The SNP CYP2C19*2 (rs4244285) GENOTYPING USING TETRA-PRIMER ARMS PCR METHOD

Submitted by:

Afia Adiba Rupoma

ID: 16136031

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of
Bachelor of Science in Biotechnology.

Department of Mathematics and Natural Sciences

Brac University

August 2020.

Declaration

I hereby declare that the thesis project titled "**The SNPCYP2C19*2 (rs4244285) GENOTYPING USING TETRA-PRIMER ARMS PCR METHOD**" has been written and submitted by me, Afia Adiba Rupoma and has been carried out under the supervision of **Md. Mahmudul Hasan Akash**, lecturer, Department of Genetic Engineering and Biotechnology, Dhaka University, and Professor **Dr.Mahboob Hossain**, Coordinator of Microbiology Program of MNS Department, BRAC University, Dhaka.

It is further declared that this thesis has been composed solely by me and it has not been submitted, in whole or in part, in any previous institution for a degree or diploma. All explanations that have been adopted literally or analogously are marked as such.

Afia Adiba Rupoma

(Afia Adiba Rupoma)

Student

Approval

The thesis/project titled **The SNP CYP2C19*2 (rs4244285) GENOTYPING USING TETRA-PRIMER ARMS PCR METHOD** submitted by

1. Afia Adiba Rupoma (ID: 16136031)

of Spring, 2016 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Biotechnology on 26th of August 2020.

Examining Committee:

Supervisor:
(Member)

Mahboob Hossain

_____

M. Mahboob Hossain, PhD
Coordinator of Microbiology Program of MNS Department,
Brac University

Co-Supervisor:
(Member)

_____

Md. Mahmudul Hasan Akash, lecturer,
Department of Genetic Engineering and Biotechnology,
Dhaka University

Program Coordinator:
(Member)

_____

Iftekhar Bin Naser, PhD
Coordinator of Biotechnology Program,Department of Mathematics and Natural Sciences
Brac University

Departmental Head:
(Chair)

_____

A F M Yusuf Haider, PhD
Professor and Chairperson, Department of Mathematics and Natural Sciences
Brac University

Acknowledgment

First and foremost, I would like to express my thanks to almighty Allah because He has given me the opportunity and strength to finish this research. I am also thankful for His blessings to my daily life, good health and a healthy mind.

I acknowledge my esteem to Professor **A F M Yusuf Haider**, Chairperson of MNS Department and Assistant Professor **Iftekhar Bin Naser**, Coordinator of The Biotechnology Program of MNS Department of BRAC University for allowing me and encouraging me to complete my undergraduate thesis.

My regards, gratitude, indebtedness and appreciation goes to my respected supervisors **MD. Mahmudul Hasan Akash**, Lecturer of Department of Genetic Engineering and Biotechnology, Dhaka University and Professor **Dr. M.Mahboob Hossain**, Coordinator of Microbiology Program of MNS Department, BRAC University for their constant supervision, constructive criticism, expert guidance, enthusiastic encouragement to pursue new ideas and never-ending inspiration throughout the entire period of my research work. I would like to thank and express my deepest gratitude for guiding me in my report writing and providing time to time suggestions regarding the setting of my experimental designs, interpretation of results and subsequent directions for the whole work without being a bit of impatient. It would have been impossible to submit my report without their cordial help.

I would like to extend my appreciation to the respective Lab officers **Asma Binte Afzal** and **Md. Nazrul Islam** for their suggestions and moral support during my work.

I also appreciate my thesis partners **Wasifa Ar Rahman**, **Anika Tahsin**, **Rifaquat Ahmed** for their kind cooperation and active support throughout my work.

Finally, I would like to extend my gratitude to the members of my family and friends for their prayerful concerns and supports.

Afia Adiba Rupoma.

Abstract

Drugs are used to treat and cure disease. But the same drug with the same dose might not be useful for other individuals. Through Pharmacogenetics study, how people respond differently to drug therapy based upon their genetic makeup or genes can be determined. The latest research eliminates the trial-and-error approach to medication options and therefore restricts patients' access to medications that are either not effective or harmful to them. Single Nucleotide Polymorphisms (SNPs) holds the key for identifying the probability of an individual's vulnerability to various diseases and drug reaction. Most SNPs are functionally silent, occurring in non-coding or non-regulatory genome regions. Some of the SNPs cause altered protein structure or expression, however. These biologically functional SNPs in both health and disease are considered the essence and substrate of human diversity.

A clinically significant SNP CYP2C19*2 (rs4244285) was selected for this study based on its heterozygosity. This SNP is a poor metabolizer of the drug Clopidogrel. Tetra ARMS PCR primer set was designed to detect the allelic condition of this SNP in Bangladeshi individuals. These primer sets can detect both wild type and mutant alleles. Buccal samples of healthy volunteers were collected using 0.9% NaCl in distilled water. DNA from these samples was extracted using The Phenol Chloroform method. PCR reaction was optimized by trial and error. The first PCR reaction gave clear bands at the allocated area. From the PCR reaction and afterward gel electrophoresis it was found that one sample was heterozygous mutant (G/A) and two samples were homozygous wild type (G/G) for this SNP.

This SNP genotyping protocol is effective and less expensive. As this experiment is conducted over Bangladeshi individuals, the measurement of the occurrence of this allele will help to evaluate the medication dosage to prevent side effects or overdose of the medication in the Bangladeshi population.

Table of Content

List of Figures

List of Tables

1 INTRODUCTION

1.1 Role of SNP in Pharmacogenetics

1.1.1 About Pharmacogenetics

Pharmacogenetics discusses the impact variations of the medications caused by genetic variability. Genetic metabolism variability can lead to high concentrations of drugs and an increased risk of adverse effects in slow metabolizers (Dervieux& Bala, 2006). The single nucleotide polymorphism, or SNP, is the most important cause for genetic variability and hence the foundation for a pharmacogenetics approach to drug therapy.

Pharmacogenetics addresses the role of single nucleotide polymorphisms (SNPs) in genes encoding drug-metabolizing enzymes, drug receptors, drug carriers and other proteins involved in disease pathogenesis in deciding the response of a patient to a medication. Such polymorphisms in effect change the protein product's functionality and contribute to significant phenotypic variations in therapeutic reaction or susceptibility to carcinogenesis (Alwi, 2005). Genetic factors are thought to represent important determinants of drug efficacy and toxicity. The same doses of medication can cause adverse drug effects and toxicity across populations. That is why, Pharmacogenetics study is beneficial in an effort to individualize treatment in the pursuit of answers to the biological basis of individual variations in drug response (Maitland *et al.*, 2003).

1.2 Single Nucleotide Polymorphism

The most common form of genetic variation in humans is single nucleotide polymorphisms, also called SNP. Polymorphisms are what make individuals different from one another. DNA sequence consists of a series of 4 nucleotide bases: A, C, G, and T. Each SNP represents a difference in a single nucleotide base. Not only every base disparity between two people is an SNP; a deviation is considered a polymorphism if it happens within 1 % or more of the population (Karki *et al.*, 2015).When polymorphism is

stable, it's an SNP. When there is an SNP inside a gene, the gene is defined as having more than one allele. In such cases, amino acid sequence variations can occur with SNPs. However, SNPs are not just linked to genes; they can also occur in non-coding DNA regions. A SNP site at which the base pair change leads to an amino acid substitution, or otherwise affects a protein, is a coding SNP (cSNP). There are tens of thousands of cSNPs, at least one in almost every gene(Pinglang *et al.*, 2006).

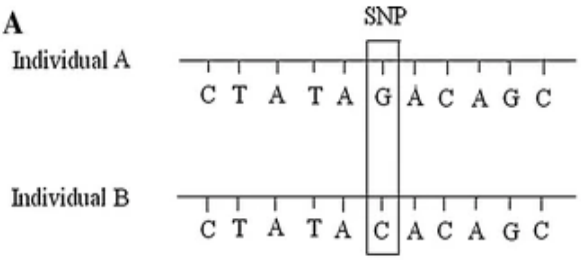

Figure 1.1: A schematic representation of a single nucleotide polymorphism (SNP). Individual A and individual B contain similar genetic information, but at a specific location of the gene, there is a single base-pair change creating the polymorphism.

1.3 Importance of SNP

The importance of SNPs comes from their ability to influence disease risk, drug efficacy and side-effects. SNPs are probably the most important category of genetic changes influencing common diseases. In phenotypic expression single nucleotide polymorphisms have been stated to be of great importance. An SNP within a gene's coding region may alter the sequence of amino acids and thus affect the role of proteins. These alterations in gene control regions can influence a gene's expression like the presence of SNP in the promoter region can affect gene expression patterns SNPs can be used to understand the molecular mechanisms of sequence evolution (Asthana & Sunyaev, 1970) furthermore, SNPs can track the inheritance of disease genes within families, SNPs can be used for distinguishing people in forensic cases, and most importantly, SNPs have a role in drug metabolism and toxicity. The features of the SNPs make them more useful than other polymorphic markers for genetic pleiotropy studies in diverse traits and disorders, and population-based gene identification. SNPs play a significant part in biomedical science.

SNPs have been used as high-resolution markers of disease-related gene maps or natural traits of genome-wide association studies (GWAS) (Ballesta *et al.*, 2020).

1.4 Importance of SNP in Pharmacogenetics and Pharmacogenomics studies

Drug activities are determined by the interplay of multiple genes encoding various proteins involved in specific biochemical pathways. With rapidly evolving technical tools for SNP discovery and extensive understanding of the function of SNPs in disease susceptibility and drug response variability, the pharmacogenetics approach to therapy is expected to take off in the not too distant future. SNP biomarkers are potentially important for wide-ranging clinical applications of Personalized Medicine, including risk assessment of disease predisposition, disease screening, diagnosis, prognosis, selection of medications, monitoring of drug response and assessment of drug resistance. There is strong evidence established that links SNPs to inter-individual differences in drug response. Patients with more active drug-metabolizing enzymes that require increased doses of the drug and those without an active enzyme may show toxicity.

For example, Secondary leukemia can occur at a later age when epipodophyllotoxins (inhibits the enzyme topoisomerase II and block DNA replication in dividing tumor cells) are used to treat juvenile cancer patients. It has been shown that an SNP in the 5V CYP3A4 promoter region, known as CYP3A4*V, decreases enzyme activity and thus epipodophyllotoxin metabolization. This is a genetic reason for the increased risk of leukemia in pediatric oncology after treatment (Herrero *et al.*, 2013).

The integration of pharmacogenetics research into actual clinical practice would result in a substantial medication system for patients carrying defective alleles.

1.5 Role of SNP in Mapping human diseases and risk prediction

SNPs have become an important tool for genetic disorder or other phenotypic trait development and study. A simple comparison of patterns of genetic variations between patients and normal individuals may provide a method of identifying the loci responsible for disease susceptibility. One advantage of this method is that it does not need a large family (Gray *et al.*, 2000). The mapping of the human genome has allowed the construction of a haplotype map to better define the heterogeneity of human SNPs. Haplotypes are clusters of SNPs usually inherited together which is shown in **Figure:1.2.** The haplotype map or HapMap is a method to identify human genetic variation that can

affect health and disease Haplotypes may have stronger associations with diseases or other phenotypic effects compared to individual SNPs and can thus in some cases have improved sensitivity in the diagnosis (Morgan, 2009).

The explanation for the current tremendous interest in SNPs is the expectation that they can be used as markers to identify genes that predispose individuals to specific, multifactorial disorders using mapping of the linkage disequilibrium (LD).

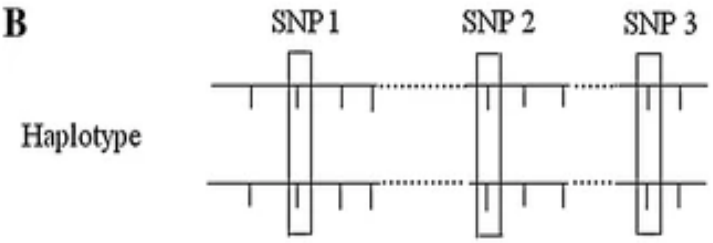

Figure1.2: A haplotype in a long stretch of DNA with distinctive patterns of SNPs at a given location of a chromosome is shown. Haplotype diversity may be generated by new SNP alleles.

1.6 Role of SNPs in Forensic Studies

SNPs provide opportunities to genetically classify a sample of data or to identify an unknown individual. They work as a genetic marker which gives genetic information even if the forensic samples are degraded. SNP analyzes are used in mass disasters and in situations where the DNA may be extensively fragmented (Shastry, 2007). SNPs have relatively low mutation rates and are also reliable genetic markers for lineage-based studies, such as inherited cases, cases of missing persons, and circumstances where no clear comparison sample is available.

1.7 SNP Genotyping

Genotyping is a method used to determine an organism's genetic structure. SNP Genotyping is done to detect clinically significant single nucleotide polymorphisms (SNPs). Most SNPs are biallelic, which means that the genotyping method for a single SNP usually consists of deciding which one of the two nucleotide bases is present at the SNP locus as only one nucleotide change is responsible for causing disease or metabolizing drug(Budowle *et al*., 2018).

SNP genotyping has immense significance. Evaluation and detection of SNP may help predict the reaction of a person to some medications, sensitivity to environmental conditions and the likelihood of developing diseases. Additionally, SNPs may be used to track disease gene inheritance within families and provide insight into the SNP linkage (Kim & Mishra, 2007).

1.7.1 SNP Genotyping Methods

Several good genotyping methods are available to meet the needs of many study designs. In order to move forward the study of genome variability, genotyping technologies need to be continually developed in terms of their accuracy, cost-efficiency and precision. And these approaches depend on the theory upon which they are based.

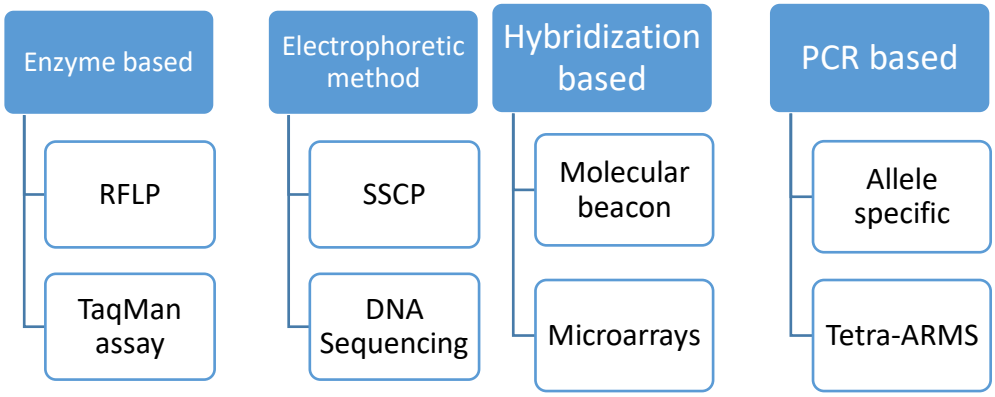

Figure1.3: Some major SNP Genotyping methods with examples of categories.

Here, **Figure 1.3** enlists some methods to detect SNPs. SNP identification method must be chosen based on the study or experiment conducted.

1.8 Enzyme Based Methods

1.8.1 Restriction fragment length polymorphism (RFLP) Method

It is an enzymatic SNP genotyping process in which a region containing a particular SNP allele is targeted and amplified by PCR and then digested with endonuclease enzyme (Brettschneider, 1998).The findings are described based on the resulting length variability of the digested PCR fragments, which are separated by gel electrophoresis which is briefly shown in **Figure 1.3.1.**

In comparison to more complicated SNP genotyping methods, the advantages of SNP-RFLP are fairly simple, fast, inexpensive and convenient methods.

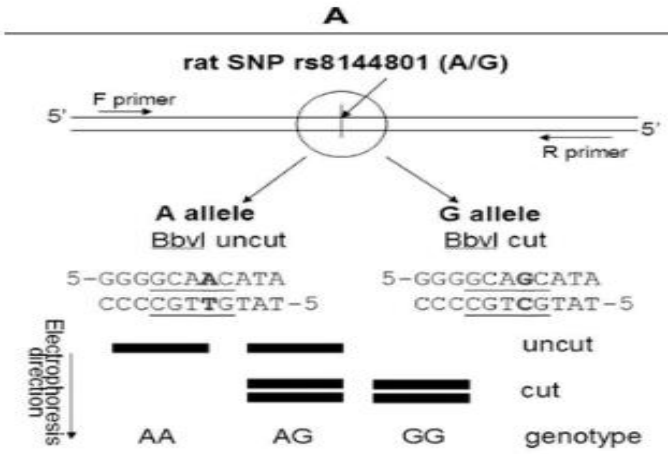

Figure1.3.1: (A) Natural RFLP. The rat rs8144801 SNP with genotype A/G is used as an example. The recognition sequence for restriction enzyme BbvI is 5-GCAGC. Accordingly, the A allele with 5-GCAAC is not recognized for enzyme digestion. As restriction endonuclease did not cut the site in gel electrophoresis AA genotype gave one band.

1.8.2 TaqMan Assay Method

The TaqMan SNP genotyping technology utilizes Taq polymerase's 5 'nuclease activity to produce a fluorescent signal during PCR. The assay uses two TaqMan Probe for each SNP, The technique uses the FRET technology that covalently binds a 5 'reporter dye and

a 3' quencher dye to the wild- and variant allele probes (Shen *et al.*, 2009). TaqMan SNP genotyping assays also require a double-stranded DNA template, the Taq polymerase enzyme, and two primers—forward and reverse—which are specific to the sequence to be amplified.

During PCR extension, the reporter and quencher dyes are released due to the 5' nuclease activity of the Taq polymerase, resulting in increased characteristic fluorescence of the reporter dye (Woodward, 2014).**Figure 1.3.2** summarizes the steps of the TaqMan Assay. The signal ratio will be representative of the specimen genotype.

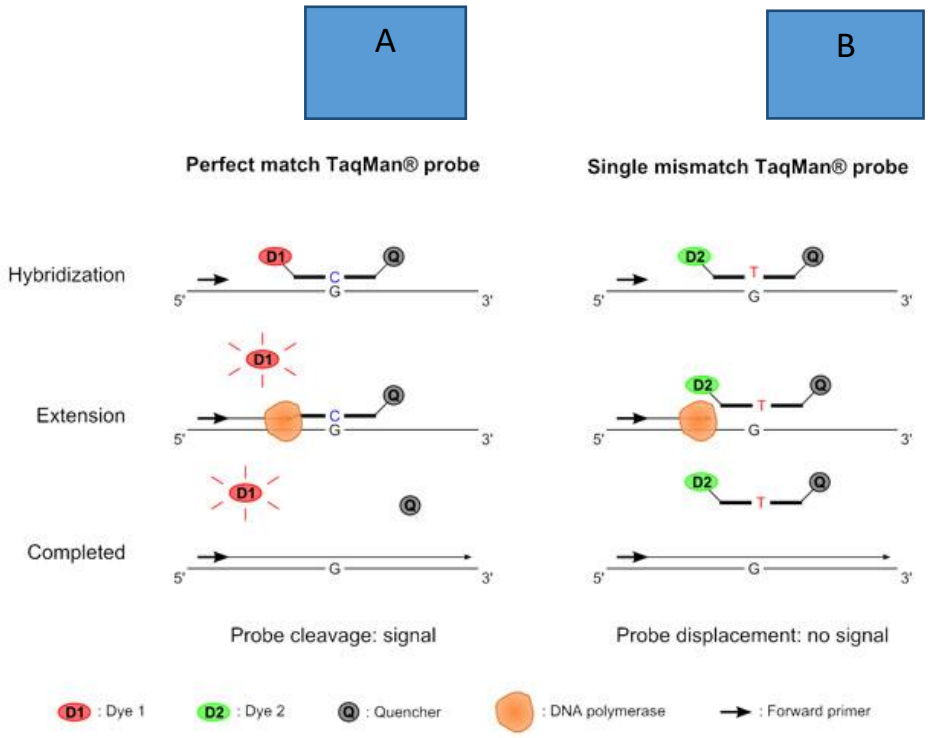

Figure1.3.2: Schematic representation of the TaqMan probe Assay. Here figure A shows fluorescence activity as the probe properly hybridizes with strand and 5' nuclease activity releases quencher dye. In figure B no fluorescence activity is detected. Due to a mismatched base, 5' nuclease activity did not happen.

The TaqMan assay is a suitable and cost-effective approach for studying large populations of SNPs.

1.9 Electrophoretic Method

1.9.1 Single-strand conformation polymorphism (SSCP)

Single-strand conformational polymorphism (SSCP) is still a commonly used genotyping tool for the detection of SNPs across various fields. DNA containing a gene mutation or SNP has a significant mobility gap when exposed to non-denaturing conditions compared with wild type DNA. The SSCP method involves PCR amplification of the target fragment, denaturation of a double-stranded PCR component with heat- and formamide, and non-denaturating polyacrylamide gel electrophoresis (Tahira *et al.*, 2009).**Figure1.3.3** shows the brief steps of SSC. The susceptibility to SSCP ranges greatly from 70% to 95%.

The advantages of SSCP can be considered as rapid, reproducible and quite a simple method that does not require specialized expensive equipment or reagents. The downside of this method is that it is fairly labor-intensive, low throughput, and involves a large sample cohort Sanger sequencing to validate the nucleotide sequence.

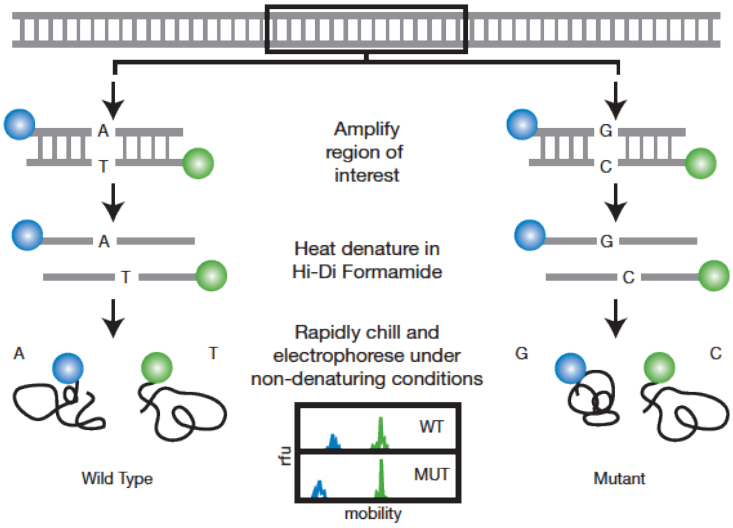

Figure1.3.3: This schematic figure shows the mobility difference of the genetic sample containing SNP.

1.9.2 DNA Sequencing Method

Sequencing techniques of the next decade have stimulated the detection of novel SNPs and other variants. Sanger sequencing results in the formation of extension products of different lengths terminated at the end of 3′ with dideoxynucleotides. Capillary Electrophoresis or CE then separates the extension products. The molecules are pumped into a long glass capillary packed with a gel polymer using an electrical current. After that laser excitation and detection are done by a sequencing machine. One advantage of Sanger sequencing is it has high accuracy. The limitation of Sanger sequencing is it has low throughput.

Whole-genome sequencing (WGS) is an efficient way to produce data for SNP discovery. WGS can be used to evaluate the genotypes of the directly significant SNPs, although the same data can also be used to perform an expanded genome-wide association study (Srilakshmi& Jyotsna, 2019).

Next-generation sequencing (NGS) is another technique. This is less expensive and allows for shorter processing time. In NGS approaches, an entire genome or selected parts of the genome are uniformly digested into tiny fragments that are sequenced and then matched with a reference genome. 'SNP calling' recognizes variable sites while 'genotype calling' determines the genotype at each site for each person.

Next-generation sequencing requires advanced bioinformatics, rapid data processing and robust data storage capabilities that can be costly.

1.10 Hybridization Based

1.10.1 Molecular Beacon

Molecular beacons (MBs) are useful tools for molecular biology, clinical diagnosis and analytical chemistry.

Molecular beacons are single-stranded nucleic acid molecules with a stem-and-loop structure that incorporates a fluorophore and a quencher. The energy that the fluorophore absorbs when bound to the goal is transmitted to the quencher and emitted as heat. There

is a fairly large range of temperatures in which perfectly compatible probe-target hybrids emit a fluorescence response while non-comparable molecular beacons remain dark shown in **Figure 1.3.4.**

Molecular beacons are useful for identifying SNP and SNV since they identify their targets with slightly greater accuracy than traditional oligonucleotide samples (S. Bhattacharyya &P. Ducheyne, 2011).

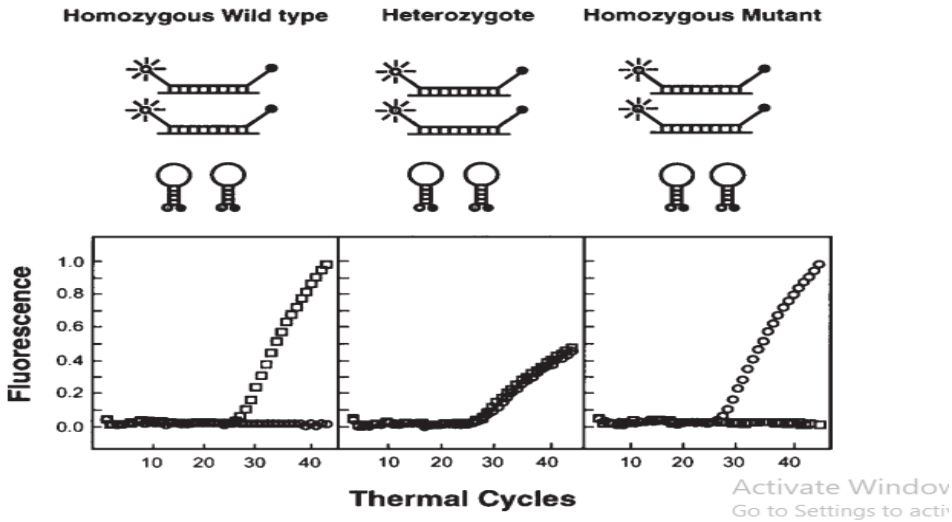

Figure1.3.4: Genotyping Principle for molecular beacons. For DNA from homozygous wild-type individuals, only the molecular beacons labeled for FAM(dye) hybridize into the amplicons, producing fluorescence, which can be measured from the graph. If there is a mutated allele, the molecular beacon cannot hybridize to the template and remain in its hairpin form. So no indication for the presence of an SNP will be observed.

1.10.2 SNP Microarray Method

SNP microarray uses recognized nucleotide sequences as probes to hybridize with the examined DNA sequences, allowing for qualitative and quantitative SNP analysis by sensing a signal.

The SNP microarray method requires many procedures. Those are processing of SNP chips, preparing sample genomic DNA, hybridizing, and screening for fluorescence (LaFramboise, 2009).

SNP microarray is also less costly then sequencing. Disadvantages of SNP Microarray can be a limited resolution, no point mutation information.

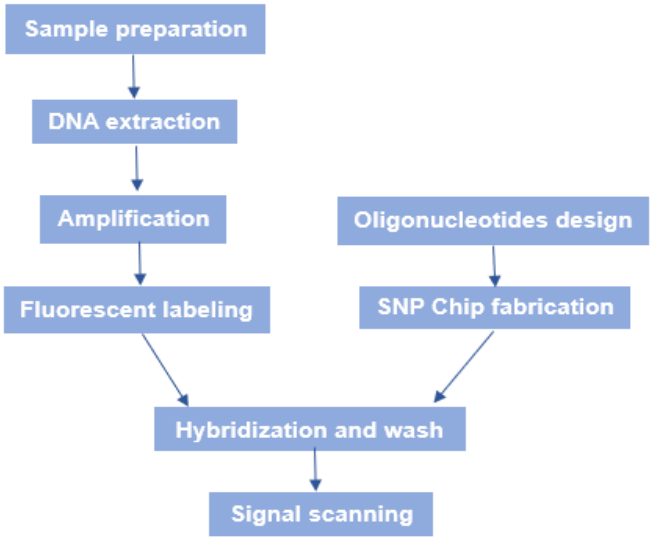

Figure1.3.5: Schematic diagram of SNP Microarray procedure

Compared with the traditional single-cell diagnostic method, the SNP microarray is a high throughput method, which is capable of performing thousands of reactions on the surface of the oligonucleotide chip at one time.

1.11 PCR Based Method

1.11.1 Allele-Specific PCR

This PCR method is based on two AS primers designed to include an SNP and each (reverse) primer to amplify PCR fragments. In the AS primers, a tail is added, thereby allowing separation of the alleles on agarose gel via the size of the PCR products. The varying sequences of the tails allow for differentiation of both AS PCR products. The trick to successful selection between the two alleles is the inclusion of a destabilizing

mismatch, which is different between the two AS primers, within the five bases of the 3 'priming edge (Gaudet *et al.*, 2009). The whole process is shown in **Figure 1.3.6.**

The advantage of the basic AS-PCR is the low cost and fast detection of the amplification products on the agarose gel.

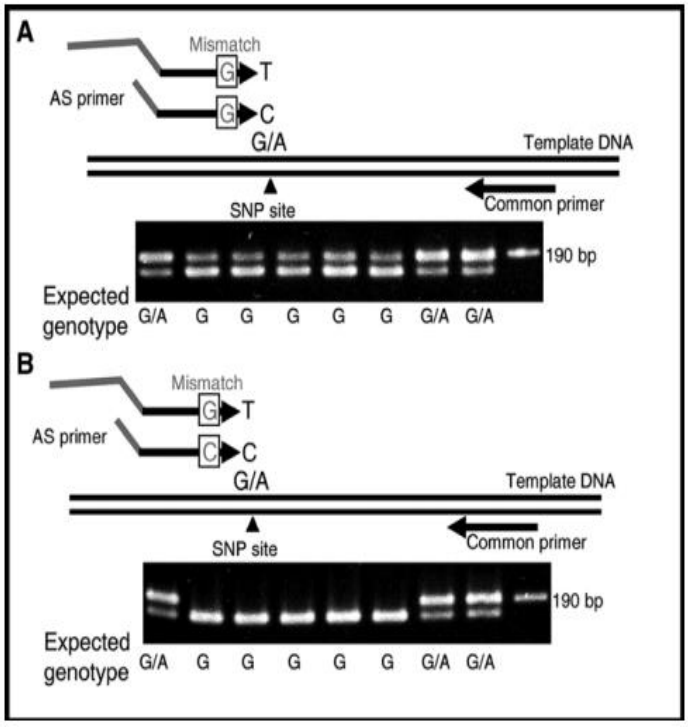

Figure1.3.6: AS- general scheme for a SNP with a similar or different destabilizing mismatch between the two AS primers. The same samples for a particular SNP are genotyped with two separate pairs of AS primers. The expected genotypes are indicated under the gel pictures. The amplification product for the allele A is shorter than the PCR product for the allele G.

1.11.2 Amplification Refractory Mutation System (ARMS) Method

Tetra-primer amplification refractory mutation system – PCR (ARMS – PCR) is a simple and economical SNP genotyping process that only involves a single PCR followed by gel electrophoresis. It is based on the principles of the tetra-primer PCR and ARMS techniques. The allele-specific strategy is based on the use of allele-specific primers with

a mismatch in their 3' terminus, allowing such primers unique to only one SNP allele and refractory to the other. Only when the 3 'end is completely complementary to the template, DNA polymerase complete the priming. A PCR amplicon is formed when this condition occurs. Through placing the outer primers at varying distances from the polymorphic nucleotide, the two fragments unique to the alleles can be identified in an agarose gel through their different sizes shown in Figure **1.3.7.**

The use of tetra-primer ARMS – PCR meets the standards of current genomics research and enables SNPs to be tested rapidly, accurately and cost-effectively. The Tetra-ARMS PCR method is considered as a simple and less expansive method to genotype SNPs as it uses four primers in a single PCR and is followed just by gel electrophoresis. The optimization phase, however, can be very laborious and time-consuming.

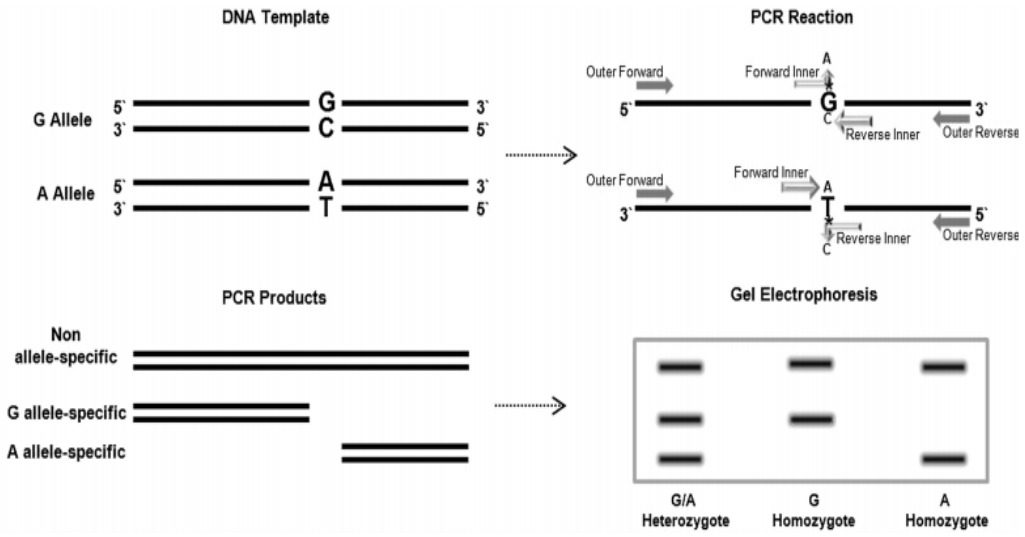

Figure1.3.7: Schematic illustration of the tetra-primer ARMS–PCR assay for SNP genotyping. One example is the C / T substitution of a heterozygote individual. Upper Two allele-specific amplicons are generated using two pairs of primers, one pair (Outer Forward and Reverse Inner) generating an amplicon representing the C allele and the other pair (Forward Inner and Outer Reverse) producing an amplicon representing the T allele. The specificity of the inner primers is imposed by two mismatches, one between the 3' terminal base of the inner primer and the template, and the other from the 3' terminal at position-2. By aligning the two outer primers at different distances from the

polymorphic nucleotide, they differ in length between the two allele-specific amplicons, allowing them to be discriminated upon by gel electrophoresis (Undi *et al.*, 2019)

The most popular technique for low- and medium-throughput SNP genotyping is polymerase chain reaction (PCR). Several PCR methods are available, and a crucial move for successful analysis is to choose the most appropriate one for each experiment. Form of polymorphism, genotyping precision, number of samples and the available PCR equipment are considerations to remember when making the decision. So keeping these factors in mind and after knowing about all other methods to detect SNP, the Tetra-ARMS PCR method was considered as ideal for this experiment.

1.12 The Drug –Clopidogrel

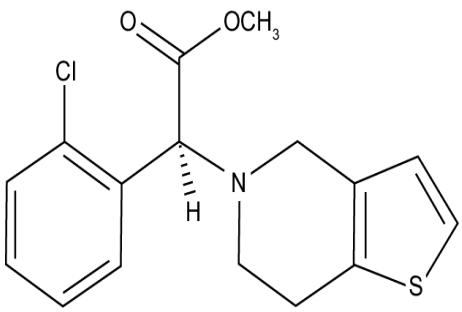

Figure 1.4:Chemical structure of Clopidogrel.

Clopidogrel (brand name Plavix) is an antiplatelet agent. Clopidogrel reduces the risk of myocardial infarction (MI) and stroke in patients with the acute coronary syndrome (ACS), and in patients with atherosclerotic vascular disease.

Clopidogrel is a second-generation thienopyridine antiplatelet drug, which irreversibly inhibits the P2Y12 receptor to exert an antiplatelet effect. Dual antiplatelet therapy (DAPT), especially aspirin and clopidogrel became the recommended treatment of CAD for its reduction of more than 80% cardiovascular events and improvement in the long-term clinical outcomes. Clopidogrel is an inactive prodrug that requires hepatic

bioactivation via several cytochrome P450 enzymes, including CYP2C19. The active metabolite irreversibly inhibits the platelet ADP receptor, P2Y12 (Mark, 2016).

1.12.1 Pharmacokinetics and Pharmacodynamics of Clopidogrel

Clopidogrel relies on numerous genetic and environmental factors to decide both its circulating concentrations of metabolites and its antiplatelet effect. As a consequence, there is sufficient inter-individual variation in clopidogrel response. The inter-individual variability in clopidogrel reaction is informed by the drug's pharmacokinetics and pharmacodynamics pathways involved.

Clopidogrel, which is an inactive pro-drug, needs to be absorbed from the intestine and converts into an active metabolite (H4/R-130964) in the liver by a two-step metabolism.H4 is an isomer of CYP thiol H4 competes with ADP for P2Y12 receptors and inhibits activation of the downstream signaling pathway, and further exerts antiplatelet activity. A recent study has shown that about 50% of clopidogrel is absorbed from the intestine and only 10%-15% of the absorbed clopidogrel is metabolized into an active thiol metabolite by several cytochrome P450 (CYP450) micro-enzymes such as CYP2C19, CYP3A4, CYP2C9, and CYP1A2. Hepatic esterases, which are present in both metabolism stages, compete with the CYP enzymes to produce inactive metabolites. As a result, the consumed clopidogrel is eventually transformed to R-130964 from just 5 to 10 %. Through the liver, R-130964 moves into the general bloodstream, where it irreversibly attaches the ADP receptor to the platelet surface, blocking the activation of glycoprotein IIb/ IIIa which is shown in the **Figure 1.4.1** (Jiang et al., 2015). The P2RY12 gene encodes the ADP receptor, and in some studies, sequence variants within this gene have been identified with decreased antiplatelet activity. CYP2C19 plays a substantial role in the bioactivation of clopidogrel.

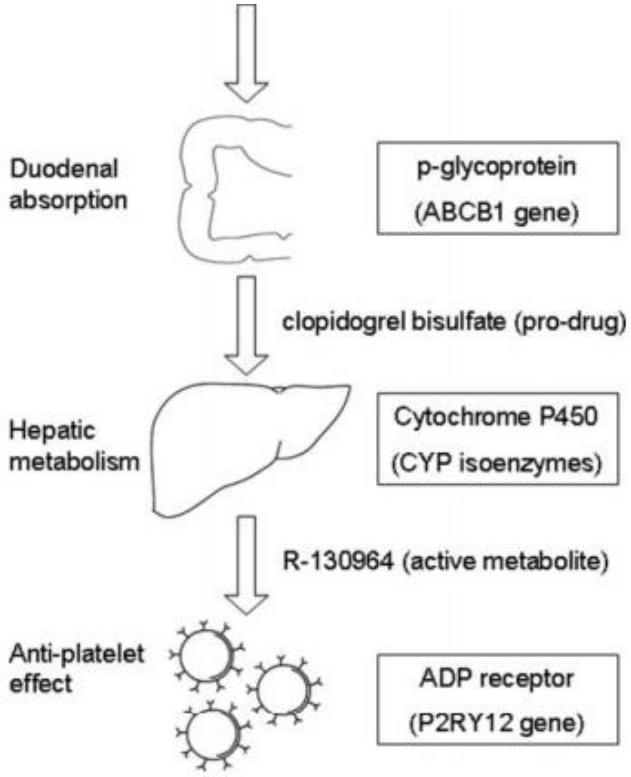

Figure1.4.1: Clopidogrel metabolism. Proteins and genes listed in boxes play the dominant role in the pharmacokinetics and pharmacodynamics of clopidogrel (Anderson *et al.*, 2010).

1.13 CYP2C19 Enzyme Activity

CYP2C19 is the main enzyme active in the hepatic synthesis of various medications. The CYP2C19 gene, which includes nine exons and eight introns, is located at the 10q24.1–10q24.3 locus of chromosome 10, where coding sequences is 1,473 bp and resulting in a protein of 490 amino acid residues.

CYP2C19 genetic polymorphism has been the focus of extensive research and has contributed substantially to the interpretation of interindividual and inter-ethnic variations in the pharmacokinetics of drugs metabolized by this enzyme. CYP2C19, the major hepatic enzyme involved in transforming clopidogrel to its primary activator of the active metabolite, is a highly polymorphic enzyme with more than 25 alleles. The enzyme is

classified into four main phenotypes based on the different alleles: Ultra-rapid metabolizer (UM), Extensive metabolizer (EM), Intermediate metabolizer (IM) and Poor metabolizer (PM).

Research shows normal homozygous genotypes CYP2C19*1/*1 represents for extensive metabolizer, mutation homozygous genotypesCYP2C19*2/*2 represent for poor metabolizer, heterozygote genotypes CYP2C19*1/*2 represent for intermediate metabolizer. The * 1 is the wild-type allele, correlated with the normal function of the enzyme and the phenotype of the 'extensive metabolizer.' The * 17 allele is associated with increased enzyme production as a result of accelerated gene transcription and individuals with one or two copies of the * 17 allele are usually known as 'ultra-fast metabolizers.' Individuals who carry one and two reduced-activity or non-functioning CYP2C19 alleles are 'intermediate' and 'poor metabolizers', respectively (Lee, 2013). They have no enzyme activity and cannot activate clopidogrel via the CYP2C19 pathway, which means the drug will have no effect.

1.14 The SNP-rs4244285

In a GWAS analysis in healthy people, CYP2C19 SNPs were the only genetic variant that had genome-wide significance in relation to drug rates and the CYP2C19 * 2 alleles were correlated with increased risk of cardiovascular events in patients. The risk allele is rs4244285(A).

The substitution of G681A in exon 5 of variant allele CYP2C19 * 2 produces an anomalous splice site resulting in modification of the mRNA reading frame and consequently a truncated nonfunctional protein. The heterozygosity of this SNP is 0.302, which is quite high value. Pro>Pro amino acid change occurs due to this polymorphism. Genetic change can be indicated by G>A / G>C. Data from multiple experiments showed that individuals with the CYP2C19 * 2 allele had compromised the pharmacodynamics response to Clopidogrel when seen with the various assays used for platelet functions (Dehbozorgi *et al.*, 2018). Heterozygous mutant (A/G) individual is an intermediate

metabolizer, homozygous mutant (A/A) is a poor metabolizer and homozygous wildtype (G/G) individual is a normal metabolizer of Clopidogrel.

From different studies, the clinical significance of this SNP has been found. Carriers of CYP2C19*2 (rs4244285), showed a lower level of active clopidogrel metabolite and higher on-treatment platelet aggregation during clopidogrel therapy. In 2006, it was reported a 25% decrease in platelet responsiveness to clopidogrel in the loss-of-function CYP2C19 * 2 alleles in stable voluntary carriers. The frequencies of the CYP2C19 * 2 alleles range from ~15% among Caucasians and Africans to ~29–35% among Asians. Moreover, the CYP2C19*2 allele was associated with a higher risk of stent thrombosis. Studies showed a person bearing the loss of function allele CYP2C19 * 2 significantly decreased platelet inhibition, increased residual platelet aggregation thus raised the risk of adverse cardiovascular events (Dean, 2018),(Mirabbasi *et al.*, 2017). Also, this SNP increases the concentration of the drug tacrolimus in blood. This creates complications in renal transplantation and allograft dysfunction (Lee, 2012).

1.14.1 Worldwide Scenario of CYP2C19*2 Genetic Variation

- Nearly 30 % of the US population bears a loss-of-function allele of CYP2C19, resulting in an elevated risk of significant adverse cardiovascular outcomes if treated with clopidogrel after a percutaneous coronary intervention (Ma *et al.*, 2011).
- Upon suffering a first heart attack, a study of 245 French patients under 45 years of age administered clopidogrel found that rs4244285(A) allele carriers were at 4x greater risk for future adverse cardiovascular conditions relative to non-carriers.
- An analysis of 2,208 French patients administered clopidogrel, 225 of whom died later and 94 of whom had a non-fatal heart attack or stroke, came to two conclusions.

  (1)Patients carrying any two CYP2C19 loss-of-function alleles (* 2, * 3, * 4, or * 5) had around 2x elevated risk of adverse cardiovascular conditions relative to CYP2C19 * 1 homozygotes.

(2) Of the 1,535 patients who received percutaneous coronary surgery (angioplasty) during hospitalization, patients with two loss-of-function alleles of CYP2C19 were at a much greater risk of adverse effects relative to homozygotes of CYP2C19 * 1.

1.15 Importance of the Study

Drugs are not equally effective for all patients. Drugs can also cause serious side effects; in the worst situation, a drug used to cure an illness can produce a fatal outcome. By analyzing genetic variations between people and using these results to prescribe medications, the risk of side effects can be reduced. Since SNPs are responsible for more than 80 percent of the variability between two individuals, they are perfect for the role of searching for genotype-phenotype similarities. Since certain SNPs predispose individuals to have a certain disease to respond to a medication in a particular way, they will be very useful in the development of diagnostics and drugs.

From previous studies, it has been found that the CYP2C19*2 allele is a poor metabolizer of the drug Clopidogrel which is used in heart disease. The person carrying a homozygous mutant allele is a poor metabolizer of the drug and heterozygous mutant allele is an Intermediate metabolizer. That means this drug would not be beneficial for them or can create side effects. Through SNP genotyping we can identify the genotypic trait of a population which will help to administer drug dosing in a sense to develop 'tailored' medicine for an individual.

1.16 Objective of the study

i. Developing a cost-effective method to identify CYP2C19 loss of function allele.
ii. Determination of the frequency of this allele in the Bangladeshi population for further study of population genetics.

To achieve these goals workflow has been developed to make the detection easy and less time-consuming.

- Identification of clinically significant SNP in Heart disease: CYP2C19*2 variant was selected as it is the risk allele and most studied SNP in various populations.
- Designing of Tetra-ARMS Primers for PCR: Four primers were designed to amplify the SNP for genotyping. Primers were designed by following the ideal characteristics of primers.
- PCR and Gel Electrophoresis: Right concentration of PCR reagents was maintained and optimization was done to get efficient amplification of the SNP. Through Gel Electrophoresis wild type, homozygous mutant, the heterozygous mutant can be identified.
- Detection of poor metabolizer phenotype: From the band size phenotypic traits like intermediate metabolizer and poor metabolizer can be identified.
- Determination of the frequency of the allele: Determination of Allele frequency of G and A variant for this SNP is our goal of this study.

The findings of this study can be valuable for further study of Pharmacogenetics and population genetics. Moreover, it can be said that as this method to detect SNP is cheap and easy to interpret so, SNP genotyping to design a 'tailored' drug would be appreciated and available for the Bangladeshi population very soon.

2 Materials and Methods

2.1 Study design

This experiment was designed in three approaches. Firstly, Tetra ARMS primers were designed against the desired SNP and after that PCR reaction was conducted to validate the experiment. Through PCR reaction and gel electrophoresis presence of SNP can be identified along with the homozygous and heterozygous result. Detection of allele frequency and poor metabolizer phenotype can be evaluated by data analysis.

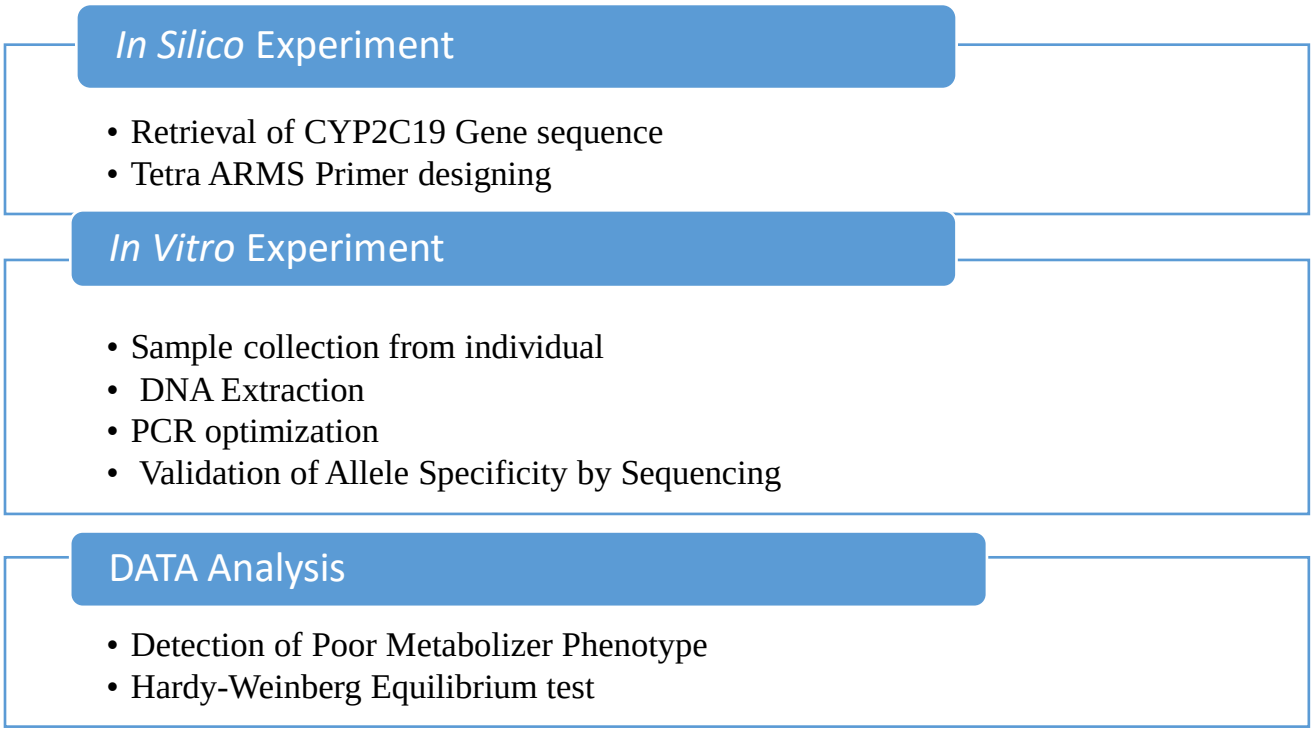

Figure 2.1: Experimental approaches for study design.

2.2 *In silico* Experiment

The purpose of this section of the experiment was to obtain the sequence of the *Homo sapiens* CYP2C19 gene, to develop the allele-specific primers and to test the adequacy of these primers for PCR results.

2.2.1 Deriving the DNA sequence of CYP2C19 Gene

The following steps were performed to retrieve the CYP2C19 gene sequence from the NCBI database:

The NCBI gateway was browsed by using the URL: https://www.ncbi.nlm.nih.gov.On the search box 'CYP2C19 Gene' was typed while the database was set to "GENE" by clicking the drop-down box on the left side of the search box. After clicking the search the first entry on the search result (GeneID:1557) was clicked for more information. On that page in the "Genomic regions, transcripts, and products" option Go to Nucleotide: FASTA option was clicked. Gene sequence in FASTA format has appeared. Send to option was clicked to reveal a drop-down box. Complete record and File were selected to download the sequence by clicking the Create file option.

2.2.2 Designing Tetra ARMS PCR Primer

BatchPrimer3 V1.0 is a gateway for designing various types of primers for experimental studies (You Frank *et al.,* 2008). In this experiment for designing ARMS PCR primers, the BatchPrimer3 V1.0 website was used. Firstly, BatchPrimer3 gateway was browsed by using the URL: https://wheat.pw.usda.gov/demos/BatchPrimer3/. On that page, there was three primer design server option. From the options USDA-ARS, Albany Server 1 was clicked. A new page popped up.

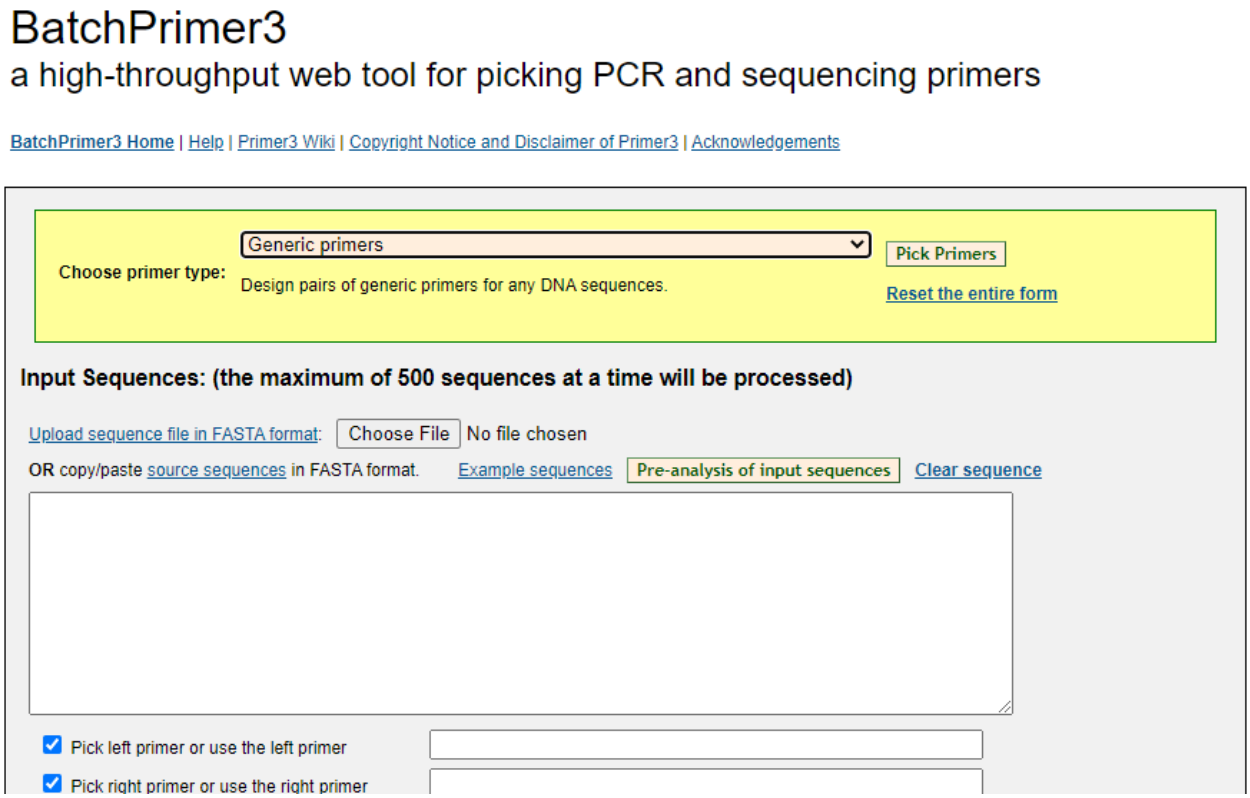

Figure 2.2: USDA-ARS, Albany Server 1 page of BatchPrimer3 Gateway.

On that page, in the 'Choose primer type' section, from the drop-down box, the Tetra primer ARMS PCR primer option was chosen. Before uploading the sequence file in the box mentioned in **figure 2.2**, the sequence must be constructed into the NCBI dbSNP FASTA format.

```
>gnl|dbSNP|rs4244285|allelePos=939|totalLen=1800|taxid=9606|mol="genomic"|cla
ss=1|allele="G/A/C"|build=123

 CTATAAAGTA CTTTGGTGAC AGCCCCAAAG CGTGCTTATA TCACTCCATG GACATCCAGG
 CACTTTGGAG TCTTCCATTA CTCACAAGGC TTGTCCTTCA ATTCACACTT TGTCATATTG
 TGTGACAGAA ATATCCTAAT CTAAAAGACA TTATCTCCTT CAAGGACAGA GAATATTTGG
 AACCACAGAA GCTGCCAAGA AACACTGAAT AGGGCAGAGG TGTTTGATGT CTCAGTTGGG
 ATTCTAGCTG ATGAGACAGC TGGTTAGGAA TGAAAAAATT ATTGTTTTTT TGGTGTATGA
 ACCATAAACA GACATCACAC TTTTACCCTG TGCTGAATTG GCATGTTTTA TACTCTGCCT
 AAATAATAAT TGTATGATTT TACAGAAGTC ATTTAACTGC TCTGGTGCAC AGTTGGAATT
 TGAAGTATCT TTGAGCCCCT CCCACTTCTA AAATACTAAT TCAATTTCAG AGGCTGCTTG
 ATAGAAATCA ATATAGCAGG GACTATCTTT GTAGTATCAA TCAGGTTGTG CAAACTCTTT
 TAACCTATGC TATCATCTCC AAAATGTTAA TGTAGTAATT CATACCATCT TATATTTCAA
 GATTGTAGAG AAGAATTGTT GTAAAAAGTA AGAGAATTAA TATAAAGATG CTTTTATACT
 ATCAAAAGCA GGTATAAGTC TAGGAAATGA TTATCATCTT TGATTCTCTT GTCAGAATTT
 TCTTTCTCAA ATCTTGTATA ATCAGAGAAT TACTACACAT GTACAATAAA AATTTCCCCA
 TCAAGATATA CAATATATTT TATTTATATT TATAGTTTTA AATTACAACC AGAGCTTGGC
 ATATTGTATC TATACCTTTA TTAAATGCTT TTAATTTAAT AAATTATTGT TTTCTCTTAG
 ATATGCAATA ATTTTCCCAC TATCATTGAT TATTTCCC
 R
 GGAACCCATA ACAAATTACT TAAAAACCTT GCTTTTATGG AAAGTGATAT TTTGGAGAAA
 GTAAAAGAAC ACCAAGAATC GATGGACATC AACAACCCTC GGGACTTTAT TGATTGCTTC
 CTGATCAAAA TGGAGAAGGT AAAATGTTAA CAAAAGCTTA GTTATGTGAC TGCTTGCGTA
 TTTGTGATTC ATTGACTAGT TTTGTGTTTA CTACGGATGT TTAACAGGTC AAGGAGTAAT
 GCTTGAGAAG CATATTTAAG TTTTTATTGT ATGCATGAAT ATCCAGTAAG CATCATAGAA
 AATGTAAAAT TAAATTGTTA AATAATTAGA ATACATAGAA GAAATTGTTT AGATAAATAT
 AATCTATCTG AACAATAAGG ATGTCAGGAT AGGAAAAGCT CTGTTCTGCA GCTTCCAGTG
 AGATCAGCAC AGGAGGAACT TAAATTTAAA AGAAAATAAA AAACATCTCC ATCAAAAAGT
 GAGTGAAGGA TATGAACAGA CACTTTTCAA AAGAAGTCAT TGGAGAAATG CAAATAAAAA
 CAACAATGAG ATACCATCTC ATGCCAGTTA GAATGGCGAT CATTAAAAAG TCAGGATATA
 ACAGATGCTG AAGAGGATGT GGAAAAATAG AAATGTTTTT ACTCTGTTTG TGGGAGCGTA
 AATTAGTTCA ATCATTGTGG AAGGCAGTGT GGCAATTCCT CAAGGATCTA GAACTAGAAA
 TACCATTTGA CCCAGCAATC CAATTACTGG ATATATATCC AAAGGATTAT AAATTATTCT
 ATTATAAAGA CACATGTACA CATTTGTTTA TTGTGCCACT ATTCACAATA GCCAAGACTT
 GGAATCAACC CAAATGTACA T
```

Figure 2.3: NCBI dbSNP FASTA format of CYP2C19 sequence for designing Tetra ARMS PCR primer.

Firstly, browsing https://www.ncbi.nlm.nih.gov/gene/,this URL CYP2C19 gene was searched. Gene ID:1557 was clicked for more information. After that, the Genbank option was clicked for detailed information about the CYP2C19 gene. In another tab, dbSNP was opened using the URL https://www.ncbi.nlm.nih.gov/snp/. SNP of interest rs4244285 was searched in the search box. All the information on rs4244285 was revealed. In the 'Genomic regions, transcripts, and products' section nucleotides near the SNP region were copied and searched in the CYP2C19 gene origin section for the exact location of the SNP. Around 2000 bp along with the SNP was copied and opened in word file. Before formatting the sequence it must be kept in mind that, the SNPs or alleles in the sequence must be masked using IUB/IUPAC nucleic acid code (G/C→S, A/T→W, G/A→R, T/C→Y, G/T→K, A/C→M), and the sequence file follows the NCBI dbSNP FASTA format. As the SNP of interest in this experiment is G/A/C it was masked as 'R' and arranged according to the NCBI dbSNP FASTA format shown in **figure 2.3.**

The second step was to select the parameters for the primers. The following are the parameters for Tetra ARMS PCR primers for the SNP rs4244285.

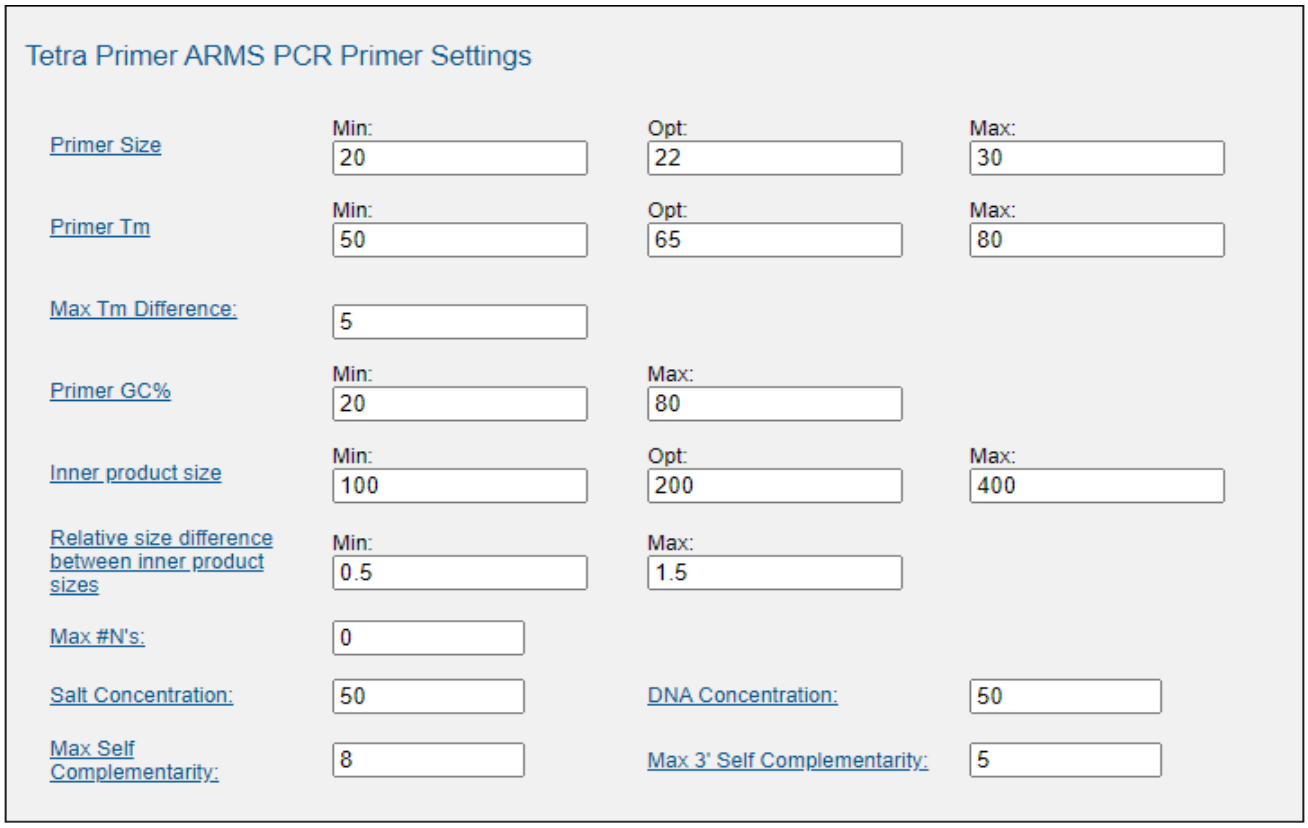

Figure 2.4: Parameters for designing primer sets

After selecting all the parameters and uploading the sequence in NCBI dbSNP FASTA format 'pick primers' option was clicked. Finally, the desired primer set was selected for this experiment.

2.3 *In vitro* Experiment

The first step of *in vitro* experiment was to collect sample. After that DNA was extracted from that sample and a PCR reaction was done to find out genotype frequencies. Finally, they were to be validated by Sanger sequencing.

2.3.1 Sample Collection from Individual

Isolation of DNA from human cheek cells is quite easy then other isolation processes. In this experiment for DNA extraction cheek cells from individuals were collected.

Firstly, 0.9% saline solution was made by adding 0.9 g NaCl in 100 ml distilled water and preserved in a sterile bottle. Healthy volunteers took around 40-50 ml saline solution and gurgled for 1min and retained the gurgle sample in autoclaved falcon tubes. Volunteer identification was performed and the required paperwork was completed.

2.3.2 DNA Extraction from Sample

Sequential steps were followed to collect DNA from the buccal sample. In this experiment for DNA extraction, a scientific paper was followed (Ghatak *et al.*, 2013). In optimization, some steps and amounts were modified. Following is the procedure for DNA Extraction from the buccal sample:

1. Samples collected from individuals were transferred to an Eppendorf tube of 1500 µl from a falcon tube. The Eppendorf tubes were centrifuged at 4500 rpm for 5 minutes. The steps were repeated 3/4 times to get a good pellet to increase cell concentration.

2. The pellet was suspended in 500 µl lysis buffer [10 mM Tris (pH 8.0),10mM EDTA, and 2.0% SDS] and 50 µl 10% SDS, followed by 5-10 µl 20 mg/ml proteinase K was added. The samples were then incubated for 2 hr at 65°C.

3. The DNA was then extracted from the samples with an equal volume of Phenol:Chloroform: Isoamyl alcohol (25:24:1) and mixed gently by inverting the tube for 3 minutes. The samples were then centrifuged for 10 min at 10.000 g (4°C) and the upper aqueous layer was transferred to a fresh sterilized micro-centrifuge tube.

4. RNase A (10 µl of 10 mg/ml) was added and the solution was incubated at 37°C for 30 min. An equal volume of Chloroform: Isoamyl alcohol was added and centrifuged again at 10,000 g (4°C) for 10 min.

5. The upper aqueous layer was transferred to a sterilized micro-centrifuge tube and double the volume of chilled isopropanol was added, along with a one-tenth volume of 3 M sodium acetate and chilled at -20°C for 1hr for precipitation.

6. After 1hr the samples were centrifuged at 10,000 g (4°C) for 10 min. After decanting the supernatant, 250 µl 70% ethanol was added and the pellet was dissolved. The mixture was centrifuged at 10,000 g for 10 min and the supernatant was decanted gently.

7. The pellet was air-dried under laminar airflow and the air-dried pellet was re-suspended in 50 µl Nuclease free water and stored at -20°C.

2.3.3 PCR Optimization

PCR optimization is the crucial and time-consuming step of this experiment. The optimization process is done to improve PCR performance and minimize failure.

2.3.4 Preparation of Primers

Before starting the PCR reaction primers must be prepared. Following are the steps of preparation of master stock and working stock for primers:

1. A short spin was performed on the lyophilized primer to gather them at the bottom of the tube.

2. To prepare 100 µM master stock primer from 25 nmole lyophilized primer, 250 µl nuclease-free water was added and mixed.

3. Before the preparation of the working stock, the solution was kept at room temperature for 10 mins.

4. 1 µl of master stock was added to 9 µl of nuclease-free water to prepare 10 µl of working stock of 10 µM in concentration.

2.3.5 The first attempt of Polymerase Chain Reaction

The following steps were taken to determine the efficacy of the primer pairs in PCR reactions:

Five Nuclease free PCR tubes were labeled. Four tubes contained DNA template and the fifth one contained all other PCR components but not the DNA template, it worked as the negative control. There were two sets of primers- two outer primers, two inner primers. The annealing temperature was determined at 58° C. This temperature was selected based on the melting temperature of the primers predicted by 'Tm Calculator' at ThermoFisher Scientific platform. The PCR master mixture was prepared according to **Table 1.1**. All the components were mixed sequentially in the same order they have been stated here. The thermal cycler instrumentation was configured with the cycling condition shown in **Figure 2.5.**

Table 1.1: Recipe for the first attempt of PCR reactions. All the volumes are taken in µL.

10X DreamTaq Reaction Buffer*	10mM dNTP mix	Outer Forward primer	Outer Reverse primer	Inner Forward primer	Inner Reverse primer	DreamTaq DNA Polymerase	PCR H_20	DNA Template	Total
2.5	0.5	0.5	0.5	0.5	0.5	0.2	16.8	3	25

*Contains $MgCl_2$

The PCR tubes were placed in the PCR machine. After setting all the conditions PCR reaction was started. The reaction took around 1 hr 30 min to be completed. Meanwhile 2% Agarose gel was prepared to resolve the PCR product. 100 kb 2 µl ladder was loaded in gel.7 µl PCR product along with 1 µl loading dye was loaded in the wells. Electrophoresis was done for 1 hr at 80 Volt. The DNA bands in the agarose gel were visualized under UV-rays and an image (**Figure 3.4**) was captured.

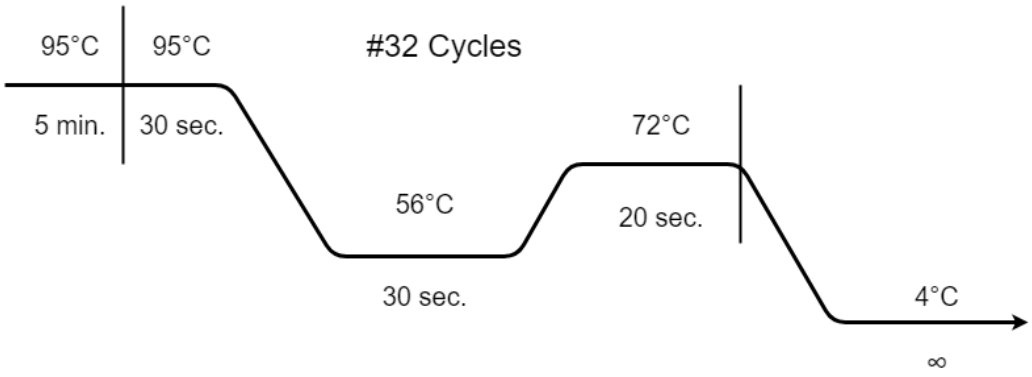

Figure 2.5: Thermal cycling condition for the first attempt at optimizing PCR reaction conditions.

2.3.6 The second attempt of Polymerase Chain Reaction

In the second attempt of the PCR reaction, the annealing temperature was changed along with the primer concentration. Five Nuclease free PCR tubes were labeled. The PCR master mixture was prepared according to **Table 1.2**. All the components were mixed sequentially in the same order they have been stated here. The thermal cycler instrumentation was configured with the cycling condition shown in **Figure 2.6**. After PCR reaction Gel electrophoresis was done for 1 hr at 80 Volt. The DNA bands in the agarose gel were visualized under UV-rays and an image **(Figure 3.5)** was captured.

Table 1.2: Recipe for the second attempt of PCR reactions. All the volumes are taken in µL.

10X DreamTaq Reaction Buffer*	10mM dNTP mix	Outer Forward primer	Outer Reverse primer	Inner Forward primer	Inner Reverse primer	DreamTaq DNA Polymerase	PCR H_20	DNA Template	Total
2.5	0.5	0.5	0.5	0.75	0.75	0.2	16.3	3	25

*Contains $MgCl_2$

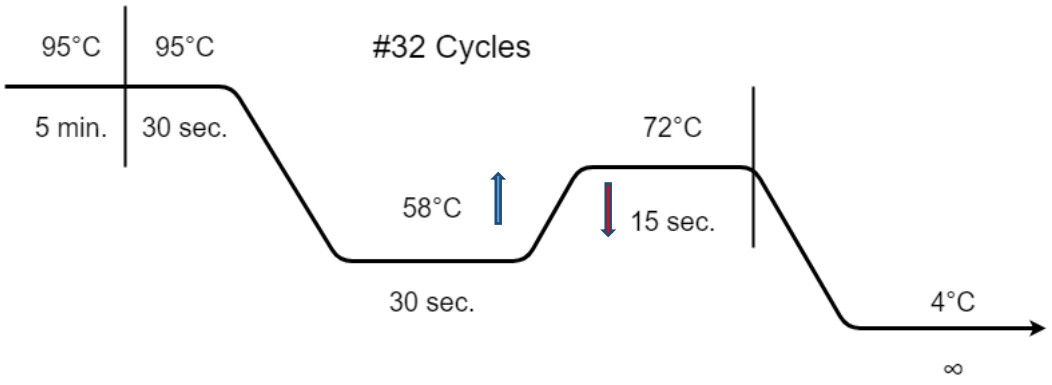

Figure 2.6: Thermal cycling condition for the second attempt of optimizing PCR reaction conditions.

2.3.7 Confirmation of PCR reaction

In the first attempt all the bands in gel electrophoresis were quite clear, but in the second attempt after changing the annealing temperature and concentration of primers no band came. That is why the recipe for the master mix and thermal cycler condition of the first attempt was finalized for PCR for further reactions.

2.3.8 Performing PCR Reaction with Individual DNA Samples

Once the PCR reaction recipe and cycling conditions were finalized, it was time to perform PCR with individual DNA samples. Around sixty samples were collected from healthy volunteers and about thirty-six DNA samples were extracted to do the PCR reaction with selected conditions.

2.4 Validation of result by Sanger Sequencing

Sanger sequencing is the process by which chain-terminating dideoxynucleotides are selectively incorporated by DNA polymerase during in vitro DNA replication. The approach was widely used to advance practical and comparative genomics, developmental genetics, and complex disease studies.

For SNP Genotyping Sanger sequencing can be very useful. In the genotyping process, primer specificity is one of the main concerns. Through PCR reaction, primer specificity can be observed but for confirmed validation of the result, Sanger sequencing is a must.

After PCR optimization and final PCR reaction, we were supposed to do the Sanger Sequencing to validate our result. Due to the COVID-19 pandemic situation, we couldn't continue further experiments.

3 RESULT

3.1 DNA Sequence of Homo sapience CYP2C19 Gene

Major attributes regarding the Homo sapiens CYP2C19 gene have been derived from NCBI's GenBank entry and summarized in **Table 3.1**. The gene is located within a cluster of cytochrome P450 genes on chromosome 10q24.

Table 3.1: The different sequence attributes of the Homo sapiens CYP2C19 gene

Homo sapience CYP2C19 Gene	
Official Symbol	CYP2C19
Official Full Name	cytochrome P450 family 2 subfamily C member 19
Accession Code	NC_000010
Gene ID	1557
Gene Type	Protein Coding
EC Number	1.14.14.51
RefSeq Status	Reviewed
Organism	*Homo sapience*
Location	10q23.33
Length of the Gene sequence	92867 bp
Region of the Gene sequence	Complement (746..11948)
Total Exon Count	9
Exon Position	(1..193,12378..12540,12710..12859,17819..17979, 19141..19317,57816..57957,80157..80344,87237..87378, 90053..92867)
Transcript ID	NM_000769.4
RefSeq Protein ID	NP_000760.1

CYP2C19 has a total of 9 Exons. rs4244285 (c.681G > A) is the defining polymorphism of the CYP2C19*2 allele (previously referred to as CYP2C19) and is a synonymous G > A transition in exon 5 that creates an aberrant splice site. This change alters the mRNA reading frame, which results in a truncated, nonfunctional protein. **Table 3.2** describes the features of this SNP retrieved from Ensembl and SNPedia.

Table 3.2: Different attributes of SNP rs4244285

Characteristics	Description	
SNP name	CYP2C19*2	
rsID	Rs4244285	
Synonyms	Archive dbSNP	rs60361278, rs17879456, rs116940633
	ClinVar	VCV000016897, VCV000633856
	OMIM	124020.0001
	PharmGKB	PA166154053
	ClinGen Allele Registry	CA126959 (A), CA211681888
Transcript strand	ENST00000371321.9, forward	
Strand	Plus (+)	
Consequence type	Synonymous Variant	
Position in Chromosome	94,781,859	
Position in Exon	681	
Codon	CCG>CCA	
Amino acid	Proline>Proline	
Variation Type	SNV (Single Nucleotide Variation)	

3.2 Designing Tetra ARMS PCR Primers

Before designing the primer set CYP2C19 gene sequence must be retrieved from NCBI. Figure 3.1 shows the search result in dbSNP from NCBI gateway which indicates the 5' flanking sequence of the CYP2C19*2 SNP:

5 '–TCATTGATTATTTCCC -3'.

Figure 3.2 shows the schematic representation of the CYP2C19 gene along with the SNP, deliberate mismatch and primer positions.

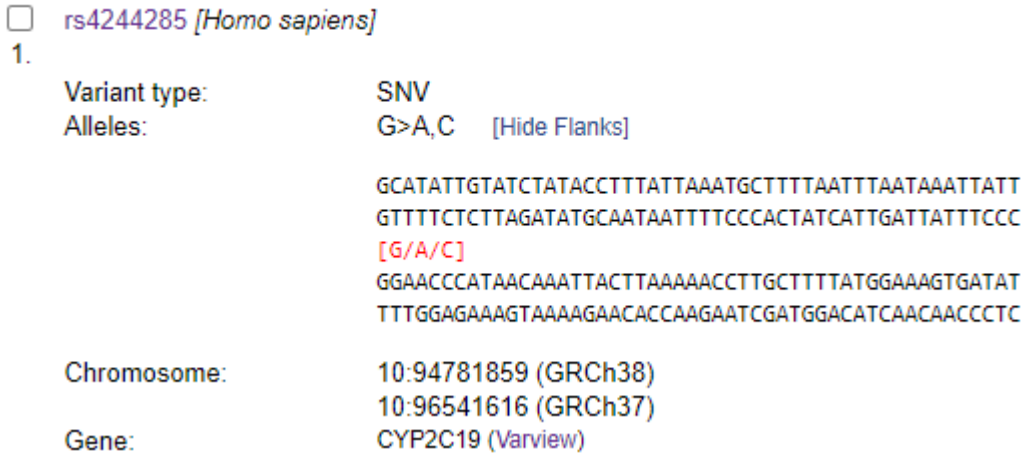

Figure 3.1: Search result in dbSNP from NCBI gateway indicating the flanking sequence of the CYP2C19*2 SNP

Here is a figure of the word document containing the CYP2C9 gene sequence, highlighted with important portions and positions.

GTATCTATACCTTTATTAAATGCTTTTAATTTAATAAATTATTGTTTTCTCTTA
GATATGCAATAATTTTCCCACTATCATTGATTATTTCCCGGAACCCATAACA
AATTACTTAAAAACCTTGCTTTTATGGAAAGTGATATTTTGGAGAAAGTAAA
AGAACACCAAGAATCGATGGACATCAACAACCCTCG

Figure3.2: Word document containing CYP2C19 gene sequence, highlighted with important portions and positions. 5′ flanking sequence, mismatch position and SNP locus is highlighted with blue, yellow and red respectively.

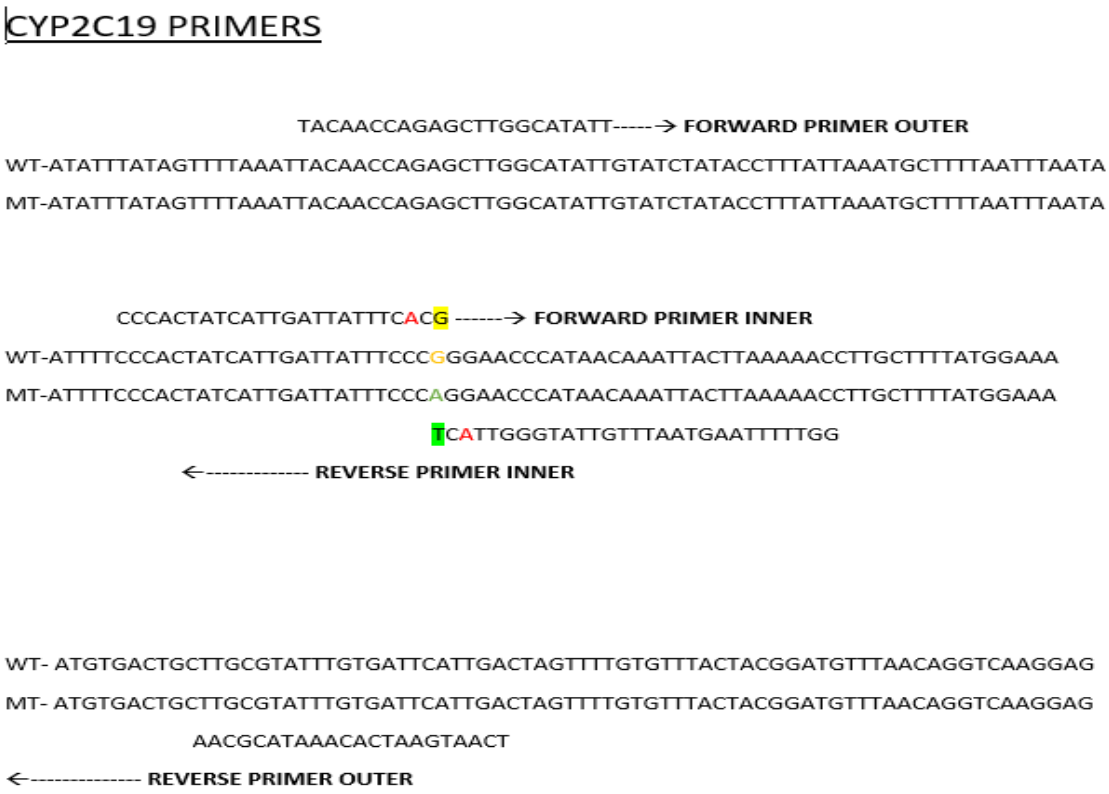

Figure 3.3: Schematic representation of the CYP2C19 gene. Here wild type allele, mutant allele, a deliberate mismatch is highlighted as yellow, green and red respectively.

Tetra ARMS PCR primers designed using the gateway BatchPrimer3 V1.0 is demonstrated in **table 3.3**. Four primers are used to amplify a larger fragment of SNP-containing template DNA, and two smaller fragments for each of the two AS products From the primary design context, it is possible to design two sets of tetra-primers for each SNP possibly according to the alignment of the AS primers. A batch module was introduced by BatchPrimer3 v1.0 to conveniently build two sets of tetra-primers for an SNP.

Table 3.3: Attributes of Tetra ARMS PCR Primers designed for this study

Primer Type	Length bp	Tm	GC%	3' Compl	Allele	Position	Primer Seq	Product Size
Outer Forward	22	58.79	40.91	2			TACAACCAGAGCTTGGCATATT	311
Outer Reverse	22	59.08	31.82	0			TCAATGAATCACAAATACGCAA	
Inner Forward	24	59.64	37.5	2	G	699	CCCACTATCATTGATTATTTCACG	219
Inner Reverse	29	58.84	27.59	3	A	699	GGTTTTTAAGTAATTTGTTATGGGTTACT	144

From t**able 3.3**, we can see that the deliberate mismatch of the Inner forward and the inner reverse primer was marked red. The length of primers is within 30 bp. The Tm value of primers is around 59°C and GC content varies from 28-40%. As there is a favorable product size difference they can be resolved by agarose gel electrophoresis easily.

3.3 Result of the first attempt of PCR reaction

In the first attempt of the PCR reaction using the primer set, annealing temperature was determined at 58°C. The concentration of Primer was 0.5 µl for each. PCR cycle was determined 32. The image of product bands in agarose gel is mentioned in **figure 3.4.**

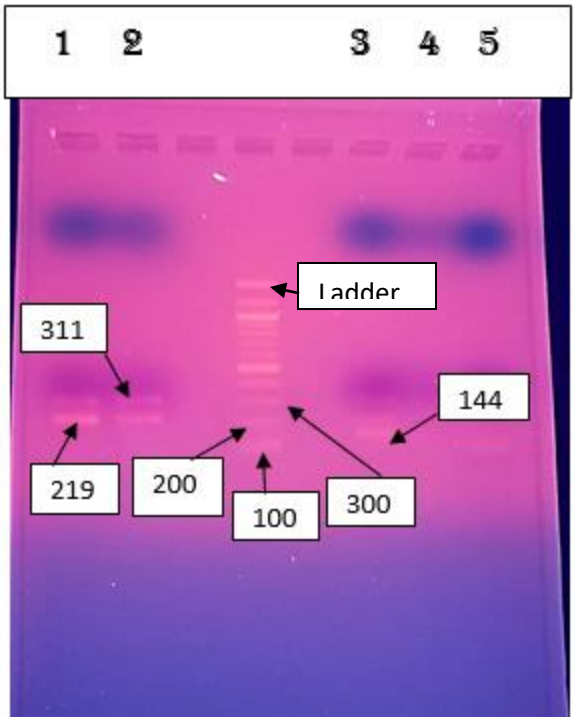

Figure 3.4: Result of the first attempt of PCR reaction. A 100 kb ladder was used to track the DNA bands.

From the **figure3.4** we can see that three samples gave clear bands. The PCR reaction yields an amplicon of 311 bp which is a common amplicon of all genotypes. 219 bp amplicon was designed for wild type allele (G) and 144 bp amplicon was designed for the mutant allele (A). Therefore, it can be said that- Sample (1) and (2) were Homozygous wildtype (G/G WT). Sample (3) gave amplicon at 144bp which was for the mutant allele, so sample (3) is Heterozygous mutant (G/A MT). Sample (4) didn't give any band. Sample (5) was a negative control.

3.4 Result of the second attempt of PCR reaction

In the second attempt of the PCR reaction, the annealing temperature was determined 56°C. Elongation time was increased to 20 sec. Primer concentration was 1:1.5 µl. The image of product bands in agarose gel is mentioned in **figure 3.5.**

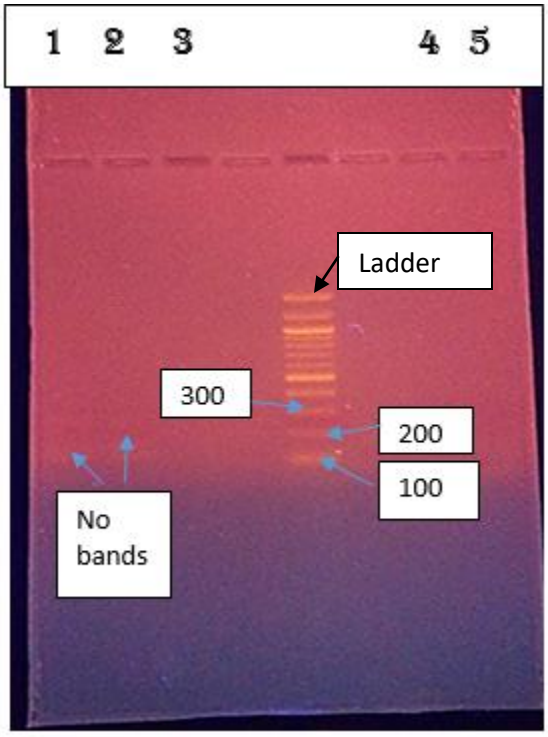

Figure 3.5: Result of the second attempt of PCR reaction. A 100 kb ladder was used to track DNA bands.

From **figure 3.5**, we can see that there are no visible bands. That means the PCR condition was not suitable for PCR reaction.

3.5 Final Result of PCR Optimization

From the two attempts of PCR reaction, we can come to the point that the conditions of the first attempt of PCR reaction are suitable for this study. The annealing temperature of 56° C, forward and reverse primers at 0.5 µl concentration for each type and elongation at 72° C for 15 sec are the proper condition for this experiment. The final PCR reaction with extracted DNA from individual's samples couldn't be conducted. Due to the COVID-19 Pandemic situation our university closed for an uncertain period of time. For this reason, the further experiment couldn't be carried out.

4 Discussion

4.1 General Discussion

Pharmacogenetics is a pharmacological branch associated with the effects of genetic variables on drug reactions. There are so many genetic factors influencing drug response and single nucleotide polymorphism is one of them. An SNP occurs when a single nucleotide is exchanged for another at a point in an individual's genome. These polymorphisms in turn modify the structure of the protein component and lead to major phenotypic changes in the therapeutic response or carcinogenic susceptibility. The same dose of drugs may induce adverse drug effects and toxicity across populations containing the risk factor. Thus, Pharmacogenetics research is useful in an attempt to individualize therapy while finding answers to the molecular basis of contextual drug response variations. SNP Genotyping is done to detect clinically significant single nucleotide polymorphisms (SNPs). There is a variety of genotyping approaches available to meet the needs of several studies which is mentioned in **section1.1.4.** Every approach has its pros and cons. For this experiment, the Tetra ARMS PCR method was very suitable. We attempted to use this method to genotype one of the most clinically significant SNP (CYP2C19*2) in the gene CYP2C19. Primarily the key objective of this research was to implement a basic method of detection for this clinically important CYP2C19 * 2 SNP, which would be simpler to execute, cost-effective and can be used for routine diagnostic purposes.

Clopidogrel is a blood thinner or antiplatelet drug that makes blood less likely to produce a harmful blood clot. It is a prodrug of a platelet inhibitor used to reduce the risk of myocardial infarction and stroke. Clopidogrel is an inactive prodrug that needs hepatic bio-activation by multiple enzymes of cytochrome P450 including CYP2C19. The active metabolite inhibits the platelet ADP receptor P2Y12 irreversibly. CYP2C19 is the major hepatic enzyme involved in the transition of clopidogrel into its key active metabolite.CYP2C19 genetic polymorphism has been the subject of comprehensive study and has greatly contributed to understanding inter-individual and inter-ethnic differences

in the pharmacokinetics of drugs metabolized by this enzyme. Finding shows typical homozygous genotypes CYP2C19 * 1/*1 for comprehensive metabolizers, homozygous mutants CYP2C19 * 2/*2 for poor metabolizers, heterozygous genotypes CYP2C19 * 1/*2 for intermediate metabolizers. So, CYP2C19*2 is the risk allele. The individual having this risk allele has no enzyme activity and is unable to activate clopidogrel via the CYP2C19 pathway. From this experiment, we found that sample (1) and (2) were homozygous wildtype (G/G), sample (3) was heterozygous mutant (G/A). That means, samples (1) and (2) do not contain the risk allele, but sample (3) has a risk allele and intermediate metabolizer of Clopidogrel. Therefore, sample (3) must be prescribed an adjusted dose ofclopidogrel if diagnosed with cardiovascular disease.

CYP2C19 * 2 allele frequencies ranged from ~15% among Caucasians and Africans to ~29–35% among Asians. This SNP raises the concentration level of the drug 'Tacrolimus' in blood. This creates complications in renal transplantation and dysfunction of the allograft. That is why individual having this risk allele is given a reduced dose of this type of drug when a suitable alternative is absent and prescribed for an alternative when present. Through SNP genotyping we can classify the genotypic character of the population that will help in administering drug dosing in a sense to develop individually tailored medicine. This study design is developed to genotype a very clinically significant SNP using the Tetra ARMS PCR method. Unfortunately, because of this pandemic situation with COVID-19, we could not finish our experimental work, but at least we developed a cost-effective approach for genotyping this SNP which can be used in diagnostic purposes with further validation.

4.1.1 Remark on *In silico* Primer Designing

The first objective of this study was to design allele-specific tetra primers that will identify the alleles of the SNP in a simple PCR reaction. BatchPrimer3 V1.0 was used to design the tetra primers, the result is shown in **table 3.4**. In general, an ideal primer must contain some features. The length of primers must be within 18-25 bp. The lengths of the designed primers were within this range. GC% is a very important feature before designing primers. GC content should be within 40-60%. Our designed primer also

followed this feature. It is needed to keep in mind that the length and composition of primers directly affect the PCR annealing temperature (Ta). A melting temperature (Tm) of 52° C to 58° C is a good starting range when designing primers. A deliberate mismatch was introduced into the primers. This deliberate mismatch ensures primer specificity. Finally, it was found that the primers built in the in silico study are theoretically relevant to the accomplishment of our SNP detection objective.

4.1.2 Remark on DNA Extraction from the Buccal sample

Firstly, we tried to extract DNA from the buccal swab using the protocol of the PureLink Genomic DNA kit. We collected buccal samples with the swab and kept the swab in lysis buffer for 24hr. We conducted the extraction procedure as described in the PureLink Genomic DNA Kit user guide, but couldn't retrieve the desired DNA bands. As the method is time-consuming and we were facing problems with the extraction process we tried another process of DNA extraction.

We made a solution using 0.9g NaCl in distilled water to collect DNA samples from individuals. For the further extraction process, the Phenol Chloroform extraction method was used. The lysis buffer used in this method contains Tris, EDTA and SDS. Tris is used to maintaining stable pH. Additionally, it plays an important role in cell lysis. EDTA works as a chelating agent. It reduces the activity of DNase and RNase. SDS is an anionic detergent that solubilizes protein and lipids. Proteinase K used in this method is widely used for digesting protein and extracting contaminants from nucleic acid preparations. The aim of using Chloroform along with Phenol is to ensure that the aqueous and organic phases are easily differentiated. It is useful when the aqueous phase is separated from the solution to obtain a pure sample of nucleic acid. Isoamyl alcohol helps in minimizing foaming during interphase in the extraction process of DNA. This prevents a solution from emulsifying. The liquid phase contains DNA and lipid, protein and other impurities that are separated in the organic phase. RNase A is an important enzyme for the removal of RNA for RNA free DNA purification which is used in this method after the phenol-chloroform isoamyl alcohol treatment. After this for more pure DNA, the sample is treated with Phenol Chloroform. To extract some of the salt from the

pellet, DNA is treated with 70% ethanol. Since 100% of ethanol precipitation contributes to the removal of all water molecules from DNA and total dehydration, which leaves them insoluble. Pellet is then re-suspended in nuclease-free water for a better result in PCR.

4.1.3 Remark on PCR Optimization

The Tetra primer sets, manually designed using the BatchPrimer3 V1.0, should be calibrated in the laboratory before being used for individual genotyping purposes. Various parameters are contributing to the success or failure of PCR reaction, like primer ratio, annealing temperature, the concentration of $MgCl_2$. All these parameters must be optimized before conducting the final PCR reaction. Nonspecific band, primer dimer problems can be minimized by optimizing these parameters. In this experiment, a negative control was used in the PCR reaction. From **Figure 3.4** we can see that sample (5) containing the negative control didn't give any band. Which means PCR was operated successfully. There was a common band in every lane at the 311bp region so this band acted as the positive control. Thermal cycling is an integral feature of PCR, which ensures the reaction is exposed to a range of correctly specified temperatures. The reaction mixture initially heats up to 95 ° C. It allows the hydrogen bonds between the strands of template DNA molecules to melt or denature. It creates two single-stranded DNA molecules out of each double helix. Then cool down for the primers to be annealed. After that again temperature is raised and a new strand of DNA is made. So determining the PCR cycle number is crucial. The annealing temperature for PCR condition was primarily 56ºC, 57 ºC, 58 ºC according to the primers' Tm. We were supposed to optimize the annealing temperature. At the first attempt of PCR, the annealing temperature was 56°C for the 30 sec. We got a clear band at this temperature. So, PCR conditions were optimized properly for this study.

4.2 Future Perspective

Various methods are available for single nucleotide polymorphism (SNP) genotyping each with its benefits and limitations. Some of them can be costly and time consuming for on the spot diagnostic testing. In this research project, we have developed a SNP genotyping protocol to find out whether an individual is a wild type homozygous, homozygous mutant, or heterozygous mutant for a particular SNP.

Unfortunately, because of this pandemic situation, we couldn't implement our full study design and accomplish the goal. But if we could continue our experiment, we might genotype 100 samples. For instance, from 100 genotyping results, 10% of them were a homozygous mutant,20% of them were a heterozygous mutant, and 70% them of wild type homozygous for this SNP. From this result, we could say that 10% of these volunteers have no enzyme activity to metabolize the drug clopidogrel. So, they must be prescribed alternative medication if they diagnose with Coronary artery disease in the future. Through this genotyping presence of loss of function allele CYP2C19*2 can be identified and an appropriate dose can be prescribed to patients.

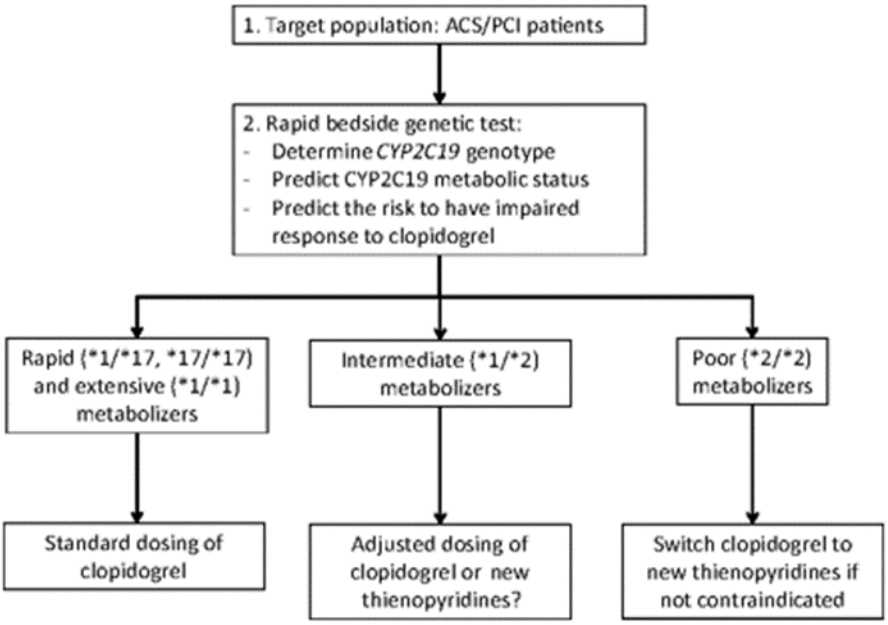

Figure 4.1: An algorithm to refine clopidogrel treatment and dose with respect to CYP2C19 genotype for the target population of patients with ACS / PCI (Alexopoulos, 2011).

CYP2C19 genetic testing is capable of recognizing patient subgroups who do not benefit from regular clopidogrel therapy. The first step is to determine the presence of etiological alleles forms the patient sample. This experiment represents a rapid, accurate and cost-effective method to detect the presence of the CYP2C19 SNP variants.

Moreover, this genotyping method can be used to determine the allele frequency of this SNP. Allele frequency (also known as gene frequency) is the concept used to characterize the fraction of gene copies in a given population that is of a specific allele. For population genetics, allele frequencies indicate the genetic variation of a species population or the abundance of its gene pool equivalently (Cross *et al.*, 2010). For example, if all the alleles in a pea plant population were purple alleles, W would be 100%, or 1.0. Nevertheless, if half of the alleles were W and half were W, the frequency of each allele would have a 50 %or 0.5. Changes in allele frequencies over time can indicate that genetic drift is occurring or that new mutations have been introduced into the population. Detection of disease-associated or drug-metabolizing allele frequency would help to administer population genetics and pharmacogenetics studies. As this experiment is done over the Bangladeshi population, detection of the frequency of this allele would help to determine drug dose so that drug toxicity,side effects, or overdosing of the drug can be avoided in Bangladeshi people.

Conclusion

In recent years, Pharmacogenetic studies have become popular due to the increasing awareness about the influence of genotypic reaction to a given drug. People became more interested in having their own personalized medicine that would work most efficiently. Therefore genotyping to determine the presence of risk alleles is a must before taking personalized medication (Zhang & Nebert, 2017). The hepatic enzyme CYP2C19 metabolizes important drugs including Clopidogrel. If an individual lacks this enzyme he/she may face problems metabolizing this drug and that could be life-threatening.

We developed a PCR technique with Tetra Primer ARMS to detect the risk allele. This approach is cost-effective and simple to enforce. The optimization of PCR condition is time-consuming, other than this method is easy to execute in routine diagnostic approaches.

With further experiments, the data form of this study would help in the field of population genetics. Through this experiment allele frequency of this SNP can be determined in the Bangladeshi population.

References

1. Alexopoulos, D. (2011). CYP2C19* 2 alleles: guiding medication choice after percutaneous coronary intervention. Interventional Cardiology, 3(1), 1-4.

2.Alwi, Z. B. (2005). (1). The Malaysian journal of medical sciences: MJMS, 12(2), 4.

3.Anderson, C. D., Biffi, A., Greenberg, S. M., &Rosand, J. (2010). Personalized approaches to clopidogrel therapy: are we there yet?. Stroke, 41(12), 2997-3002.

4.Asthana, S., &Sunyaev, S. (1970, January 1). Understanding the Functional Importance of Human Single Nucleotide Polymorphisms. Retrieved from https://link.springer.com/chapter/10.1007/0-387-36747-0_11.

5.Ballesta, P., Bush, D., Silva, F. F., & Mora, F. (2020). Genomic Predictions Using Low-Density SNP Markers, Pedigree and GWAS Information: A Case Study with the Non-Model Species Eucalyptus cladocalyx. Plants, 9(1), 99.

6.Brettschneider, R. (1998). RFLP analysis. In Molecular tools for screening biodiversity (pp. 85-95). Springer, Dordrecht.

7. Budowle, B., Daal, A. van, Budowle, B., Budowle, B., Daal, A. van, Sensabaugh, Faculty of Health Science & Medicine. (2018, May 16). Forensically relevant SNP classes. Retrieved from https://www.future-science.com/doi/10.2144/000112806.

8. Cross, D. S., Ivacic, L. C., Stefanski, E. L., & McCarty, C. A. (2010). Population-based allele frequencies of disease-associated polymorphisms in the Personalized Medicine Research Project. BMC genetics, 11, 51. https://doi.org/10.1186/1471-2156-11-51

9. Dean, L. (2018). Clopidogrel therapy and CYP2C19 genotype. In Medical Genetics Summaries [Internet]. National Center for Biotechnology Information (US).

10.Dehbozorgi, M., Kamalidehghan, B., Hosseini, I., Dehghanfard, Z., Sangtarash, M. H., Firoozi, M., Ahmadipour, F., Meng, G. Y., &Houshmand, M. (2018). Prevalence of the CYP2C19*2 (681 G>A), *3 (636 G>A) and *17 (806 C>T) alleles among an Iranian

population of different ethnicities. Molecular medicine reports, 17(3), 4195–4202. https://doi.org/10.3892/mmr.2018.8377

11.Dervieux, T., &Bala, M. V. (2006). Overview of the pharmacoeconomics of pharmacogenetics.

12.Gaudet M., Fara AG., Beritognolo I., Sabatti M. (2009).Allele-Specific PCR in SNP Genotyping. In: Komar A. (eds) Single Nucleotide Polymorphisms. Methods in Molecular Biology™ (Methods and Protocols), vol 578. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-60327-411-1_26

13.Ghatak, S., Muthukumaran, R. B., &Nachimuthu, S. K. (2013). A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. Journal of biomolecular techniques: JBT, 24(4), 224–231. https://doi.org/10.7171/jbt.13-2404-001

14.Gray, I. C., Campbell, D. A., &Spurr, N. K. (2000). Single nucleotide polymorphisms as tools in human genetics. Human molecular genetics, 9(16), 2403-2408.

15.HerreroMaría José, Rojas, L., Bea, S., Plumed Jaime Sánchez, Hernández Julio, Poveda, J. L., &Aliño Salvador F. (2013, February 13). Practical Pharmacogenetics and Single Nucleotide Polymorphisms (SNPs) in Renal Transplantation. Retrieved from https://www.intechopen.com/books/current-issues-and-future-direction-in-kidney-transplantation/practical-pharmacogenetics-and-single-nucleotide-polymorphisms-snps-in-renal-transplantation.

16.Jiang, X. L., Samant, S., Lesko, L. J., & Schmidt, S. (2015). Clinical pharmacokinetics and pharmacodynamics of clopidogrel. Clinical pharmacokinetics, 54(2), 147-166.

17.Karki, R., Pandya, D., Elston, R. C., &Ferlini, C. (2015). Defining "mutation" and "polymorphism" in the era of personal genomics. BMC medical genomics, 8, 37. https://doi.org/10.1186/s12920-015-0115-z

18.Kim, S., &Misra, A. (2007). SNP genotyping: technologies and biomedical applications. Annu. Rev. Biomed. Eng., 9, 289-320.

19. LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic acids research, 37(13), 4181-4193.

20.Lee, S. J. (2013). Clinical application of CYP2C19 pharmacogenetics toward more personalized medicine. Frontiers in genetics, 3, 318.

21. Lee. (2012, December 20). Clinical Application of CYP2C19 Pharmacogenetics Toward More Personalized Medicine. Retrieved from https://www.frontiersin.org/articles/10.3389/fgene.2012.00318/full.

22. Ma, T. K., Lam, Y. Y., Tan, V. P., & Yan, B. P. (2011). Variability in response to clopidogrel: how important are pharmacogenetics and drug interactions?. British journal of clinical pharmacology, 72(4), 697-706.

23.Maitland-van der Zee, A. H., Puckett, B. J., Terra, S. G., & Walker, J. (2003). SNP s: Single Nucleotide Polymorphisms and Pharmacogenomics: Individually Designed Drug Therapy. Burger's Medicinal Chemistry and Drug Discovery, 181-204.

24.Mark G. Papich,(2016).Clopidogrel, Editor(s): Mark G. Papich, Saunders Handbook of Veterinary Drugs (Fourth Edition), W.B. Saunders,2016, Pages 176-178, ISBN 9780323244855.

25.Mirabbasi, S. A., Khalighi, K., Wu, Y., Walker, S., Khalighi, B., Fan, W., Kodali, A., & Cheng, G. (2017). CYP2C19 genetic variation and individualized clopidogrel prescription in a cardiology clinic. Journal of community hospital internal medicine perspectives, 7(3), 151–156. https://doi.org/10.1080/20009666.2017.1347475

26.Morgan, L. (2009). Haplotype mapping in human disease. The Obstetrician &Gynaecologist, 11(4), 277–283. doi:10.1576/toag.11.4.277.27532

27.Pinglang Wang, Manhong Dai, Weijian Xuan, Richard C. McEachin, Anne U.Jackson, Laura J. Scott, Brian Athey, Stanley J. Watson, Fan Meng,(2006).SNP Function Portal: a web database for exploring the functional implication of SNP alleles, Bioinformatics, Volume 22, Issue 14, 15 July 2006, Pages e523–e529.

28.S. Bhattacharyya, P. Ducheyne,(2011).3.327 - Intracellular Probes, Editor(s): Paul Ducheyne, Comprehensive Biomaterials, Elsevier, 2011, Pages 447-470, ISBN 9780080552941.

29.Shastry, B. S. (2007, October 11). SNPs in disease gene mapping, medicinal drug development and evolution. Retrieved from https://www.nature.com/articles/jhg2007117.

30.Shen GQ., Abdullah K.G., Wang Q.K. (2009) The TaqMan Method for SNP Genotyping. In: Komar A. (eds) Single Nucleotide Polymorphisms. Methods in Molecular Biology™ (Methods and Protocols), vol 578. Humana Press, Totowa, NJ.

31.Srilakshmi Srinivasan, Jyotsna Batra,(2019).Single Nucleotide Polymorphism Typing, Editor(s): Shoba Ranganathan, Michael Gribskov, Kenta Nakai, Christian Schönbach, Encyclopedia of Bioinformatics and Computational Biology, Academic Press,2019, Pages 432-440, ISBN 9780128114322,

32.Tahira, Tomoko &Kukita, Yoji &Higasa, Koichiro& Okazaki, Yuko &Yoshinaga, Aki & Hayashi, Kenshi. (2009). Estimation of SNP allele frequencies by SSCP analysis of pooled DNA. Method in molecular biology (Clifon, N. J.). 578. 193-207. 10.1007/978-1-60327-411-1_12.

33.Undi, R., Lim, H. Y., & Wang, W. (2019). Rapid and reliable identification of insulin 2 gene mutation in Akita diabetic mice by a tetra-primer-ARMS-PCR method. Heliyon, 5(1), e01112.

34. Woodward J. (2014). Bi-Allelic SNP Genotyping Using the TaqMan® Assay. In: Fleury D., Whitford R. (eds) Crop Breeding. Methods in Molecular Biology (Methods and Protocols), vol 1145. Humana Press, New York, NY

35.You Frank, M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y. Q., Hane, D., ... & Anderson, O. D. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinforma, 9, 253.

36.Zhang G, Nebert DW,(2017).Personalized medicine: Genetic risk prediction of drug response. Pharmacology & Therapeutics. 2017 Jul;175:75-90. DOI: 10.1016/j.pharmthera.2017.02.036.