High Resolution Melt (HRM) Based Quantitative Real-Time Polymerase Chain Reaction (qPCR) Primer Design Tool for Point Mutation Detection

A dissertation submitted for the Degree of Master of Information Technology

M. KAJAN
University of Colombo School of Computing
2018

DECLARATION

The thesis is my original work and has not been submitted previously for a degree at this or any other university/institute.

To the best of my knowledge it does not contain any material published or written by another person, except as acknowledged in the text.

Student Name : M. KAJAN

Registration Number : 2015 / MIT/ 021

Index Number : 15550219

_____

Signature: Date:

This is to certify that this thesis is based on the work of

Mr. M. KAJAN

under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Certified by :

Supervisor Name : Ms M.W.A.C.R Wijesinghe

_____

Signature: Date:

ABSTRACT

Single nucleotide polymorphisms (SNPs) are point mutations in the genetic material of the population which result different characteristics and disabilities in the population. Determining the abundance of these mutations may pave way for better clinical diagnostics and towards personalized medicine. Amplification Refractory mutation system – Polymerase Chain reaction (ARMS-PCR) and qPCR based Competitive amplification of differently melting amplicons (CADMA) are the most cost efficient techniques that can be used to detect these mutations. CADMA is a technique which incorporate additional mutations to increase Melting temperature (Tm) gap between different allele specific products.

Tools were developed to design these different assays which facilitate the molecular biology researchers and healthcare professionals in Molecular diagnostics sector to design their HRM, ARMS-PCR and Allele specific PCR assays with just 2 clicks. These systems were developed using the PHP server site script which hides the inside functionality from the user to keep the methodology confidential. The NCBI e-utils were used to retrieve the DNA sequence and other information from the dbSNP server and soton.ac.uk server was used to address few primer design challenges. The combination of codes were used to design this effective tool which give wide range to information to the user and enhance the usability. WordPress framework and themes were used to identify the real users and their identities in order to get in touch with them to get feedbacks. Test cases were identified and manual grey-box testing and automated integrated system testing using Katalon studio were performed to ensure the smooth functioning of the tool system.

Overall the tool works fine and makes the user easy and convenient in designing their own DNA tests for Single Nucleotide polymorphisms (SNPs) by just inserting the dbSNP ID of their interest into the tool.

ACKNOWLEDGEMENT

I would like to take the opportunity and thank all the people who spent their time to encourage and shared their knowledge to help me to complete this project. This project would not have been possible without their essential and gracious support.

I wish to express my most profound gratitude to my parents and supervisor Ms M.W.A.C.R Wijesinghe for their inspiring guidance, constant encouragement and support given to me so that this dissertation became a reality. Also I would like to thank my research supervisors Dr. N. V. Chandrasekharan, Prof. Raveen Hanwella and Prof. Varuni A. de Silva in University of Colombo faculty of science and faculty of medicine for their support and flexibility during my work to balance both my work related research and Masters in IT degree final year project. Without their guidance and persistent help this dissertation would not have been possible.

I also thank my lecturers both in Faculty of Science and in school of computing in University of Colombo for their immense support and the guidance they have given me to strengthen my knowledge. Also my thanks goes to staff of Biotechnology laboratory (Dept. of Chemistry, University of Colombo), Department of Psychiatry (Faculty of Medicine, University of Colombo), University Psychiatry Unit (ward 59) – NHSL and UCSC whom facilitated my research activities within and above their assigned duties and responsibilities.

I take this opportunity to record my sincere thanks to all colleagues and peers in Biotechnology laboratory, Department of Chemistry, University of Colombo and UCSC for their friendly support & guidance they have provided whenever I struggled to accomplish any tasks. I sincerely thank my colleagues for their continuous help, support and advices which made this report a success.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A	-	Adenine
ARMS	-	Amplification-refractory mutation system
C	-	Cytosine
CADMA	-	Competitive amplification of differently melting amplicons
CSS	-	Cascading Style Sheets
dbSNP ID	-	SNP database identifier
DMSO	-	Dimethyl sulfoxide
DNA	-	Deoxyribonucleic Acid
dNTP	-	Deoxy ribonucleotide triphosphate
dsDNA	-	Double stranded DNA
G	-	Guanine
HRM	-	High Resolution Melt
HRMA	-	High Resolution Melt Analysis
HTML	-	Hypertext markup language
mRNA	-	Messenger RNA
NCBI	-	National Center for Biotechnology Information
PC	-	Personal computer
PCR	-	Polymerase Chain Reaction
PHP	-	Personal Home Page (Server site language)
qPCR	-	Quantitative real-time PCR
RFLP	-	Restriction fragment length polymorphism
RFU	-	Relative Fluorescent Units
RNA	-	Ribonucleic Acid
rRNA	-	Ribosomal RNA

SNP	-	Single Nucleotide Polymorphism
ssDNA	-	Single stranded DNA
T	-	Thymine
Tm	-	Melting temperature
tRNA	-	Transfer RNA
U	-	Uracil
UI	-	User interface
UML	-	Unified Modeling Language
XML	-	Extensible Markup Language

CHAPTER 1: INTRODUCTION

Genetic information was transmitted from parents to their children using a sequence of 4 letter alphabet (A,T,G,C nucleotides) like the binary used in computation (Zeros and Ones) [1]. Single copy of the information available in DNA was copied several times in another type of Nucleic acid called RNA (T will be replaced by U in this process), when the information was needed by the cell [2]. This RNA was converted further into proteins which are the functional molecules inside a cell using a process called translation. This translation process identify adjacent 3 nucleic acids called codons and assemble appropriate amino acid to the protein chain [3].

Single nucleotide polymorphism (SNP) is a one nucleotide variation that occurs at a specific position in the genome (Genome: Complete set of Genes in an organism). Based on their position and type (replacement / insertion / deletion) it can alter the amino acid composition of the protein which can affect the function of an important protein [4]. For example SNP annotated by dbSNP id rs80338944 causes a replacement of G by A in the GJB2 gene which result TGG codon codes for tryptophan converted into TGA stop translation signal and the truncated proteins produced in the humans due to this causes deafness.

DNA is a fairly stable molecule when compared to RNA and proteins due to its double stranded nature. Each DNA inside a cell accompanied by reverse complementary DNA bound by the hydrogen bonds between complementary bases [A and T] and [G and C] [3]. Where G-C binding is stronger than A-T due to one additional hydrogen bond exist in-between them. These hydrogen bonds are breakable using the temperature (giving energy). The temperature at which half the amounts of hydrogen bonds were broken or in other words the temperature at which half the amount of DNA seen single stranded is called as the Melting Temperature (T_m) of DNA. When the temperature was lowered hydrogen bonds can be reformed [5].

Polymerase chain reaction (PCR) is technique used to amplify/increase the copies of the desired fragment of DNA inside a tubes using a DNA replicating enzyme called DNA dependent DNA

polymerase, suitable buffer for its activity (consist desired ion strength, pH and co-factors in appropriate concentrations) including sodium and magnesium ions, dNTPs (individual molecules of A, T, G and C) and oligonucleotides (small [15-35] stretch of DNA which will create a hybrid with the boundaries of desired gene to facilitate binding of DNA polymerase to the DNA) called primers [6].

A primer should be designed in a region which exclude the repeats occurring, should have a Melting temperature (T_m) around 55-65^{0}C (Recommended) and its composition (GC content) should be kept within 40-60% also it should not have self-annealing and dimerization abilities in order to do a successful PCR assay [7].

Quantitative real-time PCR (qPCR) is a PCR technique which use fluorescent dyes to monitor this amplification process in real-time and also it was possible to identify the amplification of desired product using intercalating dyes & melting temperature (T_m) instead of size without agarose gel electrophoresis. To do this identification there is an additional step called Melt-curve involved in this qPCR protocol in which after amplification the temperature raised to 95^{0}C and cooled in 0.5^{0}C temperature decrements. The instrument measures the fluorescent signal after each cycle during the amplification and after each temperature decrements during the melt curve formation. Each base replacement can cause a slight change in the product melting temperature which can be detected via high throughput qPCR devices only [8].

Competitive amplification of differently melting amplicons (CADMA) qPCR system haves 3 primers instead of 2 primer pairs used in traditional high resolution melt (HRM) assay [9]. Where 2 of those primers will be allele specific and one of them was modified with high GC content and other one will be lowered the GC content to produce differently melting (high Tm difference) DNA fragments in the PCR reactions specific to each alleles.

1.1 Motivation

Within the available qPCR instrumentation in laboratories of developing countries it was quite difficult to perform the HRM assay since they lack high precision in controlling the temperature in 0.1^{o}C or lower least counts. Investing in a new equipment with this functionality cost millions of rupees to the laboratory. Other widely used techniques to genotype single nucleotide polymorphisms (SNPs) having several shortfalls which are listed below in table 1 and compared with HRM based assays.

Genotyping technique	Instrumentation required	Cost per SNP	Success rate %	Duration
ARMS-PCR [10]	PCR machine, Gel electrophoresis unit	Low	< 80 %	3 hours
Taqman assay [11]	qPCR machine	High	> 99.5 %	2 hours
PCR-RFLP [12]	PCR machine, water bath, Gel electrophoresis, additional enzymes	Depending on the enzyme	> 95 %	4-5 hours
Allele specific PCR [13]	PCR machine, Gel electrophoresis unit, special taq polymerase	Medium	> 90 %	3 hours
HRM	**qPCR only**	**Low**	**> 99.5%**	**2-2.5 hours**

Table 1: Comparison of widely used SNP genotyping techniques in Sri Lanka

When compared to above assays HRM based assays are highly successful, cost effective and less time and effort to be used to perform. So developing this kind of an assay that would be compatible with any kind of qPCR instrumentation will be beneficial for developing countries. CADMA is one such method of assay but, there are currently no tools available to design PCR oligos for this assay for a beginner.

1.2 Aims and Objectives

Major aim of this project is to develop a tool to design DNA oligos (primers) for "Competitive amplification of differently melting amplicons" (CADMA) based high resolution melt (HRM) qPCR assay for cost effective and successful single nucleotide polymorphism (SNP) genotyping.

Goals of the project:

1. To retrieve the DNA sequence flanking to the SNP of interest from National center for biotechnology information (NCBI) external database through the tool
2. To interpreting the result from NCBI database into 3 different segments (Upstream region, alleles in SNP of interest and downstream region) using the tool
3. To select the primer orientation suitable for allele specific primer designing based on the comparison of GC content between forward primer designed in the upstream region and reverse primer in the downstream region
4. To design the 3 CADMA primers and calculate the primer and product sizes with corresponding melting temperatures for both primers and PCR products
5. To show the outputs to the user in the display

1.3 Scope

High resolution melt (HRM) is a developing area in molecular biology which is more helpful in identifying the disease causing mutations accurately and more rapidly than any conventional methods. Due to its accuracy and unique nature now leading biotech companies such as Thermo fisher is now recommending HRM assays to be performed to scan mutations prior to DNA sequencing to avoid the unnecessary sequencing costs on mutation-free DNA fragments [14]. The application of this tool was limited to detect only the point mutations which are replacements (No deletions and insertions). Even within the point mutations only 1type of replacement will be allowed. If the given point mutations have more than 2 genotypes an error message will be displayed. If the tool was unable to fetch the sequence from NCBI database (external) it will again display an error message "The given dbSNP ID was not found / There is a connection error".

CHAPTER 2: LITERATURE REVIEW

In this chapter the experimental biotechnology aspects of the tool to be developed together with the system requirement analysis and design strategies will be discussed briefly

2.1 BACKGROUND

Inside a cell nucleic acids encode and transmit information. Deoxyribonucleic acid, or DNA, is the linear macromolecule that encodes the hereditary data of all organisms, serving as the template for all cellular biochemistry. Enzymes in the nucleus and cytoplasm translate ribonucleic acid, or RNA, from DNA template. The DNA serves as a template for transcription and is not modified or destroyed by the process of transcription. In the cytoplasm of the cell, RNA is non-destructively translated into chains of amino acids by organelles called ribosomes. These amino acid chains are subsequently processed, or post-translationally modified into proteins.

A single strand of DNA is a linear chain of nucleotides which consist of a sugar, a phosphate group and one of four bases namely: Adenine (A), Thymine (T), Cytosine (C) or Guanine (G). The sugars of the nucleotides get connected through phosphodiester bonds so that ssDNA (or any nucleic acid) resembles a series of bases projecting from a phospho-sugar backbone chain. The DNA sequence means the order of in which the nucleotides occur from the 5’ end to the 3’ end of the phospho-sugar chain. The sequence is processed or “read” from 5’ to 3’ direction because the protein complexes which interact with ssDNA do so directionally, starting at the 5’ end of the sugar backbone and proceeding in the 3’ direction. The structure of DNA and the interpretation of “sequence” is depicted in Figure 2.1. While biochemically inaccurate, the terms ‘nucleotide’ and ‘base’ are often used interchangeably, and a nucleotide is frequently called by the name of the base which it contains [15].

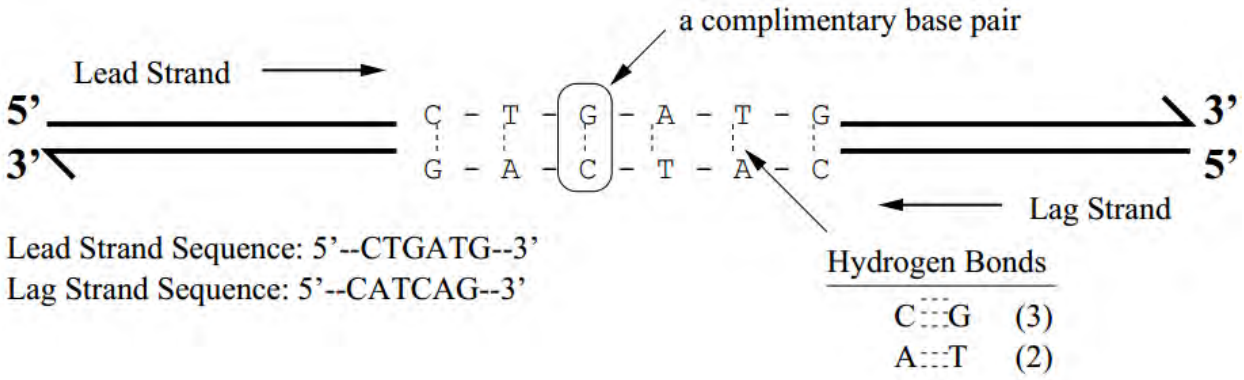

Figure 2.1: The structure of DNA

Double-stranded DNA (dsDNA) will form spontaneously whenever two reverse complementary strands of ssDNA are nearby at an appropriate temperature. "Reverse complementarity" is defined as follows: adenine is reciprocal to thymine, cytosine is reciprocal to guanine and if one strand is aligned in 5' to 3' direction then the reverse complementary strand is aligned from in 3' to 5' direction. Reverse complementarity is illustrated in Figure 2.1. When written, the top strand of dsDNA is called the "lead strand" and the bottom is the "lag strand". Also the anti-sense strand which will be copied by the RNA polymerase will be reverse complementary to the sense strand which is identical to the mRNA produced in the sequence except the thymine which will be replaced by uracil. (Refer figure 2.3) The enzymes which read dsDNA do not inherently distinguish between the lead and lag strands, they merely read one or the other strand from 5' to 3'; in this way, dsDNA can have a very high information density [16].

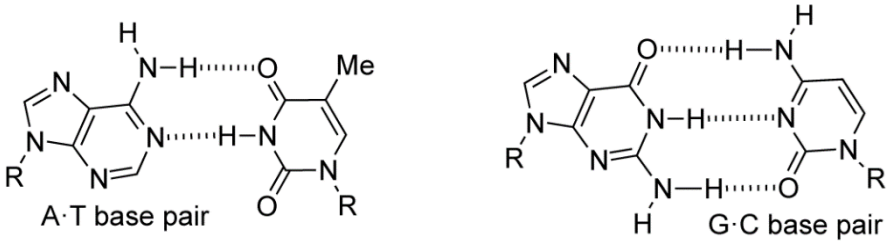

Figure 2.2: The base pairing of complementary bases through hydrogen bonding

The dsDNA formation occurs when hydrogen bond formation between complementary base pairs in ssDNA strands happens: adenine forms two hydrogen bonds with thymine and guanine forms three hydrogen bonds with cytosine as illustrated in the above figure 2.2. The process by which two strands of ssDNA combine into a single strand of dsDNA is annealing or hybridization and the reverse process is denaturation or melting. The larger the number of hydrogen bonds in a given dsDNA strand, the higher the melting temperature.

2.1.1 DNA to RNA to Protein

Construction of RNA is from a different set of nucleotides than DNA. RNA have an additional oxygen in each pentose sugar residue in the RNA backbone and also the coding sequence will have Uracil, the corresponding DNA sequence will have Thymine. For example, the DNA coding sequence 5'-a**T**gac**TT**caga-3', when transcribed into messenger RNA (mRNA), results in the sequence 5'-a**U**gac**UU**caga-3'.(Figure 2.3) During translation, an mRNA sequence was read three adjacent nucleotides at a time after the ribosomal binding; this set of 3 adjacent nucleotides is referred to as a triplet or codon. For 61/64 possible codons there is a corresponding transport RNA (tRNA) somewhere in the cell's cytoplasm that associates each that codon with a specific amino acid (AA). Remaining 3/64 codons are stop signal to terminate the protein synthesis. When translating mRNA into protein, the ribosome reads a codon, selects the tRNA corresponding to that codon, strips the amino acid off of that tRNA, adds that amino acid to the growing peptide chain, releases the tRNA and moves on to the next codon. When the ribosome reaches a termination point on the mRNA, or if the ribosome stalls or falls off of the mRNA, translation ends and the peptide chain is released to be processed into a protein [17].

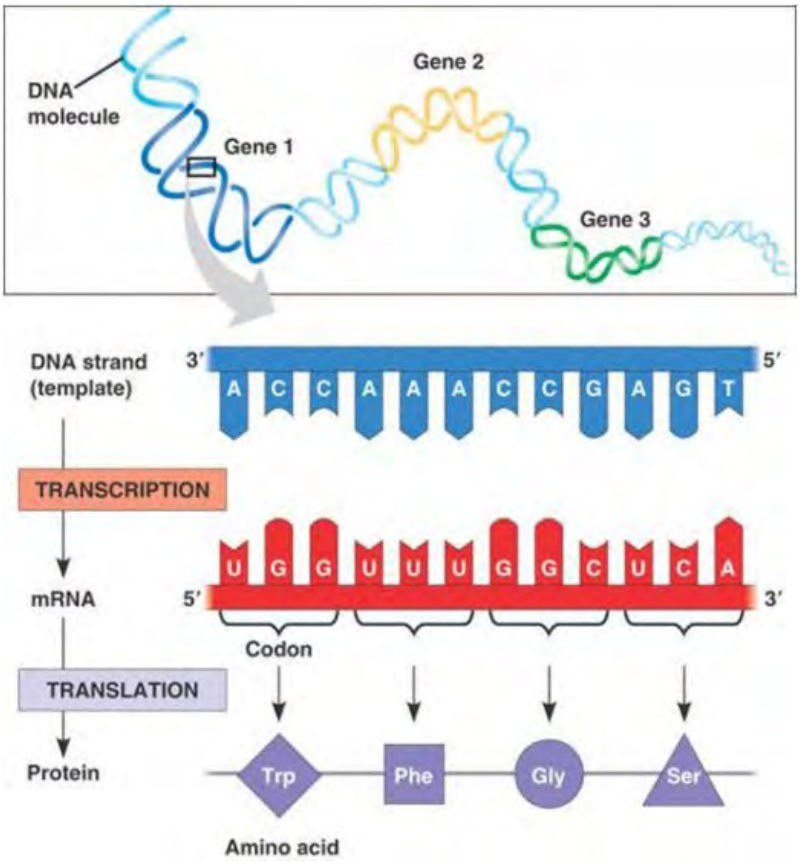

Figure 2.3: The process of transcription and translation [18]

2.1.2 Single Nucleotide Polymorphism (SNP)

Variation in a single nucleotide that occurs at a specific position in the genome (Genome: Complete set of Genes in an organism). This can be base replacement / Insertion / deletion. But in general SNP is taken as a base replacement.

Based on their position and type (replacement / insertion / deletion) it can alter the amino acid composition of the protein which can affect the function of an important protein. For example SNP annotated by dbSNP id rs80338944 causes a replacement of G by A in the GJB2 gene which result TGG codon codes for tryptophan converted into TGA stop translation signal and the truncated proteins produced in the humans due to this causes deafness.

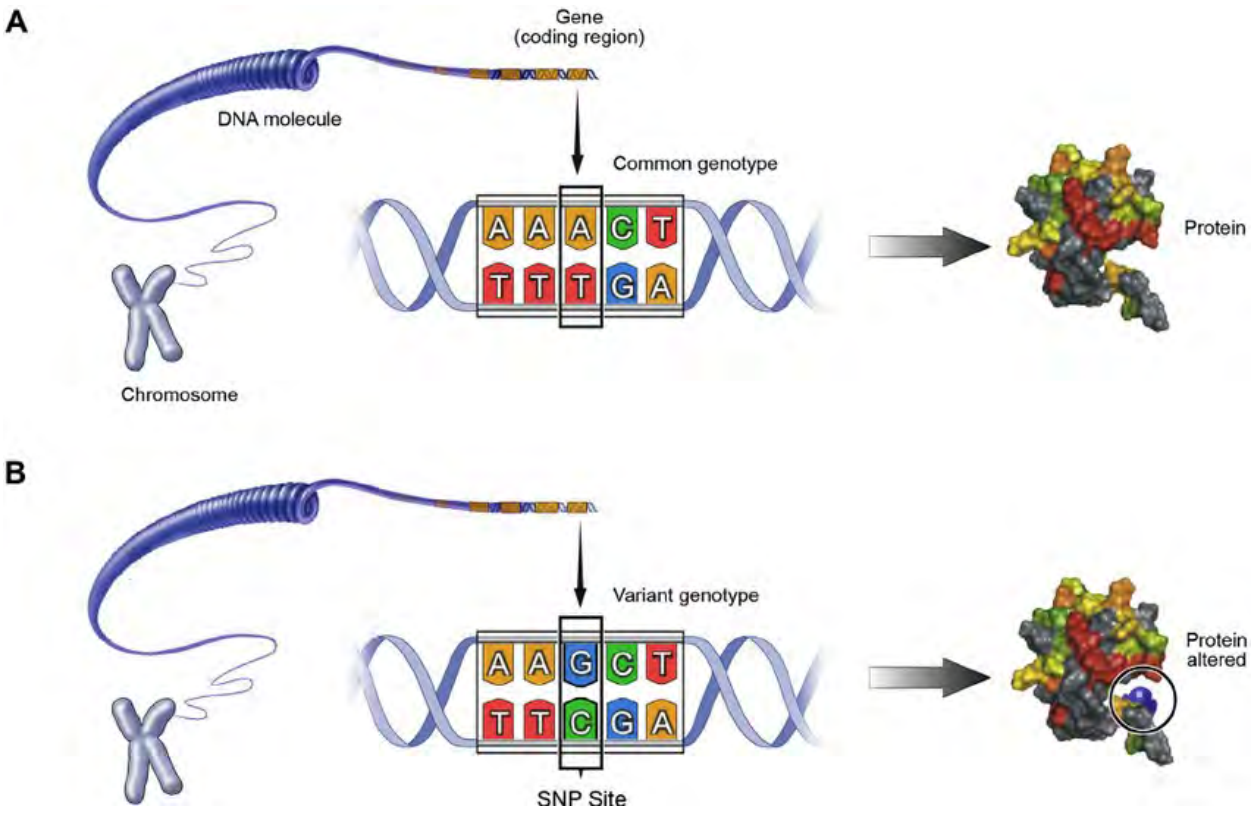

Figure 2.4: Illustration of effect of a Single Nucleotide Polymorphism (SNP) on protein [19]

1st Base		2nd Base: U		2nd Base: C		2nd Base: A		2nd Base: G		3rd Base
U		UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine	U
		UUC		UCC		UAC		UGC		C
		UUA	Leucine	UCA		UAA	STOP	UGA	STOP	A
		UUG		UCG		UAG		UGG	Tryptophan	G
C		CUU		CCU	Proline	CAU	Histidine	CGU	Arginine	U
		CUC		CCC		CAC		CGC		C
		CUA		CCA		CAA	Glutamine	CGA		A
		CUG		CCG		CAG		CGG		G
A		AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine	U
		AUC		ACC		AAC		AGC		C
		AUA		ACA		AAA	Lysine	AGA	Arginine	A
		AUG	Methionine (Start)	ACG		AAG		AGG		G
G		GUU	Valine	GCU	Alanine	GAU	Aspartic Acid	GGU	Glycine	U
		GUC		GCC		GAC		GGC		C
		GUA		GCA		GAA	Glutamic Acid	GGA		A
		GUG		GCG		GAG		GGG		G

Table 2.1: Codon table representing general nucleic acids code for different amino acids

SNPs are been classified into four classes depending on the melting temperature (T_m) difference it can cause to the DNA depending on their hydrogen bond forming ability as shown in the table I below.

SNP class	Base Change	Typical T_m Melt Curve Shift
1	C/T and G/A	Large (>0.5°C)
2	C/A and G/T	
3	C/G	
4	A/T	Very Small (>0.5°C)

Table 2.2: SNP classes and their contribution to the T_m change

2.1.3 Determining Melting Temperature

DNA melting temperature (T_M) is a measurement used to determine those assays affected by the DNA stability. Temperature at which 50% of the DNA strands in a homogenous solution become thermally denatured is identified as melting temperature (T_M). Denaturation occurs when increased thermal energy induces the dsDNA to uncoil by the breakage of the hydrogen bonds between the two complementary strands. There is an analogous measure, annealing temperature (T_A), which refers to the temperature at which 50% of strands in a solution have undergone duplex formation.

The melting temperature of a given DNA duplex can be determined using the solution salt concentration, pH, and a number of coefficients which represent the nearest-neighbor entropy of the different base pairs as determined by Santa Lucia et. al. We use the salt-corrected nearest-neighbors method to predict T_M of different primers and primer overlaps, assuming that only perfect duplexes will form. [20] The equations for calculating T_M and the tables of nearest-neighbor parameters is given below in Table 2.3.

Sequence	ΔG_0	ΔH_O	ΔS_o
AA/TT	-1	-7.9	-22.2
AT/TA	-0.88	-7.2	-20.4
TA/AT	-0.58	-7.2	-21.3
CA/GT	-1.45	-8.5	-22.7
GT/CA	-1.44	-8.4	-22.4
CT/GA	-1.28	-7.8	-21
GA/CT	-1.3	-8.2	-22.2

CG/GC	-2.17	-10.6	-27.2
GC/CG	-2.24	-9.8	-24.4
GG/CC	-1.84	-8	-19.9

Table 2.3: Table of Unified Nearest-Neighbors parameters, adapted from [21].

2.1.4 PCR amplification

The polymerase chain reaction (PCR) is a widely used biological technique that amplifies specific genes that may only be available in small quantities [7]. The advent of PCR has changed the outlook of genetic research in the community. Over the years PCR and its different variants have found an application in several fields. For example, amplification of these small DNA quantities has helped research teams involved with the Human Genome Project. [22] It could also help in experiments like identifying new DNA sequences and placing them into an already existing classification system, or in diagnosing a disease etc.

2.1.5 PCR requirements

PCR is a multi-step process and thus the success of the entire process relies on the successful completion of every step. For the successful execution each step of PCR requires some specific elements. These elements include: a DNA template that contains the region that is to be amplified; and a primer pair of two short oligonucleotide sequences called the forward primer and the reverse primer. A forward primer is an oligonucleotide sequence complementary to the DNA target anti-sense strand. The reverse primer is complementary to the 3’ end of the sense strand of the target DNA.

It also requires a DNA polymerase which is an enzyme that catalyzes the polymerization of nucleotides into a strand complementary to a given template strand. Thus for the synthesis of this new strand, the polymerase requires deoxynucleotide triphosphates (dNTPs). These dNTPs are the nitrogen bases that are the building blocks of DNA and include Adenine, Guanine, Cytosine, Thymine commonly known as A,G,C,T respectively. All these along-with buffer solution, certain ions like Magnesium, Potassium etc. help in amplifying small quantities of DNA into millions of copies [23]. Finally, PCR is performed in small test tubes inside a thermal cycler.

2.1.6 PCR amplification procedure

The PCR process can be sub-divided into two major steps: amplifying the target DNA and checking the amplified region for the desired target. Figure 2.5 [24] gives an outline of a sample DNA amplification into millions of copies using PCR.

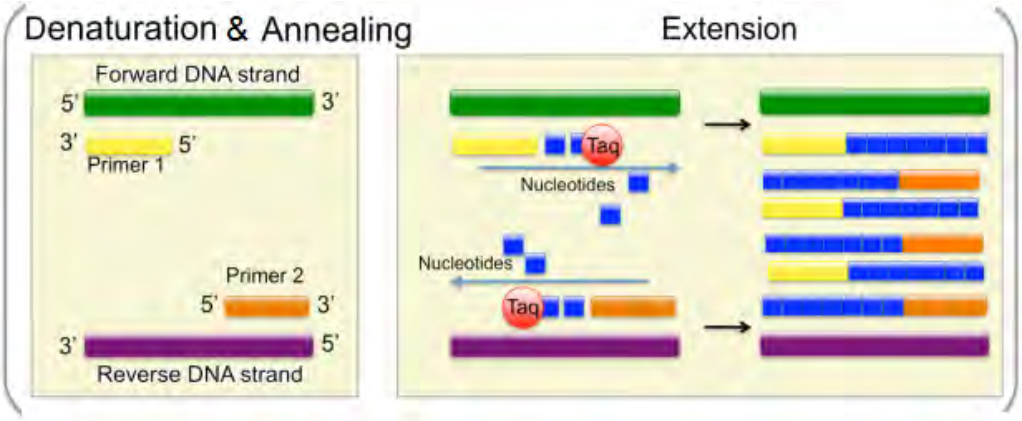

Figure 2.5: Schematic diagram for illustrating the DNA amplification during the PCR cycle

1. Amplifying the target DNA the process of amplification of DNA involves a cycle of phases that are run several times, thus amplifying the target into thousands and millions of copies. All these phases are heavily dependent on temperature changes and thus need to be carried inside a thermal cycler. This series of internal phases are: denaturation; annealing; and elongation.

Denaturation is the first phase in PCR. In this, the double stranded DNA template is exposed to a temperature of around 94°C. At this temperature, the hydrogen bonds joining the nucleotides on the complementary strands are broken, thus giving single stranded DNA molecule. This breaking up of the double stranded DNA to single stranded is called DNA melting.

Annealing is the next phase in the DNA amplification step. The temperature for this phase depends on the melting temperature of the primers. Primers (oligonucleotides) are present in the tube along-with the single stranded DNA (result of denaturation). These are freely moving and due to their instability, they try to bind (form ionic bonds) to another single stranded sequence, to a form a stable double strand. If a primer is specific to a certain DNA target, it will go and anneal to it making a short but stable connection. DNA polymerase then acts upon this to complete the double strand in the next phase.

Elongation is the phase of the amplification cycle where the DNA polymerase works to form dsDNA region. For this, the temperature is changed to 72°C which is appropriate for the working of the polymerase and also helps in breaking any unstable primer-template connections. In this phase, DNA polymerase acts on the product of the annealing phase. It starts elongating the primer bound to the DNA target by adding nucleotides to the 3' end of the primer. These nucleotides are complementary to the corresponding nucleotide on the target strand.

Since the forward primer (binding to the anti-sense strand) and a reverse primer (binding to the sense strand) for the same template exist in the same tube, in one cycle 2 copies of the target are generated. Thus, with every cycle the copy number increases exponentially. Generally 30-40 PCR cycles of run for a single target DNA generates millions of copies of the same target.

2. Checking the amplified DNA product After the DNA amplification step, it is important to check the amplified region before using it in any other application. Research teams need to make sure that the product represents the expected target. For carrying out this step, they use Gel electrophoresis. They use an agarose-based gel to check for the size of the PCR product. The amplicon is compared with a DNA ladder that consists of DNA molecules of known size. The results of this experiment assist the researchers by providing them with several characteristics of the product, like size, presence of multiple primer binding locations etc., that then help in deciding whether the PCR was successful or not.

2.1.7 Quantitative real-time PCR (qPCR)

Quantitative PCR was first developed in 1993 by constructing a system that detects PCR products as they accumulate [25]. Initially, this "real-time" system included an intercalating dye incorporated into the DNA during amplification, a modified thermal cycler with ultraviolet (UV) light source to illuminate the samples, and detection of raw fluorescence output from a CCD-camera. Amplification produces increasing amounts of dsDNA that bind intercalating dye, therefore increasing the intensity of the fluorescent signal as more copies of DNA are produced. The system plots the relative fluorescence units (RFU) on the y-axis and the cycle number on the x-axis.

Quantitative PCR can be divided into 3 distinct phases: exponential (geometric) amplification, linear amplification and the plateau region. During the exponential phase of the PCR process, the reaction results in a theoretical doubling of amplicons with each cycle. At the beginning of the exponential phase, the baseline establishes the threshold and occurs when the fluorescence signal is consistent with the normal background levels. During the amplification process of a sample, the point at which the level of fluorescence exceeds the threshold is referred to as the cycle threshold (CT). The CT value is inversely related to the template concentration at the beginning of the qPCR process, therefore it is lower for a sample with a higher initial concentration and is higher for a lower concentration sample [26].

2.1.8 Detection Chemistries

There are different approaches that can be used for fluorescence-based detection assays including fluorogenic probes, DNA binding dyes, scorpion probes, and molecular beacon technology [27]. The two most common methods used are the fluorogenic 5' nuclease assay – more familiarly known as TaqMan – and SYBR Green intercalation. The result of these assays is an increase in fluorescence intensity proportional to the amount of amplicon produced.

2.1.9 Melt Curve Analysis

The Watson and Crick double helix structure of a DNA molecule consists of two anti-parallel strands that are linked together through a process known as hybridization, where nucleotides pair up through the formation of hydrogen bonding between complementary bases [28]. In Chargaff's base pairing rule, it states that adenine (A) forms a basepair with thymine (T) and guanine (G) forms a basepair with cytosine (C), which therefore means that the percentage of bases in DNA should be equal for A and T as well as G and C [29]. Two hydrogen bonds are formed between AT basepairs and three hydrogen bonds between GC basepairs. Therefore more energy is required to break GC bonds and they have a higher thermal stability than AT bonds [28]. A DNA molecule may be rich in AT content at one region and rich in GC content at other regions. The molecule would therefore be melted in segments since the AT pairs would separate before the GC pairs as the temperature is gradually increased.

The process of denaturation is reversible. Once a fragment of DNA is exposed to high temperatures it will separate into its two strands. However, when the temperature is reduced the ssDNA will adhere to its complementary sequence and anneal to each other. The process in which the complementary strands re-form to their original conformation is known a renaturation (or reannealing).

The melt curve analysis is generally used with qPCR. It is generated after PCR amplification and indicates a change in fluorescence as temperature is raised by a small fraction of a degree, from 65^{o}C and slowly increasing to 90^{o}C. The difference in fluorescence is used to determine the melting temperature (Tm), the temperature at which amplicon dissociation occurs. The melt curve software plots the rate of change of the relative fluorescence units (RFU) with time (T) (-d(RFU)/dT) on the Y-axis versus the temperature on the X-axis.

The Tm of a giver sample is obtained by product length, sequence, GC content and strand complementarity [30]. Therefore, nonspecific amplification products such as primer-dimer artifacts, mis-priming, and inhibitor binding will result in alterations to the melting temperature and curve. The unique shape of the curve is also a function of the DNA sequence being melted. This allows amplicons containing different length and sequence variants in single closed tube to be discriminated on the basis of melt curve shape, irrespective of whether they share the same Tm.

2.1.10 High resolution melt (HRM)

Recent studies by University of Utah and Idhao technologies found that melt curve analysis can be used to study high throughput genotyping and epigenetic changes in DNA. A technique being adopted currently for these purposes is high resolution melting (HRM) analysis, which is a relatively new and rapid, post-PCR analysis method used to identify variations in nucleic acid sequences. Similar to melt curve analysis, HRM is used to characterize samples according to their dissociation behavior with increasing temperature and fluorescence detection. Samples can also be discriminated according to their sequence, length, GC content or strand complementarity. However, HRM goes beyond the power of traditional melt curve analysis by allowing the study of thermal denaturation of dsDNA in much more detail, permitting the detection of subtle changes in sequence, GC content or structure [31].

2.1.11 Competitive amplification of differently melting amplicons (CADMA)

CADMA system has three primers instead of 2 primer pairs used in traditional HRM assay [32]. Where two of those primers will be allele specific and one of them was modified with high GC content and other one will be lowered the GC content to produce differently melting (high Tm difference) DNA fragments in the PCR reactions specific to each alleles as illustrated below in figure 2.6.

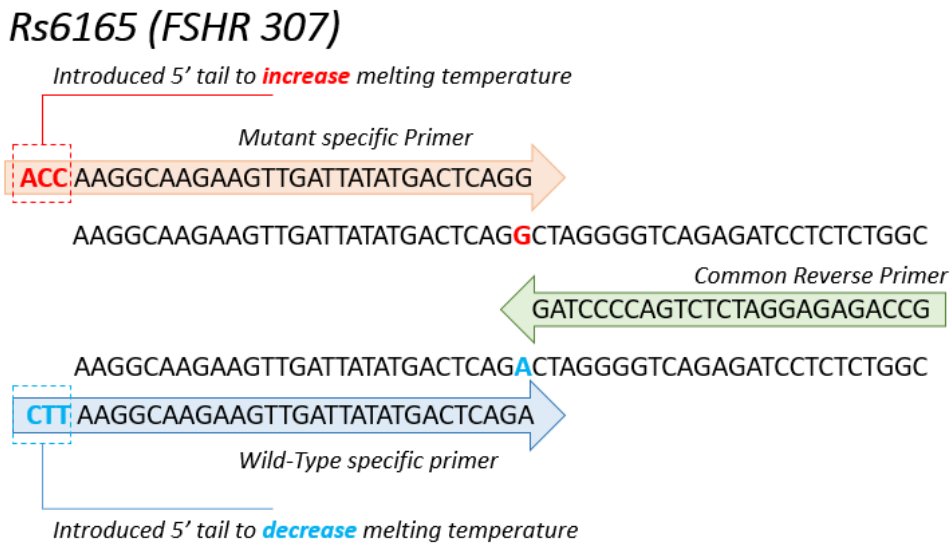

Figure 2.6: CADMA primers designed for rs6165 mutation shown in the above diagram

2.1.12 Primer Design

A primer is a single stranded oligo-nucleotide sequence varies in its length based on its application, but for regular PCR an 18-24 bp long primer is considered appropriate [33]. PCR uses two primers: forward and reverse. The forward primer is a sequence that is complementary to a region on the anti-sense strand whereas a reverse primer is complementary to some region on the sense strand of a double stranded DNA molecule.

Primer(s) are one of the requirements in PCR. The importance of primer(s) in PCR arises from the way DNA polymerase synthesizes a DNA molecule. It needs some dNTPs to be present on which it can work and synthesize the entire DNA molecule. Thus, a primer provides DNA polymerase with a short sequence of dNTPs to work on. Besides, their role in assisting DNA polymerase in

forming the double stranded molecule, primer pairs play a major role in controlling the amplification of the target region. These pairs help in restricting the amplification in the desired target region. Thus it is safe to say that primer(s) play a very important role in PCR based DNA amplification.

During sequencing of a DNA molecule, sometimes the base composition at a location is not clear. Thus there is a possibility that certain locations may have any of the four nucleotides and their different combinations. Such a condition leads to the presence of a degenerate base at those locations. As a result, besides having the four regular nucleotide bases (A, G, C, T) a primer can also contain degenerate bases [34]. Such primer(s) are called degenerate primer(s). Degenerate primer(s) can be used to target a similar (not identical) gene from different organisms. Also, if a primer is being designed using an amino acid sequence, then also designing degenerate primer(s) is useful as it helps in reducing problems that arise due to degeneracy in the triplet codon(s) where the different codons code for the same amino-acid. An example of a degenerate primer would be CGCAGGCGGTTWKRTAAGTCTG, where W means that at this position either an A or T can be observed; K represents the presence of either a G or T; and R represents either an A or a G.

2.2 REQUIREMENT ANALYSIS

This section provides a functional overview of the system. This will again be properly be divided into two parts namely functional and nonfunctional requirement analysis. Firstly we can have a look at the functional requirements of the system. The tool must be usable by any user irrespective of their level of education in biotechnology. By only entering the dbSNP ID the user will be able to design the oligo nucleotide primers for High Resolution Melt assay (qPCR based) and also able to modify the desired Melting temperature and GC content fields for getting optimum primers in exceptional situations.

System must be able to retrieve the sequences flanking to the SNP of interest from NCBI external server simply from the user input and also by editing the dbSNP ID field user can design next assay without restarting the application. System must be able to calculate the GC content and melting temperature while performing the primer design to keep them within the optimum level. At the end of the design software should be able to calculate the melting temperatures of the resulting amplicons and display it to the user. Tool should be able to identify the appropriate

flanking side that having a GC content of 50% or closest to 50% to introduce the allele specificity and also able to reverse complement the DNA sequences from the downstream region by itself.

Non-functional requirements of the tool include usability in any PC on any platform, User able to copy the context sequence and designed primer sequences from the tool as text, user ability to distinguish the differently melting amplicons with their respective primers, alleles and their melt curve using a color code.

2.3 REVIEW OF SIMILAR SYSTEMS

Absolutely there are no identical system to be compared with the HRM design tool, ARMS-PCR design tool that was developed. These are the value added versions of the existing systems where few shortfalls of the existing systems were addressed during this project.

uDESIGN (v2.4.0) is a tool developed by The university of Utah for designing primers for high-end qPCR instruments with high precision, temperature control and resolution which cannot be utilized by most of the low resolution qPCR instruments. Thus this is the first tool that adopt CADMA methodology to address this issue with similar features as uDesign such as Sequence retrieval and ease of use for the user with adjustable $MgCl_2$ concentration T_m and GC content preferences. Figure 2.7 shows the user interface of uDESIGN.

Currently existing tool to design ARMS-PCR assay using the soton.ac.uk server (Link: http://primer1.soton.ac.uk/primer1.html) requests the user to enter the sequence, mutations and mutation position manually using a conventional HTML form. In this ARMS-PCR design tool developed, an attempt was made to obtain the sequence directly from the NCBI server with just the dbSNP ID from the user input in order to minimize the potential human errors that could be caused.

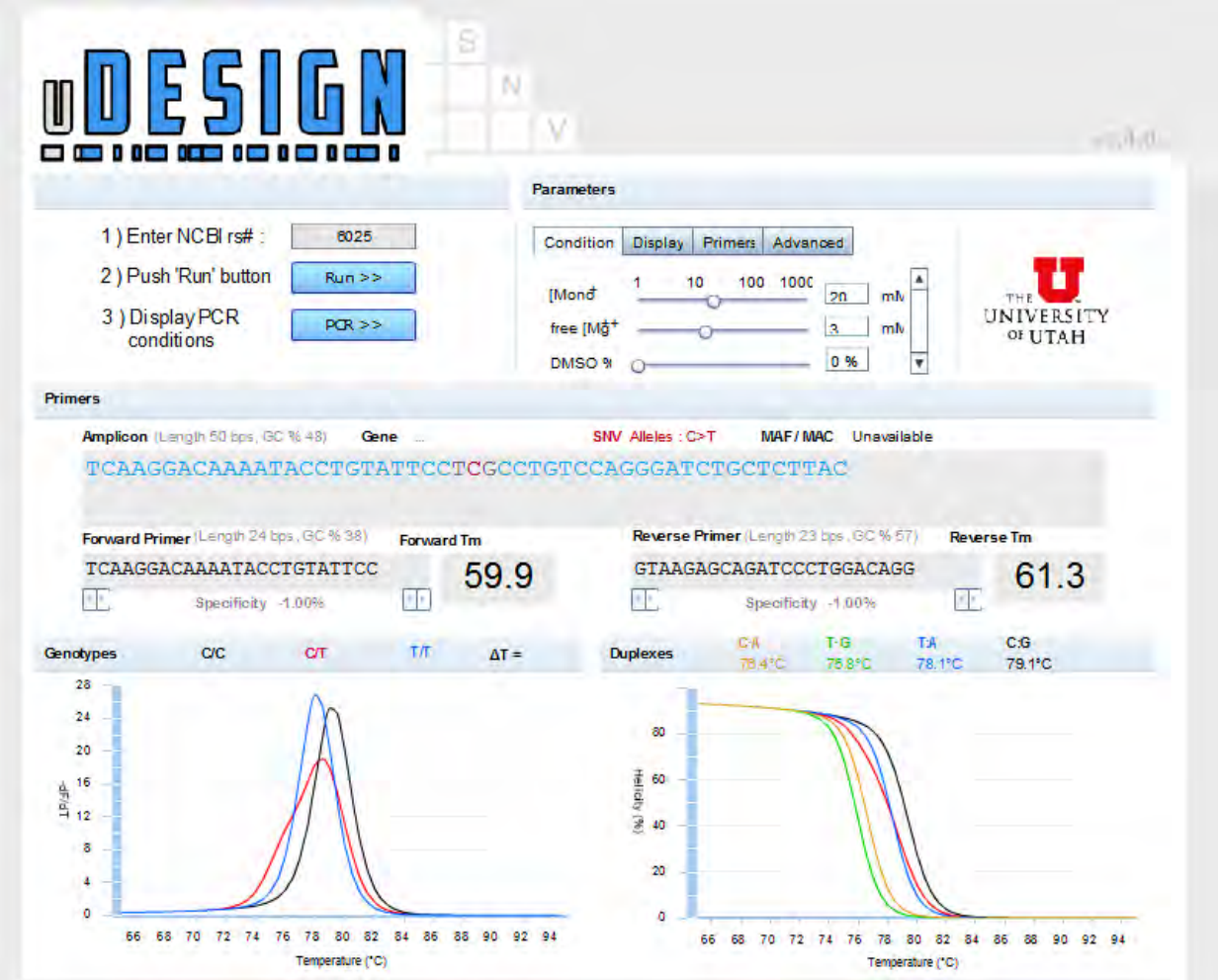

Figure 2.7: uDESIGN user interface [35]

2.3.1 Identification of research problem

The major challenge behind the CADMA based HRM primer design is the side of placement for allele specific primers in critical instances such as low and high GC content resulting instances with the modifications to increase the melting temperature difference between the allele specific PCR products. Together with few other primer design challenges and CADMA primer designing problems this gives a complex problem to be solved.

2.3.2 Primer design challenges

Due to the role the primer(s) play in carrying out a successful PCR experiment, the designing of primer(s) is of utmost importance and thus needs to be done very carefully. Over the years [36]

[37], researchers have laid out certain important parameters that need to be taken into consideration while designing primer(s), a few important ones are:

(i) Primer melting temperature,

(ii) Guanine (G) + cytosine (C) content in the primer sequence,

(iii) Length of the primer,

(iv) Size of the product that is being amplified,

(v) Absence of complementarity within a primer (self-complementarity) or complementarity with the other primer (leads to primer-dimers),

(vi) Temperature at which the primer binds to the given DNA template (annealing temperature),

(vii) The nucleotide residue at the 3'end of the primer and also the residues in the 3' clamp (3-4 residues on the 3' end).

There are certain other factors that need to be considered and cannot be left out like, in case of PCR, the difference between the melting temperature of the forward and reverse primers needs to be taken into account, also, the absence of multiple primer binding sites on the same template could also affect the product being amplified. All these parameters help in designing effective primer(s), which in-turn help in successful PCR.

2.4 Alternative Design Strategies

Best to generate three alternative design strategies namely Low-end (Provide all required functionalities user demand with a system minimally different from existing), High-end (solve problems and provide extra features users desire) and Mid-Range (Compromise features of both high and low end). In case of this the tool intended to be developed follows the mid-range design strategy which satisfy the minimal difference in case of ARMS-PCR design tool just by saving time taken for the obtaining and entering DNA sequence, Alleles and position from dbSNP database to Primer1 tool in soton.ac.uk server using the automation of the process using NCBI e-Utils.

The high-end solution was given to users in the HRM-Assay design tool by providing all necessary information required to perform the CADMA type HRM assay, (Primer sequences, Melting Temperatures, GC content and primer length) to analyze the results (Melting temperatures of expected amplicons to be generated by the Polymerase chain reaction, amplicon size to detect it on a agarose gel), to compare with already available data (Minor allele frequency and minor allele in the population if available) and finally to present the results in scientific journals with relevant background information (Chromosome, Gene, type and link to additional information from dbSNP server)

CHAPTER 3: PROBLEM ANALYSIS & DESIGN METHODOLOGY

3.1 REQUIREMENT ANALYSIS

There are two types of requirements that should be addressed in all the systems / tools. This section provides a functional overview of the system in both user and the system perspective called functional and nonfunctional requirements.

3.1.1 Functional requirements

- The tool must be usable by any user irrespective of their level of education in biotechnology
- User must be able to enter the dbSNP ID and able to modify the desired Melting temperature and GC content fields
- System must be able to communicate with NCBI external server to retrieve the flanking sequences for the given dbSNP id from the user input
- User must be able to modify the dbSNP ID for the next design without restarting the application
- System must be able to calculate the GC content and melting temperature while performing the primer design to keep them within the optimum level

3.1.2 Non-functional requirements

- Tool must be usable in any PC
- User can copy the context sequence and designed primer sequences for his reference and to order custom primers
- User must be able to see the differently melting amplicons with their respective primers and melt curve using the color code
- Application must not be platform independent
- Project will be implemented mainly in PHP

3.1.3 Identification of research problem

The major challenge behind the CADMA based HRM primer design is the side of placement for allele specific primers in critical instances such as low and high GC content resulting instances with the modifications to increase the melting temperature difference between the allele specific PCR products. Together with few other primer design challenges and CADMA primer designing problems this gives a complex problem to be solved.

a. Common Primer design challenges

Due to the role the primer(s) play in carrying out a successful PCR experiment, the designing of primer(s) is of utmost importance and thus needs to be done very carefully. Certain important parameters that need to be taken into consideration while designing primer(s), a few important ones are:

1. Primer melting temperature,
2. Guanine (G) + cytosine (C) content in the primer sequence,
3. Length of the primer,
4. Size of the product that is being amplified,
5. Absence of complementarity within a primer (self-complementarity) or complementarity with the other primer (leads to primer-dimers),
6. The nucleotide residue at the 3'end of the primer and also the residues in the 3' clamp (3-4 residues on the 3' end).

There are certain other factors that need to be considered and cannot be left out like, in case of PCR, the difference between the melting temperature of the forward and reverse primers needs to be taken into account, also, the absence of multiple primer binding sites on the same template could also affect the product being amplified. All these parameters help in designing effective primer(s), which in-turn help in successful PCR.

b. CADMA Primer design problem

In addition to above mentioned conventional PCR primer design problems CADMA based primer design faces few other challenges such are:

1. Determining the direction which should be selected for allele specific primer designing depending on the GC content
2. Identifying the sites to be modified to widen the melting temperature difference between differently melting amplicons
3. Determining the melting temperatures of the amplicons to be produced
4. Repeating the procedure if the widened gap between differently melting amplicons does not exceed the threshold

3.2 SYSTEM DESIGN

System design is the most crucial thing for a tool due to its ability to dominate the usability of the tool. The use-case diagram for the purposed system is illustrated below in Figure 3.1

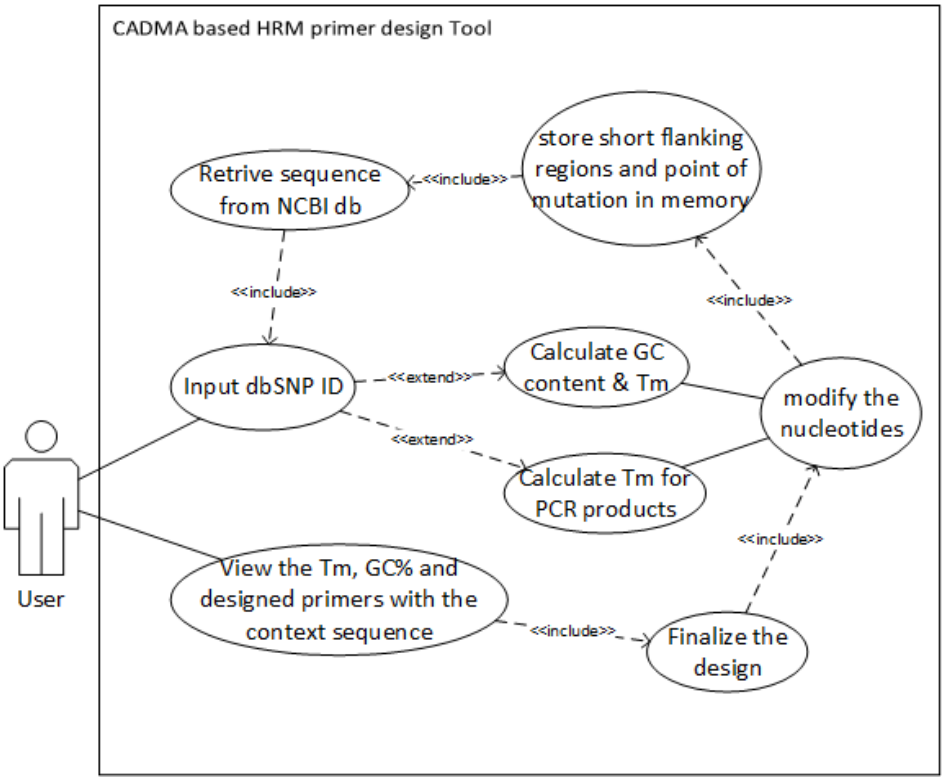

Figure 3.1: Use case diagram for purposed tool

3.2.1 HRM oligo design tool workflow

The below figure shows a general workflow for HRM oligo design tool. The user inputs a dbSNP ID / rs number and the other parametric values in case they vary from the default parameters. HRM oligo design tool then checks the oligos for all the parameters. The oligo that best match with all these conditions will be displayed in their respective fields together with expected resulting amplicons and their melting temperatures.

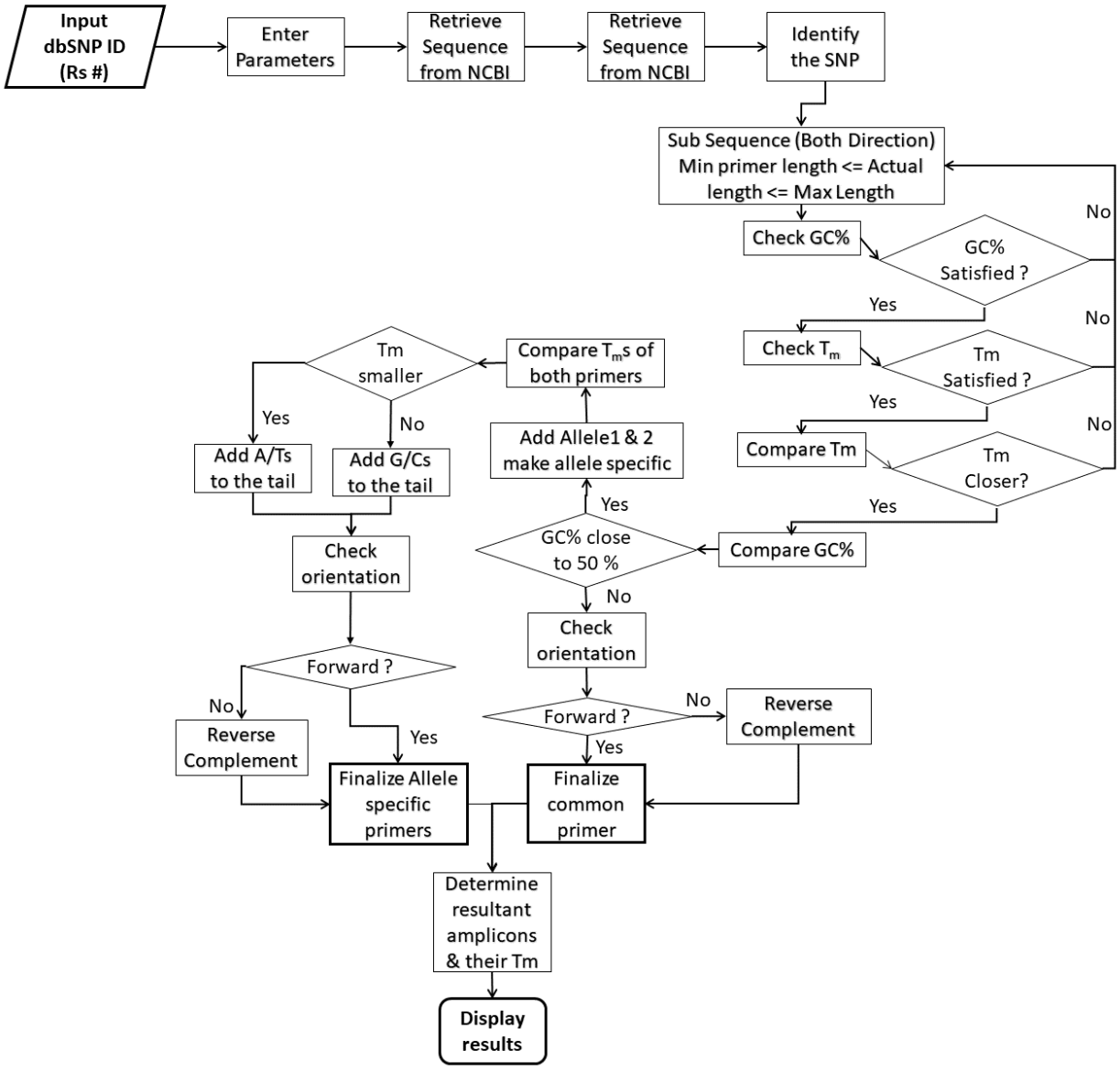

Figure 3.2: Workflow of HRM oligo design tool

This tool doesn't have a database and the data used will we retrieved from an external (NCBI) database upon user enters the dbSNP ID on the field provided by the tool. So that there is no database tables / fields are available and hence no database model diagrams such as ER or EER diagrams and database schemas not relevant to the intended project.

3.2.2 UI design

User interface of the "High Resolution Melt (HRM) – Oligo design Tool" was constructed by improving the existing tool called "uDesign" designed by University of Utah by requiring the minimal data from the user input (Only Rs number / dbSNP ID required and others have the optimum number by default). Once the user press design button after entering the required data two allele specific primers and one common primer will be deigned according to the CADMA principles. Also the melting temperatures (Tm), GC content, Orientation, Length and Melting temperatures of the products will be displayed. Each allele specific product will be displayed by a distinguished color for the user to identify the matching parameters very quickly in a short period of time. Figure 3.3 illustrates the proposed system design.

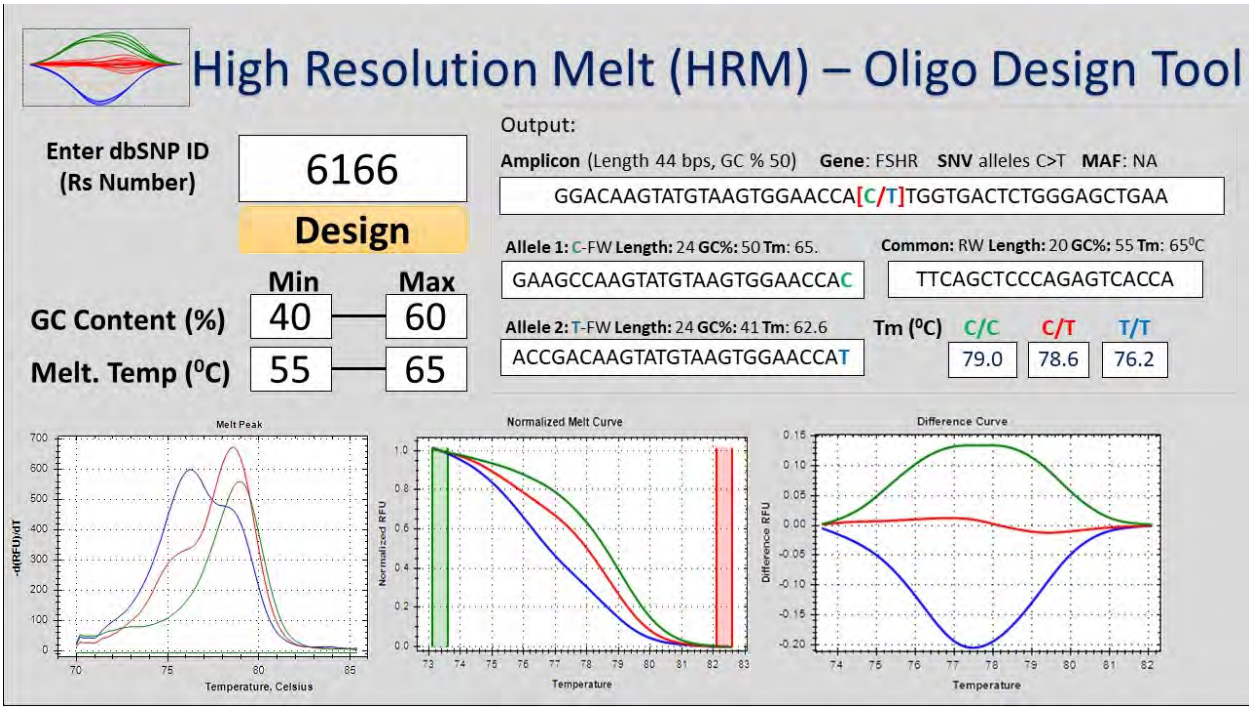

Figure 3.3: UI design of the purposed system

3.3 Quality of Solution

The most accepted and updated nearest neighbor method for Melting temperature (Tm) calculation is used in this process to calculate the Tm of the primers. The algorithm similar to Primer3 used for finding the most appropriate pairs by the repetitive calculation of Tm, GC content and length adjustment during this tool development to keep the GC content within 40-60% and Tm within 60-70 ^{o}C for the best results.

Furthermore this algorithm compares the primers and select the pest primer to introduce allele specific bases and mutations to better separate the melt curves by increasing the Tm difference between 2 allele specific amplicons. This was done by estimating the closeness of Tm to 50% by comparing the primers from both orientations. To add the mutations in the better way again the calculated Tm of different allele specific products taken into account and G and/or C residues will be added to the allele specific primer with highest Tm. And A/T residues will be added to the allele specific primer with the lowest Tm. These additions will be made to the tail (5' end) of the primer to prevent the disruptions that can be caused to the primer binding. The tool was expected to produce best results except GC and AT rich regions. In addition the last 5' nucleotide will be placed to enhance the effect of the Tm shift using pre-calculated patterns according to the nearest neighbor enthalpy and entropy values.

The high value customer segment who is interested in the HRM based qPCR SNP genotyping assay design could be captured if there are login and user only functions enabled in the tool. So that the social login system was selected in mind to get the real user information then forcing them to sign up to the system using different usernames.

3.4 Implementation methodology

The agile methodology or iterative approach will be followed in making the tool / program. This is because the tool itself contains many modules and complex algorithms. So starting on a small backbone and expanding the coding and integrating different modules will be meaningful and safe way to construct this total tool.

Product development work will be broken into small increments that minimize the amount of up-front planning and design. Iterations, or sprints (short time frames). All functions including planning, analysis, design, coding and unit testing will be carried out for each iterations. This minimizes overall risk, allows the product to adapt to changes quickly and to learn this process by the developer in more meaningful way. An iteration might not add enough functionality, but the goal is to have an available release (with minimal bugs) at the end of each sprint. The unit testing and overall testing will be done by an experienced biotechnology person and the results were evaluated manually using the other available tools for Tm and GC content calculations within the given parameters to identify the bugs.

CHAPTER 4: IMPLEMENTATION

4.1 Introduction

In an IT project the implementation is the crucial part of the project. The way we design the user interface, how the efficient is our algorithm and how improved the reusability of the coding will contribute to the success of the implemented project. In this project we have paid more attention on code reusability, UI, Ease of use / User friendliness as the core factors.

4.2 Individual modules

Individual modules for this program was built and tested separately using manual testing methods. Few of the modules which have been identified to perform as a separate tool for researchers has been implemented both as separate tool and also integrated with the main tool developed.

4.2.1 Module to retrieve DNA sequences from dbSNP dataset (NCBI server)

The algorithm to retrieve the DNA sequence flanking to the SNP of interest was developed. By giving the dbSNP ID (or Rs number) the upstream, downstream sequences with the mutation and mutation position can be retrieved from NCBI database.

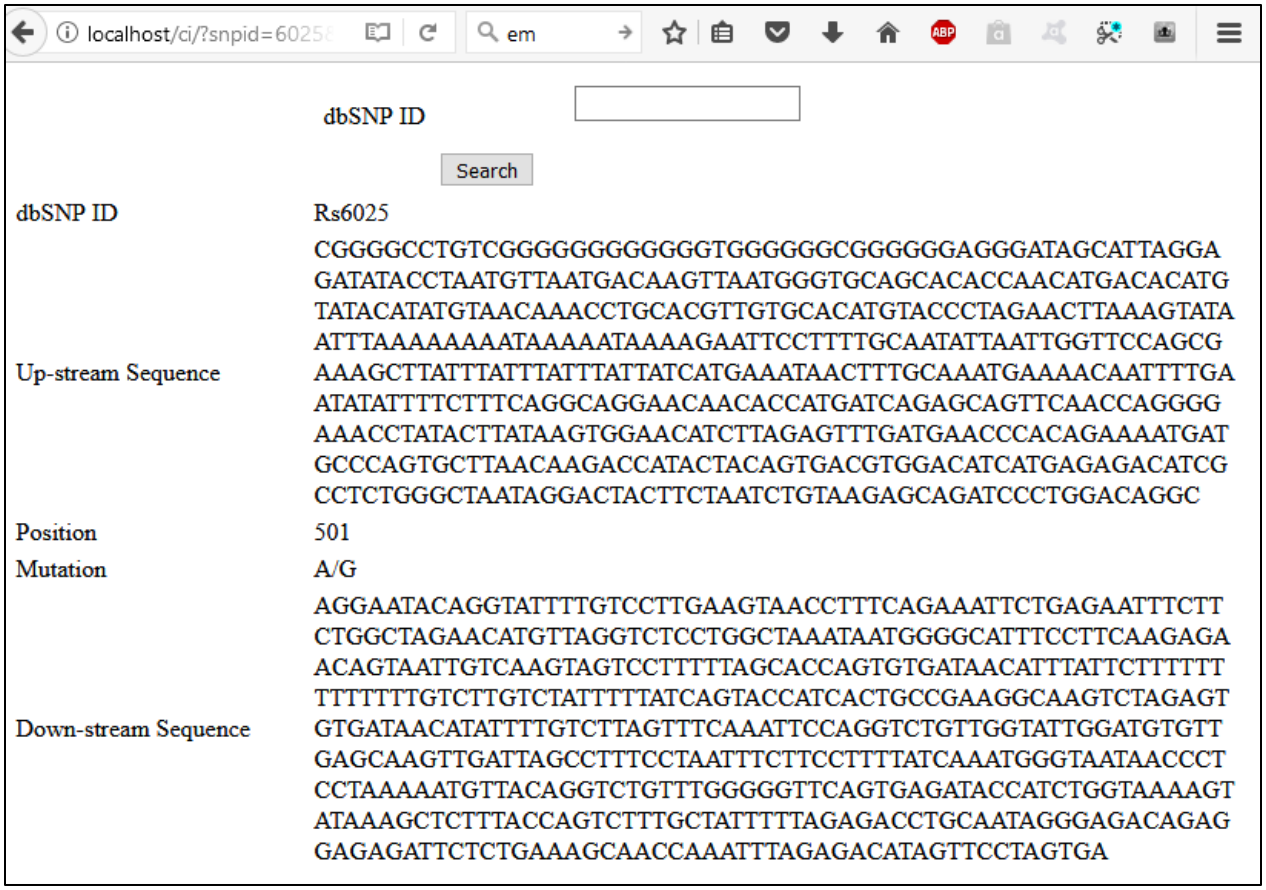

Figure 4.1: Module that retrieve reference sequence for the SNP of interest

4.2.2 Module to submit the sequence to soton.ac.uk to retrieve primer sequences

A part of the above shown DNA sequence was used as an input for this module to submit the DNA sequence, mutation position and mutation alleles along the default parameters to the soton.ac.uk server to get the most appropriate the ARMS-PCR primer pairs from the external server code execution.

```
<?php
function readHtmlLine($html, $n) {
    $dom = new DOMDocument();
    $dom->preserveWhiteSpace = false;
    $dom->loadHTML($html);
    $dom->formatOutput = true;
    $lines = explode(PHP_EOL, $dom->saveHTML());
    return $lines[$n-1];
}

$file = file_get_contents(
'http://primer1.soton.ac.uk/runprimer1.cgi?CGGGGCCTGTCGGGGGGGGGGGGTGGGGGGCGGGGGGAGGGATAGCATTAGGAGATATACCTAATGTTAATGACAAG
TTAATGGGTGCAGCACACCAACATGACACATGTATACATATGTAACAAACCTGCACGTTGTGCACATGTACCCTAGAACTTAAAGTATAATTTAAAAAAAATAAAAATAAAAGAATTCCT
TTTGCAATATTAATTGGTTCCAGCGAAAGCTTATTTATTTATTTATTATCATGAAATAACTTTGCAAATGAAAACAATTTTGAATATATTTTCTTTCAGGCAGGAACAACACCATGATCA
GAGCAGTTCAACCAGGGGAAACCTATACTTATAAGTGGAACATCTTAGAGTTTGATGAACCCACAGAAAATGATGCCCAGTGCTTAACAAGACCATACTACAGTGACGTGGACATCATGA
GAGACATCGCCTCTGGGCTAATAGGACTACTTCTAATCTGTAAGAGCAGATCCCTGGACAGGCCAGGAATACAGGTATTTTGTCCTTGAAGTAACCTTTCAGAAATTCTGAGAATTTCTT
CTGGCTAGAACATGTTAGGTCTCCTGGCTAAATAATGGGGCATTTCCTTCAAGAGAACAGTAATTGTCAAGTAGTCCTTTTTAGCACCAGTGTGATAACATTTATTCTTTTTTTTTTTT
GTCTTGTCTATTTTTATCAGTACCATCACTGCCGAAGGCAAGTCTAGAGTGTGATAACATATTTTGTCTTAGTTTCAAATTCCAGGTCTGTTGGTATTGGATGTGTTGAGCAAGTTGATT
AGCCTTTCCTAATTTCTTCCTTTTATCAAATGGGTAATAACCCTCCTAAAAATGTTACAGGTCTGTTTGGGGGTTCAGTGAGATACCATCTGGTAAAAGTATAAAGCTCTTTACCAGTCT
TTGCTATTTTTAGAGACCTGCAATAGGGAGACAGAGGAGAGATTCTCTGAAAGCAACCAAATTTAGAGACATAGTTCCTAGTG+501+C+T+28+30+26+200+300+100+1.5+1.1
+65+80+50+80+20+8.00+3.00+50+50+10');
echo readHtmlLine($file, 3);

?>
```

Figure 4.2: Code to retrieve ARMS-PCR primer pairs from soton.ac.uk server

4.2.3 Measuring GC content function

Measuring GC content becomes vital role to make a decision on which orientation does the allele specific primer will be positioned and also its an indication of success rate of the assay.

```
function CountCG($c){
        $cg=substr_count($c,"G")+substr_count($c,"C");
        return $cg;
        }

function CountATCG($c){
        $cg=substr_count($c,"A")+substr_count($c,"T")+substr_count($c,"G")+substr_count($c,"C");
        return $cg;
        }

function GCcontent($primer){
        $cg=round(100*CountCG($primer)/strlen($primer),1);
        return $cg;
        }
```

Figure 4.3: GC content measurement functions

4.2.4 Melting Temperature measurement

Melting temperature of the primers is the key determination factor for assay attribute called PCR primer annealing temperature. Also it was used to measure the discrimination of alleles in the final melting curve generation phase in qPCR assay. The traditional equation of melting temperature calculation was no longer valid since it has more error on longer oligos. So that the Gibbs free energy equation and nearest neighbor based calculation algorithm was adopted from the biophp.org open source biotools web.

Old equation: 2x (A+T) + 4x (G+C) = Tmelt

Gibbs free energy equation: G = H – TS

Unified oligonucleotide $\Delta H°$ and $\Delta S°$ NN parameters in 1 M NaCl

Sequence	$\Delta H°$ kcal/mol	$\Delta S°$ cal/k·mol
AA/TT	−7.9	−22.2
AT/TA	−7.2	−20.4
TA/AT	−7.2	−21.3
CA/GT	−8.5	−22.7
GT/CA	−8.4	−22.4
CT/GA	−7.8	−21.0
GA/CT	−8.2	−22.2
CG/GC	−10.6	−27.2
GC/CG	−9.8	−24.4
GG/CC	−8.0	−19.9
Init. w/term. G·C	0.1	−2.8
Init. w/term. A·T	2.3	4.1
Symmetry correction	0	−1.4

Table 4.1: Enthalpy and entropy values for nearest neighbor bases

```
function tm_Base_Stacking($c,$conc_primer,$conc_salt,$conc_mg){

        if (CountATCG($c)!= strlen($c)){print "The oligonucleotide is not valid";return;}
        $h=$s=0;
        // from table at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC19045/table/T2/ (SantaLucia, 1998)
        // enthalpy values
        // entropy values
        // effect on entropy by salt correction; von Ahsen et al 1999
                // Increase of stability due to presence of Mg;
                $salt_effect= ($conc_salt/1000)+(($conc_mg/1000) * 140);
                // effect on entropy
                $s+=0.368 * (strlen($c)-1)* log($salt_effect);

        // terminal corrections. Santalucia 1998
                $firstnucleotide=substr($c,0,1);
                if ($firstnucleotide=="G" or $firstnucleotide=="C"){$h+=0.1; $s+=-2.8;}
                if ($firstnucleotide=="A" or $firstnucleotide=="T"){$h+=2.3; $s+=4.1;}

                $lastnucleotide=substr($c,strlen($c)-1,1);
                if ($lastnucleotide=="G" or $lastnucleotide=="C"){$h+=0.1; $s+=-2.8;}
                if ($lastnucleotide=="A" or $lastnucleotide=="T"){$h+=2.3; $s+=4.1;}

        // compute new H and s based on sequence. Santalucia 1998
        for($i=0; $i<strlen($c)-1; $i++){
                $subc=substr($c,$i,2);
                $h+=$array_h[$subc];
                $s+=$array_s[$subc];
        }
        $tm=((1000*$h)/($s+(1.987*log($conc_primer/2000000000))))-273.15;
        $tm = round($tm,1);
        return $tm;
       // print  "\n<font color=008800>  Enthalpy: ".round($h,2)."\n  Entropy:  ".round($s,2)."</font>";
}
```

Figure 4.4: Melting Temperature calculation algorithm adopted form Biophp

4.2.5 Additional functionalities for Tm and GC content

Like using the GC content to decide the orientation which should be made as allele specific measuring melting temperature also helps to determine which bases to be added to the oligo tail to increase of decrease the melting temperature difference between 2 allele specific products.

4.2.6 Javascript, CSS and boostrap for rich interaction / UI improvement

CSS style sheets and Javascrip was used to restrict the user within the realistic data input in the HRM primer design tool and Bootstrap was used to make the interface more colorful and user friendly. CSS also used to color code the different allele specific products after the melting temperature comparison.

Product with highest melting temperature was color coded with green and lowest with blue while when heterozygotes present (both allele 1 and 2 in same patient sample) the condition was denoted by red letter to match the melting curve output sample image for the user to easily determine the difference and keep those in mind when performing the assay.

4.3 Integrated Modules

Integration of individual modules and integration testing is the final step before releasing a beta version / prototype for the user evaluation and testing. This has been done be integrating individual modules into one php document and also by creating methods / functions and calling them from using the code wherever it was needed to be executed.

4.3.1 ARMS-PCR primer design Tool

The above mentioned individual modules in 5.2.1 and 5.2.2 was integrated together to obtain the automated ARMS-PCR design tool. This tool retrieve DNA sequence from NCBI in the 1st click after the dbSNP input and show the available genotypes (if only 2 genotypes present it automatically fix one each for each allele selection dropdown). This facilitate the user to design primers in just 2 clicks. Other parameters are provided with default parameters and the user (beginner) does not need to change those. This tools gives the best 2 primer sets as the output.

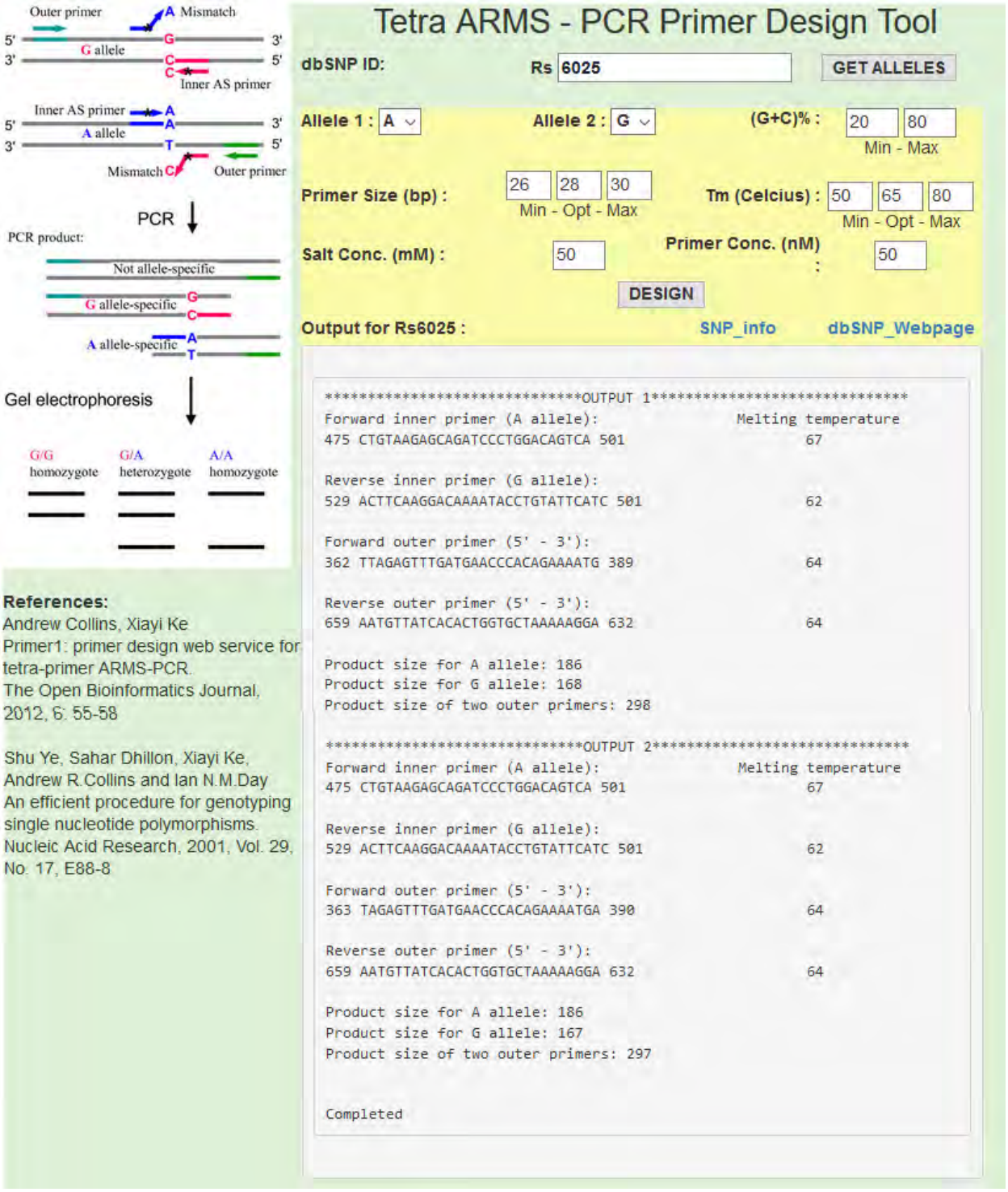

Figure 4.5: ARMS-PCR assay design tool UI with sample output

4.3.2 Integrating the modules to construct the HRM oligo design tool

Above mentioned functionalities from 4.2.1 – 4.2.6 was considered and included when making this tool functional also manual testing with few test cases was performed in order to ensure the bug free functioning of the tool.

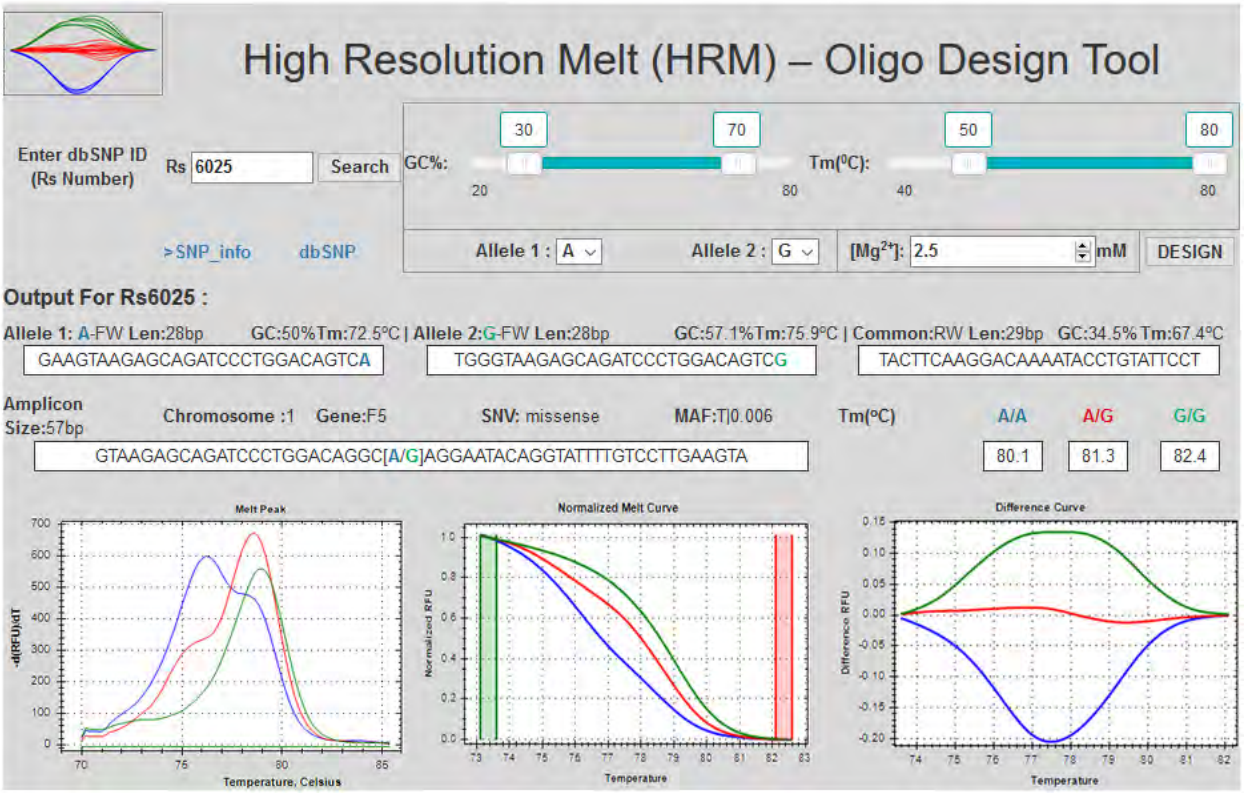

Figure 4.6: sample output of HRM oligo design tool v 1.0

4.4 Integration of tools into the WordPress

These tools were linked with the Biotechnology Forum website (www.biotech.edu.lk) webserver which was designed using the WordPress CMS tool with the aid of iframe. SNPinfo.php and ARMS-PCR design tool were integrated without any coding to allow any user to use those tools while user validation codes were used in both WordPress using “PHP Code Snippets” to make “HRM Assay Design tool” available for the subscriber only.

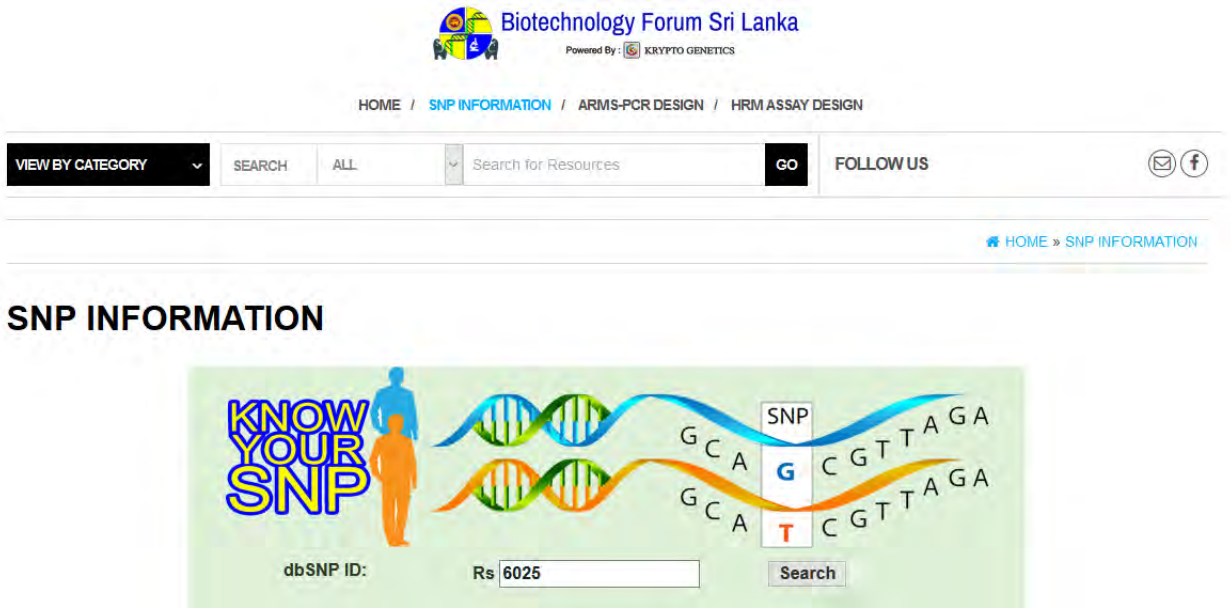

Figure 4.7: Integrated SNP information in WordPress site

4.5 Getting the list of the users

Since the real user information is a valuable resource for marketing there is an attempt was made to recruit and identify the real users using the social login function (Users able to login to the site using their social media accounts) to make them feel easier to login to the site with just few clicks. Studies suggested when there is a social login present instead of regular registration users tend to give their real information. For this purpose the customer registration function was blocked using the WordPress admin account and Social login plugin was installed. Then according to the developer instructions of the social login plugin developer accounts were created in Yahoo, Microsoft, Google and Facebook to create apps for this login functions in respective social sites and then the app ID and app secret were copied to the Social login plugin to make it enable and functional.

Figure 4.8: Social Login function placed with the login page

4.6 Manual testing

Grey box testing (combination of both white and black box) approach was adopted when performing individual unit testing manually upon building individual modules. These include:

1. Entering different dbSNP IDs and checking whether the variables loaded are correct by comparing the details with the NCBI dbSNP databse
2. Checking whether the variables are loading into the appropriate form fields in SNP information page
3. Upon development of ARMS-PCR design tool weather the correct variables are getting phrased to the soton.ac.uk server through HTTP get variables in the URL.
4. Whether the results obtained from soton.ac.uk server as the text within <pre> tags were successfully separated by the php execution and classified and displayed according to their category
5. Whether the appropriate primer was selected to function as the allele specific primer depending on the closeness of GC content to 50^{o}C

6. Using the appropriate pattern to increase the Tm difference between both PCR products from the allele specific primers was verified manually
7. Melting temperatures of primers and products was calculated using the tool available in BioPHP and compared to the values provided by the tool using the Base-stacking method which take account of concentration of different PCR constituents
8. Weather the orientations, lengths of primers and product loaded correctly in the output and high & low Tm product alleles color coded in primers, context sequence and product Tm headers according to the guideline pattern images provided by the tool

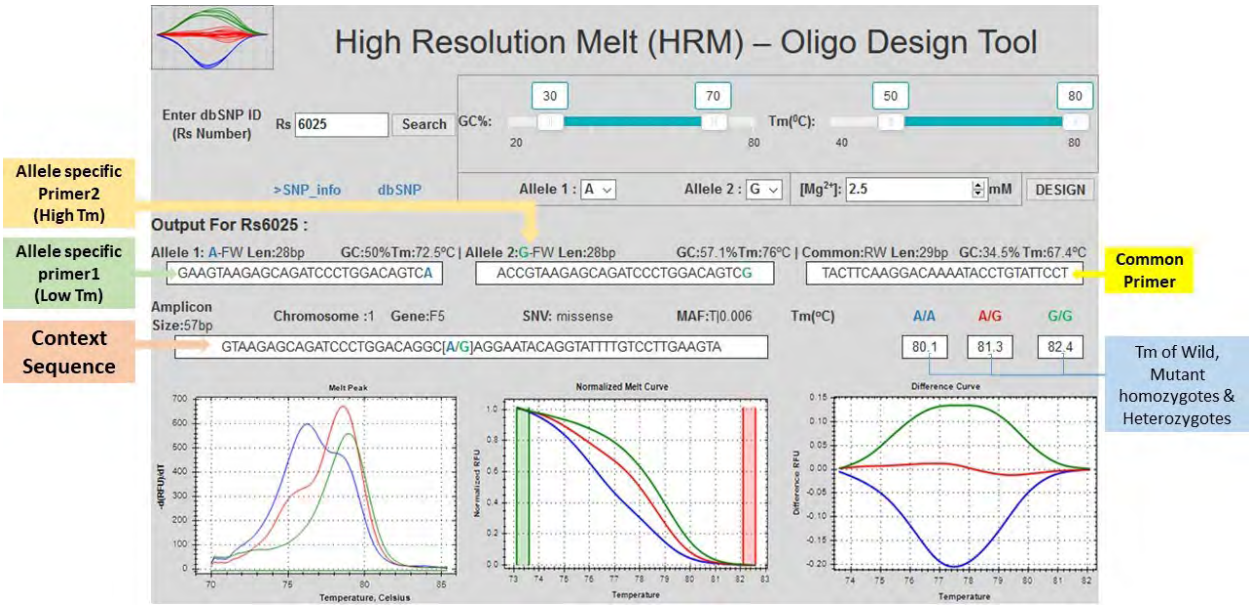

Figure 4.9: Individual components of the HRM oligo design tool

(Abbreviations: Tm-Melting temperature, $[Mg^{2+}]$- Magnesium ion concentration in PCR mixture, FW-Forward orientation, Len-Length, bp-Base pairs, MAF-Minor allele frequency; Color codes: **Blue** – Alleles that gives low Tm product and its homozygous product, **Green** – High Tm and its homozygous, **Red** – Combination of both in heterozygous state)

The following test cases used to build more functionalities of the tool to handle exceptional situations as follows:

1. db SNP ID : Rs113993960: This SNP haves indel (insertion / deletion of 2 bases) in such case the design get more complex and unable to be handled by this tool so which results an

error message by the tool. This same message also given by the ARMS-PCR design tool as well and in both tools the design button get disappeared from the tool which supposed to appear after providing the dbSNP ID.

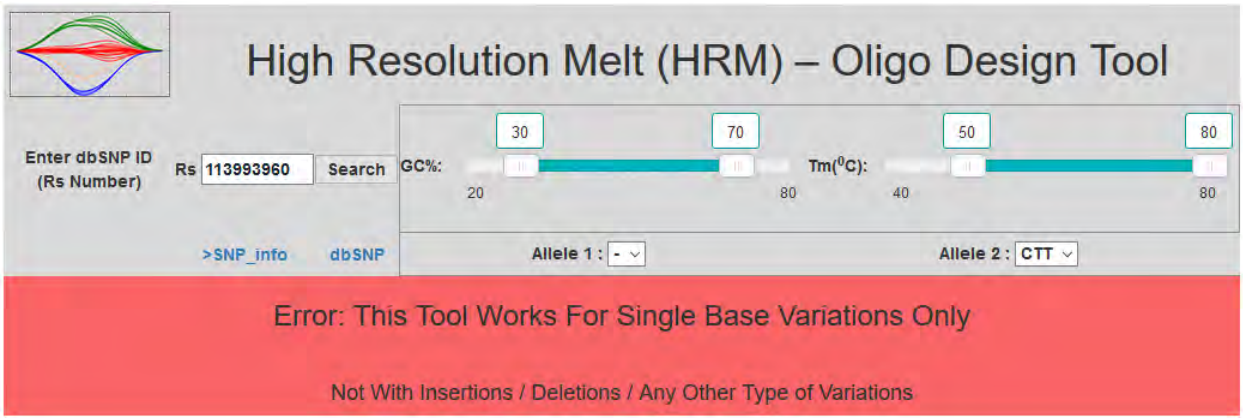

Figure 4.10: Error Message when providing dbSNP IDs for indels (Insertions / Deletions)

2. db SNP ID : Rs78655421: This SNP have more than 2 genotypes all loads into the allele selection dropdown upon input data and result an error message if both selected alleles are same

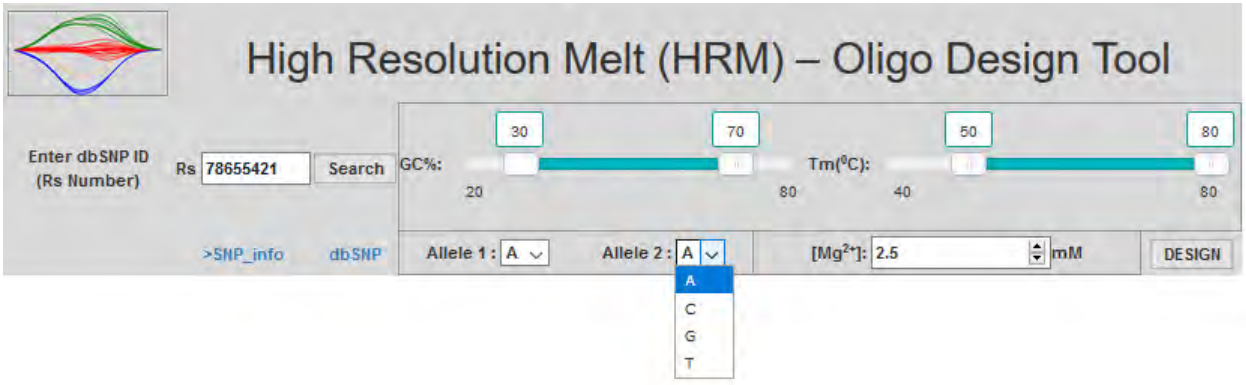

Figure 4.11: Tool provide all available alleles in both Allele 1 and Allele 2 dropdowns to let the user to select the appropriate alleles

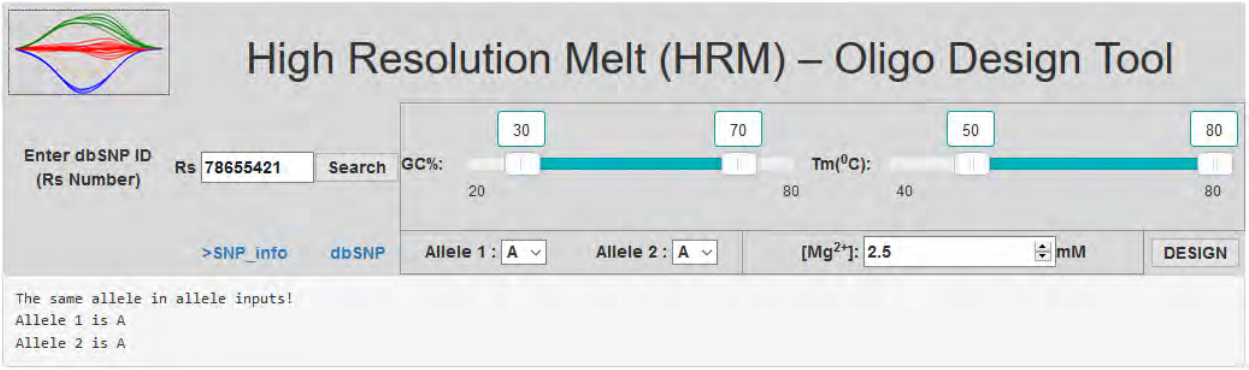

Figure 4.12: Error message provided by the tool in case the user forget to select the appropriate alleles for dbSNP IDs with multiplicity

3. db SNP ID : Rs6166: This SNP doesn't have multiplicity or doesn't have more than two genotypes so that the both values (allele 1 & 2) will be automatically selected

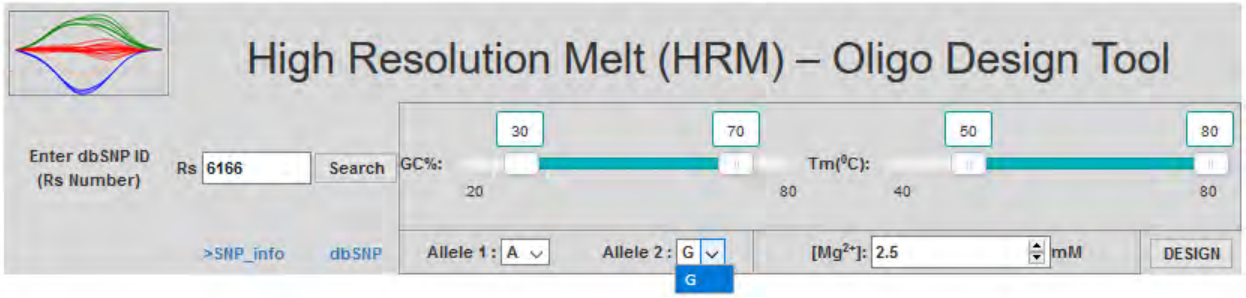

Figure 4.13: Tool select the appropriate alleles in the absence of multiplicity

CHAPTER 5: TESTING AND EVALUATION

5.1 INTRODUCTION

Software testing is a decisive component of software of quality assurance that represents the ultimate review of specifications, design, and code propagation. The testing is comprised of Verification and Validation methods. Verification - refers to the techniques of finding out whether the built software meets all the customer requirements and Validation - Refers to the techniques that ensure the software meets all the specified functional requirements. The testing procedure is considered dynamic; given that the operation is carried out on an executable part of the organization. Testing is a vital process to ensure a perfectly working system with fewer errors is delivered to the customer.

5.2 Testing Strategies

Unit Test

This type of testing was carried out for each and every atomic module or unit in the system. In that respect are some definite criteria for Quality aspects; these measures may vary project to visualize. Agreeing to our standards we prepared a unit test plan for each and every module. Functionality of each module is different from the other. So functionality test part of this design should depend on modules. Every feature of a single unit is tested using **test cases**. A test case will consist of three main components namely Test data (Inputs to the test case), Expected outcome and Objective of testing. Unit testing for each individual module / increment was performed during the course of tool development using manual testing as described earlier in 4.6.

System Test

To test all modules a proper System test plan to confirm functionality of modules was designed. When there is an issue found in test, categorized that issue and prioritized those into low, medium or high base on the sense of urgency to resolve that issue. Corrections are done according this priority.

Integration Test

In this project a combination of top-down and bottom-up testing would be utilized. Later on the integration of modules have been completed and tested, we proceeded on to testing of the entire course of study, in which validation is tested. This is, in effect, testing of functional expectations, i.e. that the software functions in the manner that the imperial reasonably expects it to. Few of the components that can be looked into are Modularity (functional interdependence of the program), Ease of operation, Security, Simplicity, Accuracy of data, Efficiency and Error tolerance.

Integration testing and unit testing were performed together when improving the system in small increments while developing the tool. At the initial stages the software called ZAPTEST Free edition was used to test the tool modules separately using its record and playback ability and also by the simple scripts. This tool was used to identify specific elements in the webpage and to go through them using the simple drag and drop function and instructions were given using the simple VB type scripts.

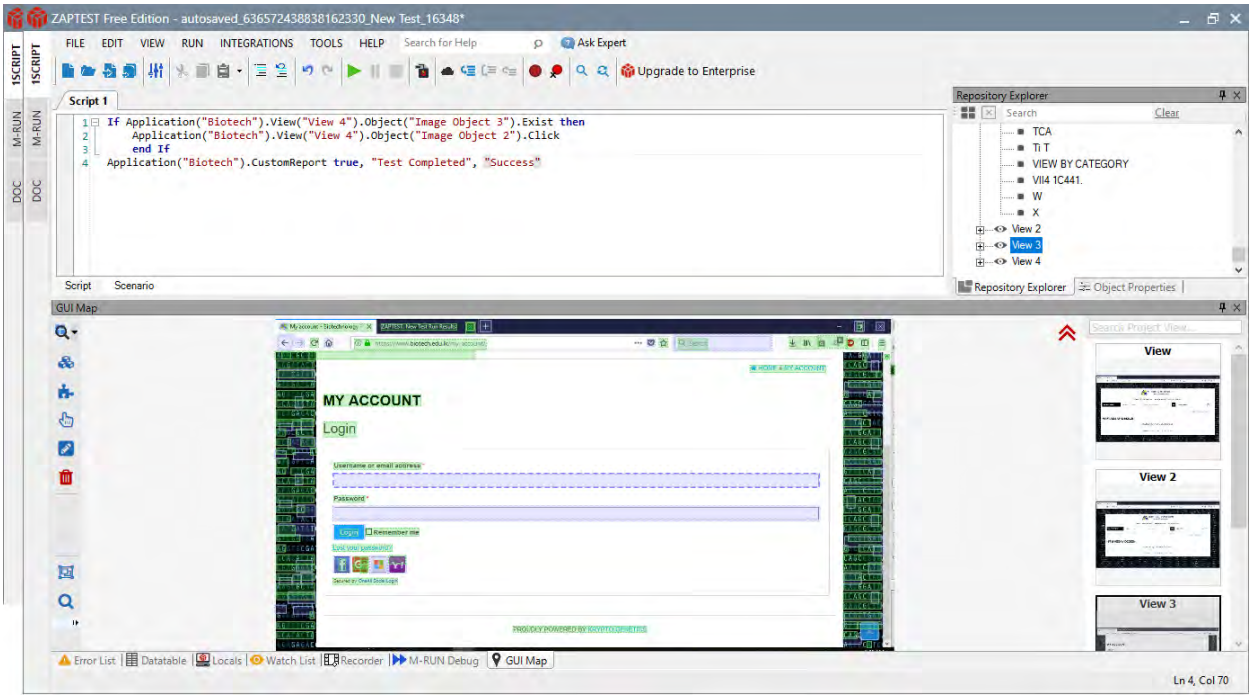

Figure 5.1: ZAPTEST UI for specific predefined basic functions

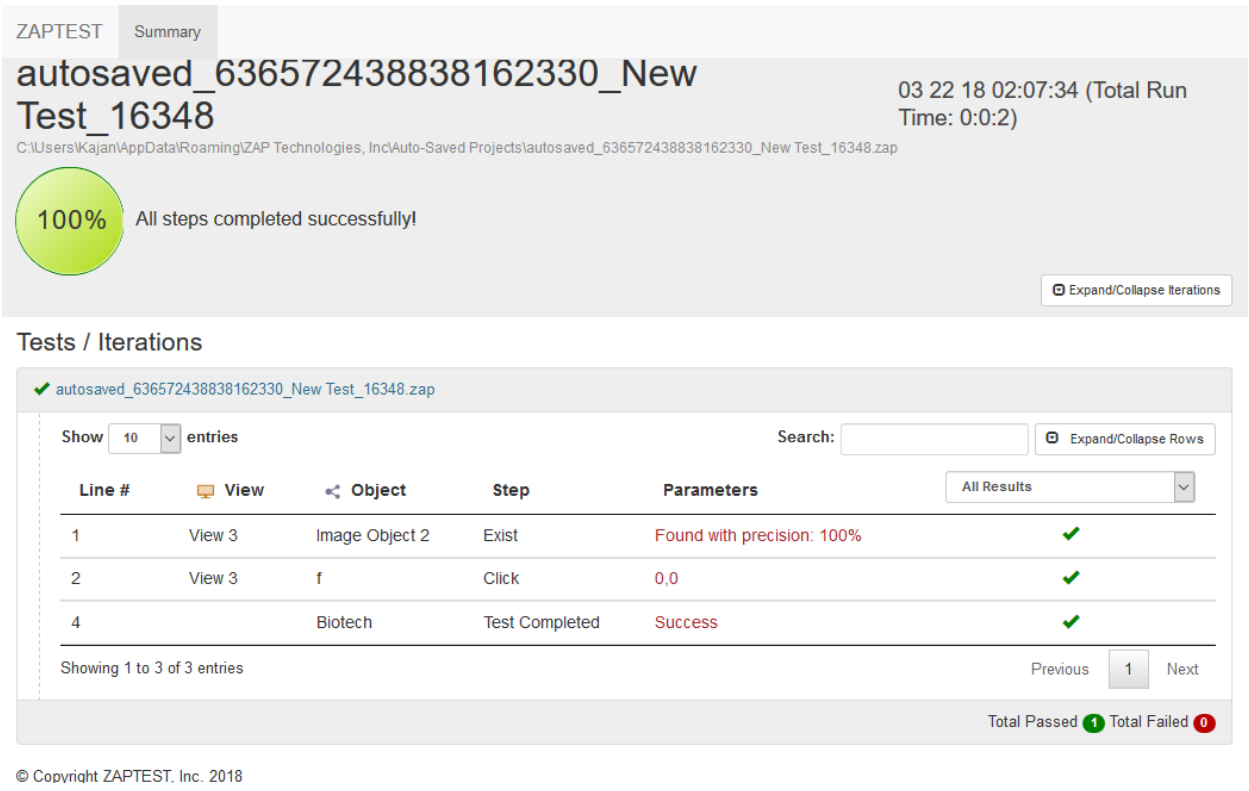

Figure 5.2: ZAPTEST results obtained for the testing shown in Figure 5.1

5.3 Test Plan & Test Cases

A Test plan, which covers all aspects of testing was used as a guide for the entire testing process. The test plan was designed before the implementation of the tool aiming at continuous testing during the development and implementation of the tool to maximize the efficiency, reduce the number of errors and to prevent the possible formation of hidden icebergs. A test plan includes: Test objectives, assumptions, test principles, user acceptance testing (UAT), Test Management and especially the test cases.

1. **Test Objectives**

   The objective of the test is to verify that the functionality of High Resolution Melt (HRM) – Oligo Design Tool v1.0 –works according to the specifications. The test will execute and verify the test scripts, identify, fix and retest the tool with various different type of Polymorphisms (SNPs) as per the entrance criteria.

The final product of the test includes:

a) A production-ready software;

b) A set of stable test scripts that can be reused for Functional and test execution.

2. Test Assumptions

Key Assumptions:

a) The End-user will be always a Molecular Biologist who have a basic knowledge on PCR conditions

b) The end-user must aware that the original Melting temperature and calculated melting temperature will be varying since the calculation was made based on few assumptions and that was just an indication only

c) User should be able to find the appropriate dbSNP id from literature or from the web before processing into the tool

d) Tool won't guarantee the results when the GC content was below 40% and above 60% when the user don't have any knowledge on PCR optimization

General Assumptions:

a) If the data for a dbSNP id was not available on NCBI database the tool getting the data from the other sources was not a functionality of the tool developed

b) Since there are more than 2 alleles for some SNPs the system only load the available alleles for the user to select even when only 2 alleles available system need that user confirmation to proceed

c) System was loaded with default values for other fields always and can be modified by the user when it's necessary

d) Tool always display the DNA sequence in which the orientation NCBI server provides the data

3. Test Principles

- Testing will be focused on meeting the business objectives, cost efficiency, and quality
- There will be common, consistent procedures

- Testing activities will build upon previous stages to avoid duplication of effort
- Testing will be a repeatable, quantifiable, and measurable activity

4. **User Acceptance Test (UAT)**

   PURPOSE: this test focuses on validating the HRM primer design logic. It allows the end users to complete one final review of the tool prior to deployment by testing the sequence generated using wet-lab practice. TESTERS: the UAT is performed by the end users.

   The user evaluation questionnaire was designed for 2 induvial tools and the whole system and sent to the users via email in PDF forms to fill them and return them to the email from which the form was sent. The sample of this questionnaire was annexed in annexure II.

5. **Test Management**

   In addition to the manual testing preformed based on acquired knowledge of different types to possible SNPs. It was also testing using automated scripts / test automation tools to test array of different SNPs from the last one year data from the University of Colombo scientific literature.

6. **Test cases**

   Test cases were created once the test plan was designed. That consists of data, procedure, and expected result, and represents just one situation under which the system or part of the system might run. Test cases were designed for each modules and functions separately to reduce complexity of the testing process. The tabulation of designed specific test cases together with test objectives, pre-conditions, steps involved, test data and expected results were given in the Annexure I. Those conditions were tested and verified manually.

5.6 TEST AUTOMATION

The automated testing for this system was performed using Katalon Studio v5.3.1 Build 1. This test comprises steps starting from opening the browser, entering the URL, click the desired tool from the system with appropriate credentials if required, providing the dbSNP ID, designing the

assay and up to closing the browser. This testing for combined functionalities can be designed simply by recording the actions performed and playback them while changing the values recorded in the testing tool. When the right values given success message is expected and also the wrong values were provided to the test tool in this testing and failure was expected to confirm the functionality of the system. This type of testing was performed repeatedly with various dbSNP IDs with different combinations of credentials to ensure the better system functionality.

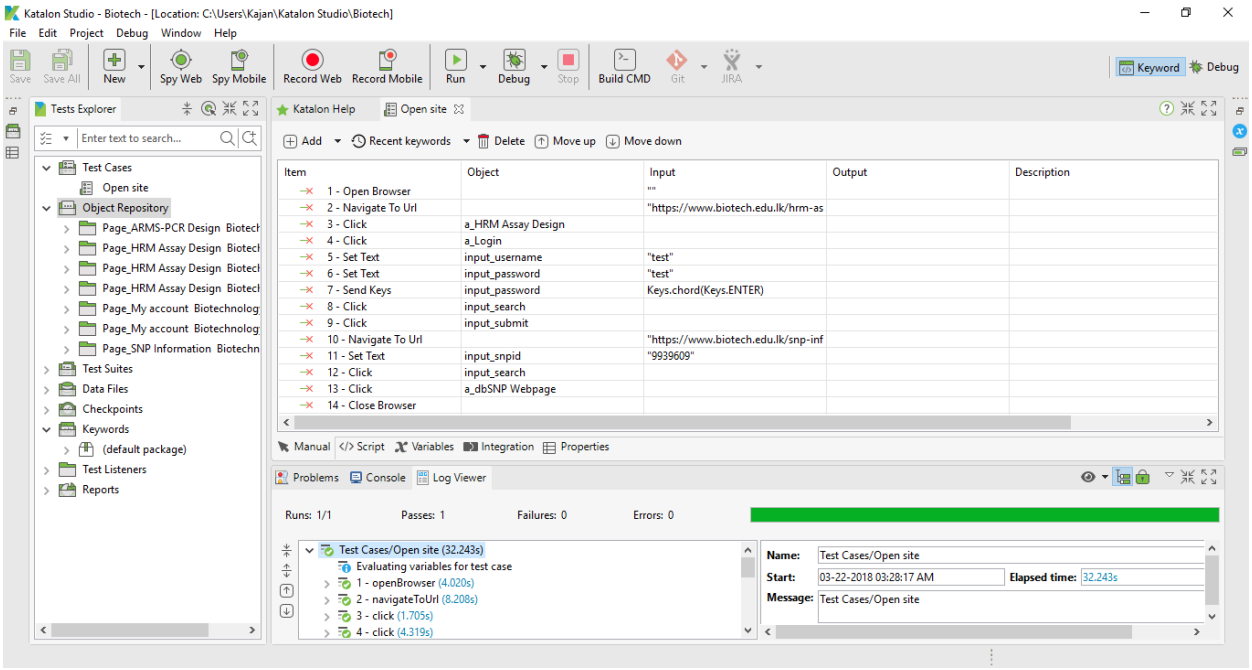

Figure 5.3: Katalon Studio UI for automated testing

CONCLUSION AND FUTURE WORK

Though the project was intended to design a tool to design HRM based assays the code components itself can be used with minimal changes to design two other tools for obtaining the information about the specific SNP using the dbSNP id and ARMS-PCR tool enable user to design their end-point PCR based SNP genotyping assay with just 2 clicks instead of visiting the dbSNP server in NCBI and copying the DNA sequence (max 1000bp) to the Primer1 tool in soton.ac.uk server and again refer the dbSNP to fill the allele and allele position in the tool form. This process takes lots of time to design an assay and it's prone to human errors and involves lot of manual work. ARMS-PCR design tool simplify this designing process and reduce the human errors.

HRM Assay design tool was made on a novel approach to design primers with shorter length to work compatibly with low end qPCR instrumentation where the uDesign server gives only the snapback primers which are expensive in terms of primer length to be synthesized chemically. This approach was came as a result of understand the nearest-neighbor method based thermodynamics for the oligonucleotides. Combining this feature with providing more details about the assay ensure better user understanding about the assay design and color codes further enhance the user acceptance.

Since the PHP was a server site script and when it was integrated to the WordPress system the security of the code was further increased and it make the developer to identify the real users of the tool in the meantime to get in touch with them to get the feedbacks and enable further enhancement of the tool.

This work will be presented in the South Asian Biotechnology Conference 2018 on 30th March 2018 as a poster presentation. The future work on this tool include further enhancement of the melting temperature gap between differently melting amplicons to eliminate the uncertainty in case of regions with low (<40%) and high (>60%) GC content and addition of few other tools will be made on www.biotech.edu.lk server.

BIBLIOGRAPHY

[1] "What is DNA? - Genetics Home Reference - NIH," U.S. National Library of Medicine, 26 06 2018. [Online]. Available: https://ghr.nlm.nih.gov/primer/basics/dna. [Accessed 27 06 2018].

[2] wikipedia, "Nucleic Acid structure," [Online]. Available: https://en.wikipedia.org/wiki/Nucleic_acid_structure. [Accessed 27 06 2018].

[3] Nature Education, "Translation: DNA to mRNA to Protein," Nature, 2008. [Online]. Available: https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393. [Accessed 27 06 2018].

[4] NCBI, "dbSNP homepage," [Online]. Available: https://www.ncbi.nlm.nih.gov/projects/SNP/. [Accessed 27 06 2018].

[5] Wikipedia, "Nucleic acid thermodynamics," [Online]. Available: https://en.wikipedia.org/wiki/Nucleic_acid_thermodynamics. [Accessed 27 06 2018].

[6] "Polymerase chain reaction," Khan Academy, [Online]. Available: https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-sequencing-pcr-electrophoresis/a/polymerase-chain-reaction-pcr. [Accessed 27 06 2018].

[7] "Polymerase Chain Reaction," [Online]. Available: http://en.wikipedia.org/wiki/Polymerase chain reaction.

[8] R. J. Sarah Maddocks, "Quantitative PCR: Things to Consider," in *Understanding PCR : A Practical Bench-Top Guide*, Academic Press, 2017, pp. 45-52.

[9] L. Kristensen, G. Andersen, H. Hager and L. Hansen, "Competitive amplification of differentially melting amplicons (CADMA) enables sensitive and direct detection of all mutation types by high-resolution melting analysis.," *Hum Mutat.*, vol. 33(1), pp. 264-71, 2012.

[10] J. S. Waye and E. Barry, "Allele-Specific Mutation Detection by," in *Molecular Diagnostics*, Burlington, Elsevier Academic Press, 2005, pp. 15-22.

[11] "SNP Genotyping Analysis Using TaqMan assays," Thermo Fisher, [Online]. Available: https://www.thermofisher.com/lk/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/snp-genotyping-taqman-assays.html. [Accessed 09 08 2017].

[12] D. Shammaa, A. Bazarbachi, H. Halas, L. Greige and R. Mahfouz, "JAK2 V617F mutation detection: laboratory comparison of two kits using RFLP and qPCR.," *Genet Test Mol Biomarkers,* vol. 14, no. 1, pp. 13-15, 2010 Feb.

[13] K. Borecki, I. Zawada, A. Pawinska-Matecka, N. Salkic, B. Karakiewicz and G. Adler, "ABCB1 3435C>T and 2677G>T/A polymorphisms in Polish and Bosnian Crohn's disease patients - A preliminary report.," *Bosn J Basic Med Sci*, 2017.

[14] "Mutation scanning," 09 08 2017. [Online]. Available: https://www.thermofisher.com/lk/en/home/life-science/pcr/real-time-pcr/real-time-pcr-applications/genetic-variation-analysis-using-real-time/high-resolution-melting-hrm/mutation-scanning.html.

[15] Jeremy, M., Berg, John, L, Tymocko, S. and and Lubert, Biochemistry, New York: W. H. Freeman and Company, 2002.

[16] T. Knight and A. Wozniak, Synthetic biology urop. notes from conversations, 2004 April.

[17] [Online]. Available: http://www.kazusa.or.jp/codon/.. [Accessed 1 08 2017].

[18] R. Baker, "Protein Production: A Simple Summary of Transcription and Translation," owlcation, 18 01 2018. [Online]. Available: https://owlcation.com/stem/protein-production-a-step-by-step-illustrated-guide. [Accessed 02 07 2018].

[19] R. C. E. T. M. R. Kathryn M.Camp MS, "Position of the Academy of Nutrition and Dietetics: Nutritional Genomics," *Journal of the Academy of Nutrition and Dietetics,* vol. 114, no. 2, pp. 299-312, 2014.

[20] J. SantaLucia and Jr, "A unified view of polymer, dumbbell, and oligonucleotide," in *Proceedings of the National Academy of Sciences of the United States of America*, February 1998 95(4):1460–1465.

[21] Cole, J. R., B. Chai, R. J. Farris, Q. Wang, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, A. M. Bandella, E. Cardenas, G. M. Garrity and M. J. and Tiedje, "The ribosomal database project (rdp-ii): introducing myrdp space and quality controlled public data.," *Nucleic Acids Res,* vol. 35, no. Database issue, pp. D169-172, 2007.

[22] "Human Genome Project Information," [Online]. Available: http://www.ornl.gov/sci/techresources/Human Genome/home.shtml. [Accessed 31 07 2017].

[23] K. B. Mullis, "The unusual origin of the polymerase chain reaction," *Scientific American,* vol. 262, pp. 56-63, 1990.

[24] "IUPAC code," [Online]. Available: http://www.bioinformatics.org/sms2/iupac.html. [Accessed 31 07 2017].

[25] R. Higuchi, C. Fockler, G. Dollinger and R. Watson, "Kinetic PCR: Real time monitoring of DNA amplification reactions," *Biotechnology,* vol. 11, pp. 1026-1030, 1993.

[26] R. E. Thompson, "AN ANALYSIS OF EFFICIENCY AND MELT CURVE EFFECTS ON QUANTITATIVE POLYMERASE CHAIN REACTION (QPCR) INHIBITION," FIU Digital Commons, Miami, Florida, 2010.

[27] B. SA., "Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction," *J. Mol. Endocrinol,* vol. 25, pp. 169-193, 2000.

[28] B. JM., "Forensic DNA Typing," *Second Edition: Biology, Technology, and Genetics of STR Markers,* vol. Elsevier Academic Press, 2005.

[29] C. E., "Structure and function of nucleic acids as cell constituents," *Fed. Proc.,* vol. 10, pp. 654-659, 1951.

[30] K. Ririe, R. Rasmussen and C. Wittwer, "Product differentiation by analysis of DNA melting curves during the polymerase chain reaction," *Anal Biochem,* vol. 245(2), pp. 154-60, 1997.

[31] "High Resolution Melting (HRM)," Thermo Fisher, [Online]. Available: https://www.thermofisher.com/lk/en/home/life-science/pcr/real-time-pcr/real-time-pcr-applications/genetic-variation-analysis-using-real-time/high-resolution-melting-hrm.html. [Accessed 31 07 2017].

[32] T. E. K. H. H. a. L. L. H. Lasse Sommer Kristensen, "Competitive amplification of differentially melting amplicons (CADMA) improves KRAS hotspot mutation testing in colorectal cancer," *BMC Cancer,* vol. 12, no. 548, 2012.

[33] C. W. Dieffenbach, T. M. J. Lowe and G. S. and Dveksler, "PCR PRIMER A LABORATORY MANUAL," *Cold Spring Harbor Laboratory Press,* 1995.

[34] "IUPAC Code," [Online]. Available: http://www.bioinformatics.org/sms2/iupac.html. [Accessed 31 07 2017].

[35] "uDesign," University of Utah, [Online]. Available: https://www.dna.utah.edu/udesign/app.php. [Accessed 31 07 2017].

[36] "Primer Design," [Online]. Available: http://bioweb.uwlax.edu/genweb/molecular/seq anal/primer design/primer design.htm. [Accessed 31 07 2017].

[37] C. W. Dieffenbach, T. M. Lowe and G. S. and Dveksler, "General concepts for pcr primer design," *Genome Res.,* vol. 3, pp. s30-37, 1993.

[38] T. Borgbo, L. Sommer Kristensen, I. Lindgren, C. Yding Andersen and L. L. & Hansen, "Genotyping common FSHR polymorphisms based on competitive amplification of differentially melting amplicons (CADMA)," *Journal of Assisted Reproduction and Genetics,* vol. 31, no. 11, pp. 1427-36, 2014.

Annex I

Test Cases

ID	Test Objective	Precondition	Steps	Test data	Expected result	Actual result	Status
1	Verify the starting of the tool	Need a server compatible with php 5.6.31 or later and browser which can render CSS and execute JavaScript	Type the URL on the browser address bar and press enter	Snp.kryptogenetics.com.lk	Browser loads the tool homepage with tabs	Success	✓
2	Tool verify the SNP id	User opened the HRM designed tool tab from the above said URL	Type alpha numeric and special characters	Incorrect dbSNP id or test with mixed characters (alpha numeric and special characters)	Display error message below the textbox	Success	✓
3	Verify the server communicating with dbSNP (in NCBI server)	NCBI server and dbSNP should be live and the server should be able load xml pages. User entered a dbSNP id	Enter a dbSNP without Rs prefix and press search button	Valid dbSNP id	Appropriate alleles are loaded in the menu	Success	✓
4	SNP multiplicity	Entered a dbSNP ID	Check the dropdown menu loaded	dbSNP id with multiple alleles	Number of options in the dropdown exceed 2	Success	✓
5	SNP indels	Entered a dbSNP ID	Check the error	dbSNP id for insertions and deletions	Error message	Success	✓
6	Combined input sent to soton.ac.uk server	Entered the dbSNP id	PCR conditions with alleles provided and pressed design	Appropriate / same alleles	If conditions match right primer sets with various parameters else error message	Success	✓
7	Allele 1 and 2 selected correctly	Entered the dbSNP id	Apply same alleles in both inputs	Same alleles	Error message “Allele 1 : X Allele 2 : X “	Success	✓
8	GC content correctness	Entered the dbSNP id	GC content adjusted	Adjusted to 50-50 from 30-70	Error message with request to adjust the parameters	Success	✓

9	Tm correctness	Entered the dbSNP id	Tm adjusted	Inappropriate Tm	Error message with request to adjust the parameters	Success	✓
10	Product GC content	Entered the dbSNP id and alleles correctly and IDT oligo analyzer tool	Check the GC content	dbSNP id and appropriate alleles	Same GC content from oligo analyzer	Success	✓
11	Primer orientation and sequence	Entered the dbSNP id and alleles correctly	Click the dbSNP link and verify the orientation and primer sequence manually	dbSNP id and appropriate alleles	Primer aligns with the dbSNP sequence	Success	✓
12	System picked the appropriate orientation	Entered the dbSNP id and alleles correctly and IDT oligo analyzer tool	Copy the Target sequence and check the GC content of both orientations manually	dbSNP id and appropriate alleles	Allele specific primers designed in the orientation where GC content close to 50%	Success	✓
13	Added Tm increasing mutation to Tm high allele	Entered the dbSNP id and alleles correctly and IDT oligo analyzer tool	Check the Tm of the Allele specific primers manually without first 3 letters	dbSNP id and appropriate alleles	Even without last 3 nucleotides the Tm of the 1st primer should be lesser than the 2nd primer	Success	✓
14	Appropriate color codes used	Entered the dbSNP id and alleles correctly and IDT oligo analyzer tool	Check the Tm of green and blue color coded primers	dbSNP id and appropriate alleles	Green – High Tm Blue – low Tm	Success	✓
15	Gene, Chromosome and SNP type	Entered the dbSNP id and alleles correctly and NCBI dbSNP database	verify the Gene, Chromosome and SNP type manually by opening dbSNP link in new tab	dbSNP id and appropriate alleles	Gene, Chromosome and SNP type loaded correctly	Success	✓

SNP-Genotyping Assay Designer
User Evaluation Form

General information

Name:

1. Educational Level:					
	Undergrad	Graduate	Postgrad	PhD.	Post-Doc
2. Computer Literacy Level:					
	Poor	Fair	Average	Good	Excellent

SNP info & ARMS-PCR design tool

3. Starting this tool was					
	V. Difficult	Difficult	Average	Easy	V. Easy
4. Using this tool was					
	V. Difficult	Difficult	Average	Easy	V. Easy
5. Resources needed for using this tool (S/W & H/W)					
	V. Low	Low	Average	High	V. High
6. Level of information needed by the tool					
	V. Low	Low	Average	High	V. High
7. Level of data provided by the system to the user					
	Poor	Fair	Average	Good	Excellent
8. Usability of the tool					
	Poor	Fair	Average	Good	Excellent
9. User Satisfaction					
	V. Low	Low	Average	High	V. High

10. Additional Comments for improvements

HRM Assay Design tool

11. Starting this tool was					
	V. Difficult	Difficult	Average	Easy	V. Easy
12. Using this tool was					
	V. Difficult	Difficult	Average	Easy	V. Easy

13. Resources needed for using this tool (S/W & H/W)	V. Low	Low	Average	High	V. High
14. Level of information needed by the tool	V. Low	Low	Average	High	V. High
15. Level of data provided by the system to the user	Poor	Fair	Average	Good	Excellent
16. Usability of the tool	Poor	Fair	Average	Good	Excellent
17. User Satisfaction	V. Low	Low	Average	High	V. High

18. Additional Comments for improvements

Overall Impression

19. Easy to understand which Page I am currently on	Strongly agree	agree	Neutral	Disagree	Strongly disagree
20. Navigation to other pages In this tool was easy	Strongly agree	agree	Neutral	Disagree	Strongly disagree
21. Use of colors gives better idea about the assay design	Strongly agree	agree	Neutral	Disagree	Strongly disagree
22. It was attractive and I will Use this tool again in future	Strongly agree	agree	Neutral	Disagree	Strongly disagree

Thank you for participating in this User Evaluation Survey

Visit us: https://snp.biotech.edu.lk

Annex III

User Evaluation Summary

Descriptive Statistics					
	Total Participants	Minimum	Maximum	Mean	Values
Edu_level	3	1.00	3.00	2.0000	1- Undergrad; 2-Graduate; 3-Postgrad; 4-PhD and 5-Post-Doc
Com_literacy	3	3.00	4.00	3.3333	1-V.low to 5-V.High
AQ3	3	3.00	5.00	4.0000	Questions on ARMS-PCR Deisgn tool 1-Least favorable answer to 5-Most favorable answer
AQ4	3	4.00	5.00	4.3333	
AQ5	3	4.00	4.00	4.0000	
AQ6	3	4.00	4.00	4.0000	
AQ7	3	4.00	5.00	4.3333	
AQ8	3	4.00	5.00	4.3333	
AQ9	3	4.00	5.00	4.6667	
HQ11	3	3.00	5.00	4.0000	Questions on HRM design tool 1-Least favorable answer to 5-Most favorable answer
HQ12	3	3.00	5.00	4.0000	
HQ13	3	3.00	5.00	4.0000	
HQ14	3	4.00	5.00	4.3333	
HQ15	3	3.00	5.00	4.0000	
HQ16	3	3.00	5.00	4.0000	
HQ17	3	3.00	5.00	4.0000	
Q19	3	4.00	5.00	4.3333	Questions on the complete system 1-Least favorable answer to 5-Most favorable answer
Q20	3	4.00	5.00	4.3333	
Q21	3	4.00	5.00	4.6667	
Q22	3	4.00	5.00	4.6667	
Valid N (listwise)	3				

SNP-Genotyping Assay Designer: For Cost Effective and Time Efficient SNP genotyping Assays

M Kajan [acd], V A de Silva[*a], B Umayal[b], C R Wijesinghe[d] and N V Chandrasekharan[c]

*[a]Department of Psychiatry and [b]Department of Obstetrics and Gynecology Faculty of Medicine, University of Colombo; [c]Department of Chemistry, Faculty of Science, University of Colombo; [d]University of Colombo School of Computing, Colombo, Sri Lanka (kajan.muneeswaran@gmail,com, *varunidesilva2@yahoo.co.uk)*

The SNP-Genotyping Assay Designer is a set of tools used to design tetra primer amplification refractory mutation system – Polymerase chain reaction (ARMS-PCR) based & High-resolution melt (HRM) based Single Nucleotide Polymorphism (SNP) genotyping assays. They provide a more convenient and easier to use technique to design an assay for both end-point PCR and qPCR (which lacks high resolution melting). The tool takes the SNP ID from the user input and it will then provide the all relevant information on the particular SNP from dbSNP database *via* the NCBI E-utils which allows the user to design an assay with just 2 clicks.

The methodology for ARMS-PCR primer design include retrieval of SNP position, flanking sequence & alleles from dbSNP database followed by manual user input into the primer1 tool available at soton.ac.uk server (http://primer1.soton.ac.uk/primer1.html). Automation of this process into a single tool facilitates faster assay design and eliminates potential human errors. Competitive amplification of differentially melting amplicons (CADMA) is a qPCR HRM based SNP genotyping technique which incorporates additional mutations to the 5' region of each allele specific PCR primer to increase Melting temperature (Tm) differences between allele specific PCR products to enable identification even with low resolution qPCR.

Penalties for incorporating additional mutations using CADMA can give non-specific amplifications in some cases. Therefore, pre-identified 3bp patterns which increase the Tm were added to the 5' tail of primers together with a specificity increasing strong mismatch in penultimate position. Nearest neighbor Tm calculation formula with salt adjustment calculations were adopted in this tool, which is freely accessible on snp.biotech.edu.lk.