Introducing Single Nucleotide variations into the Vitamin C Transporter SLC23A1

By: Rehab Alharbi

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba

In partial fulfillment of the requirements of the degree of
MASTER OF SCIENCE
Department of Human Nutritional Sciences
University of Manitoba
Winnipeg

Abstract

Background. The solute carrier SLC23A1, encoded by the *SLC23A1* gene, is the key membrane transporter mediating the reabsorption of ascorbic acid, the reduced form of vitamin C, in the renal proximal tubule. It had been shown that a single nucleotide polymorphism (SNP) in *SLC23A1* could alter the proteins ability to transport ascorbic acid. Recently, through various human re-sequencing projects, many novel single nucleotide variations (SNV) have been discovered and their information has been deposited in genomic databases. However, their impact on the proteins functions remains to be determined. In a first step for functional studies, SNV need to be introduced into an existing *SLC23A1* expression plasmids. The objective of this study was to introduce selected SNVs into the coding region, or Open Reading Frame (ORF), of the *SLC23A1* expression plasmid pReceiver-M55-*SLC23A1*.

Methodology. In the presented project Polymerase Chain Reaction (PCR) based site directed mutagenesis was chosen to introduce SNVs into the expression plasmid. Here, three slightly distinct methods were used, all based on amplification with the high fidelity polymerases Phusion™ and Q5™: a) amplification of the whole plasmid using overlapping and complementary primers; b) "two step" amplification consistent of the amplifications of separate mutated 5' and 3' pieces of the ORF and subsequent ligation; c) introduction of variations proximal to the 5' and 3' ends within one amplicon.

Results. Attempts to amplify around the plasmid never yielded the expected amplicons, even when a variety of polymerases and conditions were tested. In the two step approach, all 5' and 3' amplicons carrying variations of the *SLC23A1* ORF were achieved. However, the second step PCR to ligate these to the full length *SLC23A1* ORF failed with

Phusion™ and Q5™, and even regular Taq polymerase. Six SNVs located proximal to the 5' and 3' ends of the *SLC23A1* ORF were successfully introduced.

Conclusions. It was proven difficult to introduce site directed mutations in the middle of the *SLC23A1* ORF. Possible reasons are discussed. Variations proximal to the 5' and 3' end had been introduced. These amplicons form the basis for future functional studies.

Outlook. Functional experiments will elucidate the impact of these variations on the transporters function. Various variations in SLC23A1 are predicted to abrogate or severely impact on its functions and projects like these will aid the prioritization of variations to be investigated for an impact on vitamin C metabolism.

Dedication

Dreams do come true.

I dedicate this thesis to my beloved family for their constant support through my life journey.

To my deceased uncle,

The magazine that you brought to the 10-year-old girl each Wednesday over 15 years made the women today who can write this thesis.

You will always be alive in my heart.

Table of contents

Table of Figures

List of Tables

1- Background and Literature Review

Vitamin C is an essential micronutrient that functions as an antioxidant, as a cofactor in eight enzymatic reactions and as an agent in collagen synthesis. The human body cannot synthesize vitamin C due to the lack of gulonolactone oxidase, a required enzyme for vitamin C synthesis (Nishikimi, Fukuyama, Minoshima, Shimizu, & Yagi, 1994). Therefore, it has to be obtained from dietary sources.

Ninety percent of vitamin C in the human diet comes from fresh fruits and vegetables (Sinha, Block, & Taylor, 1993). High sources of vitamin C includes citrus fruits, guava, red and yellow peppers and red cabbage (Szeto, Tomlinson, & Benzie, 2002). Vitamin C is also commonly consumed in a form of supplements in multivitamin preparations or alone.

The current (2000) Recommended Dietary Allowance (RDA) was set at 90 mg/day for adult non-smoking males and 75 mg/day for adult non-smoking females, with an additional 35mg/day for smokers (Institute of Medicine Panel on Dietary & Related, 2000). The RDA recommendations were set at these levels because they were identified as the adequate amounts to reach the saturated levels of vitamin C in the blood (Monsen, 2000). These recommendations were regarded as adequate to prevent the vitamin C deficiency disease scurvy and to prevent other common and more complex diseases caused by suboptimal status (Levine, Padayatty, & Espey, 2011). The RDA values are set based on pharmacokinetics studies on small groups of healthy young adult Caucasians who reached plasma ascorbic acid levels of 50-55 μM when consuming these daily recommended amounts (Padayatty & Levine, 2001). Excess amount of vitamin C is

excreted in the urine. Based on the adverse effect of gastrointestinal disturbance the tolerable upper intake for adults is set at 2 g/day (Institute of Medicine Panel on Dietary & Related, 2000).

Vitamin C is transported in the human body in two forms, the reduced form ascorbic acid and the oxidized form dehydroascorbic acid (DHA) (Nishikimi et al., 1994). The transport of the oxidized dehydroascorbic acid is mediated by facilitative glucose transporters of the *SLC2A*/GLUT family, while reduced ascorbic acid is transported by the two sodium-dependent solute carries SLC23A1 and SLC23A2, encoded by the *SLC23A1* and *SLC23A2* genes (Rajan et al., 1999); (Eck et al., 2013). *SLC23A1* is expressed in the epithelial tissues mainly liver, kidneys and intestines (Hiroyasu Tsukaguchi1, 1999); (Boyer et al., 2005); (Cahill et al., 2009).
Due to the significant role of SLC23A1 in reabsorbing ascorbic acid, genetic variations in *SLC23A1* could have an impact on vitamin C status in the human body (Corpe et al., 2010a). Emerging evidence implies that even with consumption of daily recommendations of vitamin C, genetic variations may influence plasma ascorbic acid levels (Eck et al., 2013). Inadequate levels of ascorbic acid in cells play a significant role in the development of common diseases such as cardiovascular diseases and some cancers (Gupta et al., 2012).

1.1 Genetic Variations

Genetic variations are random, permanent changes in the DNA sequence (Sherry et al., 2001). The most prevalent genetic variations in the human genome are Single Nucleotide Polymorphism (SNPs), which are common and found in a minor allele frequency of more than 1% ("Genetics Home Reference," 2014). If a single nucleotide

variation occurs with a minor allele frequency of less than 1% it is considered rare and called a Single Nucleotide Variations (SNVs) (Lander1, 2001); (Becker, Barnes, Bright, & Wang, 2004; "Initial sequencing and analysis of the human genome," 2001)

Besides single nucleotide substitutions, insertions or deletions of one to fifty bases are frequent, while larger structural variations are less frequent. These events, however, are not further explained, since this project focuses on single nucleotide substitutions.

1.2 Types of Single Nucleotide Variations

Single nucleotide substitutions are by far the most frequent type of genetic variations, with a frequency of one in a thousand bases. Most of the time they have no functional consequence ("Genetics Home Reference," 2014). However, an exchange of one base for another can result in an altered codon changing an amino acid, called missense mutation. Potentially more dramatic effects on a protein should be caused when a base change introduces a premature stop codon, called a gain of stop codon.

Insertions or deletions of bases are seen less frequently; however, when they fall into the coding sequence they will cause a shift of the subsequent codons, called a frameshift. Consequently the amino acid sequence of the protein is altered and most often terminated shortly after the frameshift, most likely causing dramatic changes to the proteins functions.

1.3 The Solute Carrier Family 23 member A1 (SLC23A1) gene

The *SLC23A1* gene encodes a sodium-dependent ascorbic acid membrane transporter (P. Eck et al., 2004). The transcript was cloned for the first time in 1999 from rat kidneys complementary DNA libraries, and it was initially named sodium-dependent

vitamin C transporter SVCT1(Hiroyasu Tsukaguchi1, 1999; Tsukaguchi et al., 1999); (H. C. Erichsen, Eck, Levine, & Chanock, 2001; Hiroyasu Tsukaguchi1, 1999) It is now agreed upon that the protein is also referred to as SLC23A1, and it is required for intestinal absorption and renal reabsorption of ascorbic acid. *SLC23A1* is also expressed in the skin, liver and lungs, where its role in vitamin C metabolism is less clear (Savini, Rossi, Pierro, Avigliano, & Catani, 2008).

Although the origin of the SVCT1 transcript was prior assigned to the *SLC23A2* locus, it now corresponds to the *SLC23A1* gene, which is located on chromosome 5q31.2–31.3 (Figure 1). It extends from bases 139,367,196 to 139,384,943 spanning 17,748 bases and is transcribed from the minus strand as (http://www.ncbi.nlm.nih.gov/, 2015).

Two major transcripts (NM_005847.4 and NM_152685.3) have been recorded in GeneBank, both consisting of 15 exons (Figure 2), where NM_152685.3 encodes a longer and non-functional protein isoform with 4 additional amino acids. The major functional isoform NM_005847.4 encodes for a 598 containing protein. An alternative first exon was reported, which encodes for a longer protein isoform with severely reduced function (Amir & Eck, 2013). Therefore it is currently conclude that the SLC23A1 gene locus contains 16 exons (Amir & Eck, 2013).

Figure 1: The SLC23A1 gene located on the long "q arm" of chromosome 5 at region3 band number1 and sub-band number 2, 5q31.2 (adopted from gene cards).

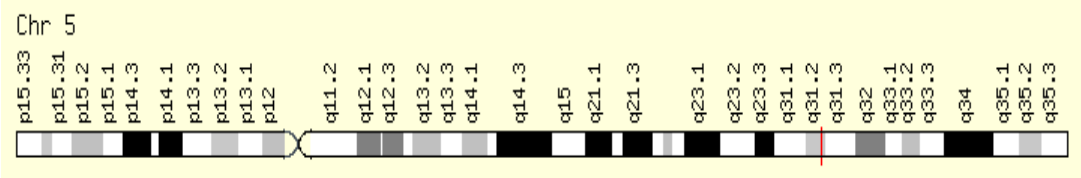


Figure 2: Exon intron structure of the two major transcripts for the human SLC23A1 gene as recorded by GeneBank (http://www.ncbi.nlm.nih.gov/)

Green boxes represent the exons, lines representing introns. Dark green indicates the coding region, light green the 5' and 3' untranslated regions.

1.4 Single Nucleotide Polymorphisms in the human *SLC23A1* gene are associated with common and complex diseases

Polymorphism in the *SLC23A1* gene are associated with the risk of preterm birth (Erichsen et al., 2006), and advanced colorectal adenoma (Hans Christian Erichsen et al., 2008). Moreover, Skibola et al. (2008) reported strong associations between polymorphism in vitamin C transporters and follicular lymphoma. Specifically, 80% elevated risk of lymphoma was associated with the rs6596473-C allele and rs11950646-GC heterozygosity (Skibola et al., 2008).

Polymorphisms in *SLC23A1* are also associated with Inflammatory Bowel Diseases. A recent case control study of 311 patients and 142 controls reported an allelic dosage effect for the SNP rs10063949-G allele, where AG heterozygosity elevated the risk for Crohn's disease by 2.5-fold, while GG homozygotes showed a 4.7-fold risk elevation (Amir Shaghaghi, Bernstein, Serrano Leon, El-Gabalawy, & Eck, 2014).

1.5 Single Nucleotide Polymorphisms in the human *SLC23A1gene* are associated with plasma ascorbate levels and complex diseases

Several human studies and *in vitro* and *in vivo* experiments suggest that genetics variations in *SLC23A1* play a role in ascorbic acid status in the plasma (Cahill & El-Sohemy, 2009; Timpson et al., 2010; Wang et al., 1999).

One study conducted on the *Xenopus laevis* oocyte system showed that ascorbic acid transport in oocytes injected with complementary RNA coding for the *SLC23A1 rs35817838-G* variant (called SNP 772G in the cited study) reduced the ascorbic acid uptake by over 80% (Corpe et al., 2010a). In contrast, the three other variations rs33972313-A, rs6886922-T, and rs34521685-G did not reduce the ascorbic acid uptake dramatically. It was predicted, that SNP rs35817838-G could cause a 75% decline in plasma ascorbate concentrations, while SNPs rs33972313-A, rs6886922-T, and rs34521685-G might decline the ascorbic acid level in plasma by 40% to 50% (Corpe et al., 2010a).

In the same study, a knock out mouse model was designed to investigate the effect of the absence of *slc23a1* on ascorbate plasma level. It was found that the absent of *slc23a1* does not impair the absorption of vitamin C, however, the ascorbate plasma levels were substantially decreased due to the lack of renal reabsorption and increased levels of vitamin C excretion in the urine (Corpe et al., 2010a). In the same knock out mouse model, liters from the female $slc23a1^{-/-}$ had a prenatal mortality rate of approximately 50%. This rate decreased to 15% with the use of vitamin C supplements (Corpe et al., 2010a).

Altogether, Corpe et al. (2010) established the proof that variation in the human *SLC23A1* gene impact on the transport capacity, and that the elimination of *slc23a1* in a

mouse model results in a disease phenotype caused by lowered plasma ascorbic acid levels.

Furthermore, human association studies were performed to investigate the link between the genetics polymorphism in *SLC23A1* and ascorbate plasma levels. In a meta-analysis of five studies from different populations including 15,087 participants it was found that the genotype SNP rs33972313-A negatively associated with circulating concentrations of ascorbic acid. Per allele of rs33972313-A plasma ascorbate level were lowered by 6 μM, regardless of the effect of other factors; such as vitamin C intake, alcohol consumption, smoking status, social status and geographical location (Timpson et al., 2010) (Duell et al., 2013).

These studies establish that genetic variations in *SLC23A1* can impact the transport function, plasma ascorbic acid levels, and can lead to disease associated phenotypes.

1.6 Single Nucleotide Variations (SNVs) reported for the SLC23A1 gene

Single Nucleotide Variations (SNVs) are rare variations, while SNPs are common with frequencies exceeding 1% on a population basis. During the initial years of the discoveries of human genetic variations many SNPs were revealed, however, with the recent progress of genome sequencing in larger population cohorts many SNVs had been discovered.

In the coding region of the human *SLC23A1* gene SNVs represent the majority of missense variations, outnumbering SNPs by 90%. SNVs are the types of variants considered to have the higher impact on genes functions, due to very low occurrence rates associated with higher penetrance. All SNPs coding for missense variations in the *SLC23A1* gene had been tested in the *Xenopus laevis* oocyte system, and the rs35817838-

G variant (772G) reduced the ascorbic acid uptake by over 80% (Corpe et al., 2010). Due to this proof of principal, it can be theorized, that SNV in the gene have similar or even more detrimental impacts.

SNVs falling into the *SLC23A1* locus could be obtained from the ENSEMBL database (ENSMBL, 2014). These data were collected from different human genome projects namely the human genome project and the 1000 genomes project (Lander et al, 2001), the NHLBI Exome Sequencing Project (ESP) (http://evs.gs.washington.edu/EVS/) and the Cancer Genome Project (http://www.sanger.ac.uk). Table 1 lists missense SNVs located in the SLC23A1 protein-coding region, of which 54 had been reported at the time of the initiation of the project in 2014.

A variety of bioinformatics programs had been developed to predict the impact of missense mutations on a proteins functions. the results for two of these prediction programs, SIFT (sorting the intolerant from the tolerant) and Polymorphism Phenotyping (polyphen) are included in table 1 to demonstrate current discrepancies in prediction methods (Adzhubei et al., 2010; Kumar, Henikoff, & Ng, 2009; Ng & Henikoff, 2003). This enforces the rationale that each missense variation needs to be tested in order to determine its functional impact. This precipitates the need for protein engineering via site directed mutagenesis, a molecular technique that involves changing one or more bases of DNA sequence to examine its impact on the gene function (Ishii et al., 1998). For this project, we aims to create intentional changes in the DNA sequence, which will mimic the change a SNV causes.

Table 1: SNPs and SNVs in the coding region of SLC23A1s, with predicted impact on the transport function. Each SNV has a unique ID consists of letters and numbers. SNVs begining with (COSM) are referred to SNVs from the Catalogue of Somatic Mutation in Cancer. SNVs and SNPs with rs-numbers were obtained from the ENSEMBL database. SNVs begin with TMP_ESP were found in the NHLBI Exom Sequencing Project and they were temporarily identified with TMP and later assigned to rs-numbers. All variations were collectively obtained from the ENSMBL database in spring of 2014.

SNV-ID	Their location on the chromosome	Amino acid #	Amino Acid change	Predicted Impact on the protein	
				SIFT	PolyPhen
COSM117494	5:138718925	6	D/E	Tolerated	Benign
rs138079930	5:138718921	8	E/K	Tolerated	Benign
TMP_ESP_5_138718915	5:138718915	10	R/W	Deleterious	Benign
TMP_ESP_5_138718909	5:138718909	12	Q/E	Tolerated	Benign
rs200503816	5:138718294	13	H/Y	Deleterious	Benign
TMP_ESP_5_138718289	5:138718289	14	E/D	Tolerated	Benign
rs143475469	5:138718277	18	D/E	Tolerated	Benign
rs182436678	5:138718272	20	S/L	Tolerated	Benign
rs199767403	5:138718266	22	P/L	Deleterious	Benign
rs114335293	5:138718219	38	V/M	Deleterious	Possibly damaging
TMP_ESP_5_138718203	5:138718203	43	L/Q	Tolerated	Probably damaging
TMP_ESP_5_138717707	5:138717707	61	A/V	Tolerated	Probably damaging

SNV-ID	Their location on the chromosome	Amino acid #	Amino Acid change	Predicted Impact on the protein	
				SIFT	PolyPhen
TMP_ESP_5_138717683	5:138717683	69	A/V	Deleterious	Probably damaging
COSM204877	5:138717681	70	L/M	Tolerated	Benign
TMP_ESP_5_138717666	5:138717666	75	D/N	Tolerated	Probably damaging
COSM308150	5:138717662	76	Q/L	Tolerated	Probably damaging
rs200631596	5:138717627	88	T/A	Deleterious	Benign
rs115153492	5:138716310	137	G/S	Tolerated	Benign
rs199550338	5:138716280	147	H/Y	Deleterious	Probably damaging
rs201746992	5:138716264	152	R/Q	Deleterious	Possibly damaging
rs142534689	5:138716250	157	G/R	Tolerated	Benign
COSM20486	5:138716072	158	G/S	Deleterious	Probably damaging
TMP_ESP_5_138716246	5:138716246	158	L/W	Deleterious	Benign
rs151090514	5:138716057	163	S/P	Deleterious	Probably damaging
rs187605953	5:138716052	164	S/R	Deleterious	Probably damaging
rs34521685	5:138715761	218	I/V	Tolerated	Possibly

SNV-ID	Their location on the chromosome	Amino acid #	Amino Acid change	Predicted Impact on the protein	
				SIFT	PolyPhen
					damaging
rs145776418	5:138715522	257	I/N	Deleterious	Probably damaging
rs35817838	5:138715520	258	M/V	Tolerated	Benign
rs33972313	5:138715502	264	V/M	Tolerated	Possibly damaging
TMP_ESP_5_138715427	5:138715427	289	R/G	Deleterious	Probably damaging
rs151281788	5:138715412	294	G/S	Tolerated	Possibly damaging
TMP_ESP_5_138715400	5:138715400	298	A/T	Tolerated	Benign
rs200925380	5:138714971	333	I/T	Deleterious	Benign
rs141189553	5:138714948	341	A/T	Tolerated	Possibly damaging
TMP_ESP_5_138714941	5:138714941	343	A/V	Deleterious	Probably damaging
rs148214051	5:138714918	351	P/T	Deleterious	Probably damaging
rs115023155	5:138714360	363	E/K	Deleterious	Possibly damaging
rs137952208	5:138714308	380	T/I	Deleterious	Probably damaging

SNV-ID	Their location on the chromosome	Amino acid #	Amino Acid change	Predicted Impact on the protein	
				SIFT	PolyPhen
rs143406252	5:138714309	380	T/P	Deleterious	Benign
TMP_ESP_5_138 714275	5:138714275	391	I/T	Deleterious	Probably damaging
rs6596474	5:138713959	421	S/A	Tolerated	Benign
rs116659253	5:138713708	470	T/M	Tolerated	Benign
TMP_ESP_5_138 713118	5:138713118	508	A/T	Deleterious	Probably damaging
rs72552250	5:138707916	526	Q/K	Tolerated	Benign
rs200264889	5:138707859	545	D/N	Tolerated	Benign
rs201565759	5:138707853	547	P/A	Deleterious	Probably damaging
rs150461696	5:138707831	554	K/R	Tolerated	Benign
rs115102991	5:138707700	598	V/L	Deleterious	Probably damaging
rs11242462	5:138720108	45	C/T	Tolerated	Benign
rs141167635	5:138715928	206	G/A	Tolerated	Benign
rs139127139	5:138715037	311	C/T	Tolerated	Benign
TMP_ESP_5_138 707912	5:138707912	527	C/T	Deleterious	Benign
rs34054084	5:138715501 -138715500	264	-/C	Tolerated	Benign
COSM217333	5:138713966 -138713965	423	-/C	Deleterious	Benign

2- Hypothesis and Objectives

Many missense mutations in a gene have no functional impact. However, it has already been shown that the SNP rs35817838-G allele reduces SLC23A1's transport function (Corpe et al., 2010). Current bioinformatics methodologies to predict the impact of a missense mutation on the proteins function are imperfect. About fifty missense variations in the *SLC23A1* gene are deposited in genetic databases, of which only 4 had been tested if they change the proteins function. However, mounting evidence for disease associations in the *SLC23A1* gene create a rationale to test all of these single nucleotide variations in order to identify disease causing mutations. It is hypothesised that variations in the *SLC23A1* gene will change the transport function, which can be tested in the *Xenopus laevis* oocyte expression system.

In order to test their functional impact, the variations need to be introduced into a suitable expression plasmid. This pilot study aims to introduce selected single nucleotide variations into the pReceiver-M55-*SLC23A1* expression plasmid, which carries a copy of the transcript representing the most common genotypes.

Three PCR based site directed mutagenesis approaches are tested for their suitability within the following sub-objectives:

A) Introducing SNVs into pReceiver-SLC23A1-M55 through amplification of the entire plasmid using overlapping and complementary primers.
B) Introducing SNVs into the middle parts of the *SLC23A1* ORF through the creation of two 5' and the 3' amplicons carrying the variation, followed by ligation of the amplicons and tagging with Gateway® flanking sites.

C) Introducing SNVs in the proximity to the 5' and the 3' ends of the *SLC23A1* ORF through within a single amplicon, followed by tagging with Gateway® flanking sites.

Products derived from sub-objectives B and C would be ready for subcloning into an entry/donor vector and transfer into a large variety of destination expression vectors using the Gateway® technology.

3- Material and Methods

3.1 Selected Single Nucleotide Variations (SNVs)

Seventeen SNVs which had been sourced from the NCBI and ENSMBLE databases (ENSEMBL, 2013; http://www.ncbi.nlm.nih.gov/, 2013) were selected. Of these, twelve were missense, three nonsense, and two frameshift variations (

Table 2). The location of the individual variations within the SLC23A1 protein is depicted in Figure 3.

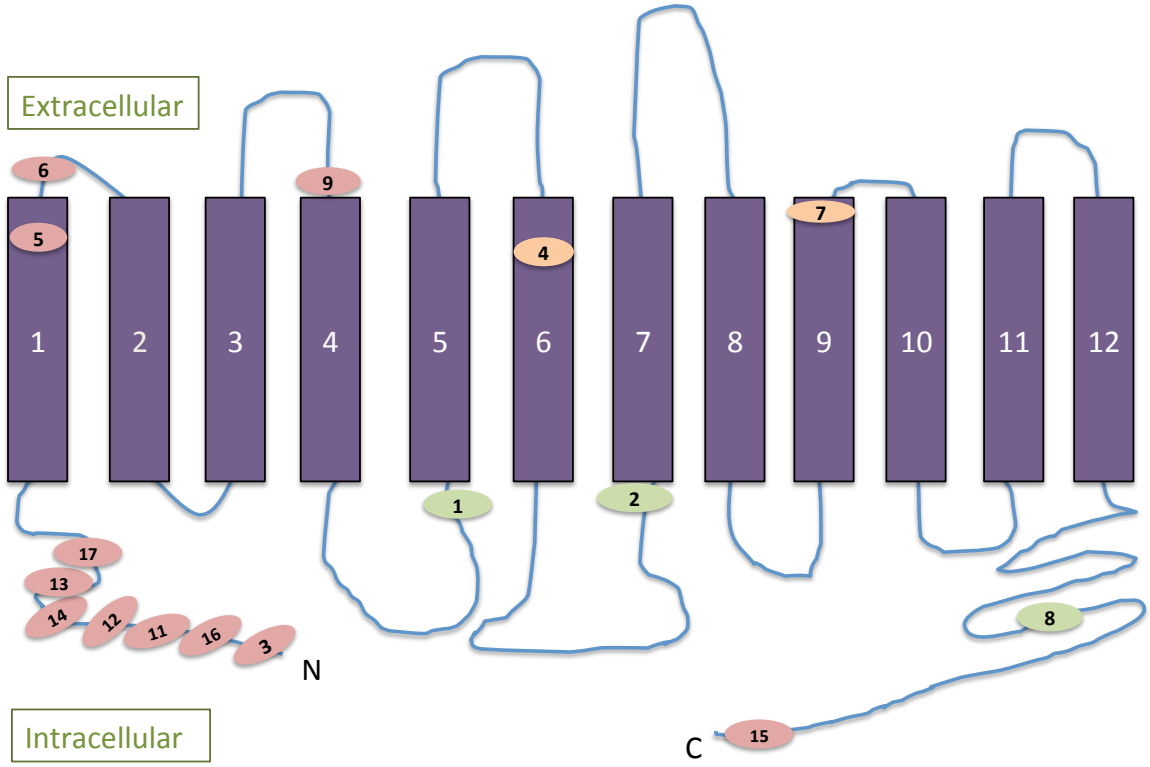
Table 2: Seventeen SNVs chosen to be introduced into the SLC23A1 ORF by sited directed mutagenesis.

SNV-ID	Type of SNV	Location on the chromosome	Exon #	Amino Acid #	Amino acid change	Allele Change	Predicted impact by Sift/ Polyphen
rs141167635	Nonsense	5:138715928	6	206	R/Stop	G/A	Tolerated/ Benign
rs139127139	Nonsense	5:138715037	9	311	W/Stop	C/T	Tolerated / Benign

COSM117494	Missense	5:138718925	1	6	D/E	G/C	Tolerated/ Benign
rs34054084	Frameshift	5:138715501	8	264	___	-/C	Tolerated/ Benign
COSM204877	Missense	5:138717681	3	70	L/M	G/C	Tolerated/ Benign
COSM308150	Missense	5:138717662	3	76	Q/L	T/A	Tolerated/ Probably damaging
COSM217333	Frameshift	5:138713966	11	423	___	-/C	Deleterious/ Benign
TMP_ESP_5_138707912	Nonsense	5:138707912	14	527	W/Stop	C/T	Deleterious/ Benign
COSM20486	Missense	5:138716072	6	158	G/S	C/T	Deleterious/ Probably damaging
rs6596474	Missense	5:138713959	11	421	S/A	G/C	Tolerated/ Benign
TMP_ESP_5_138718915	Missense	5:138718915	1	10	R/W	G/A	Deleterious/ probably damaging
TMP_ESP_5_138718909	Missense	5:138718909	1	12	Q/E	C/G	Tolerated/ Benign
TMP_ESP_5_138718289	Missense	5:138718289	2	14	E/D	A/T	Tolerated/ Benign
rs200503816	Missense	5:139382605	2	13	H/Y	G/A	Tolerated/ Benign
rs115102991	Missense	5:138718272	14	598	V/L	C/G	Deleterious/ probably damaging
rs138079930	Missense	5:138718277	1	8	E/K	C/G	Tolerated/ Benign
rs143475469	Missense	5:138718921	2	18	D/E	C/G	Tolerated/ Benign

Initially, in the year 2013, variations sourced from the Exon Sequencing Project (ESP) were assigned temporary names, indicated by TMP_ESP_5_ (TMP: temporary, ESP: the project, 5: the location on human chromosome 5). Until today, these SNVs were assigned to reference SNP numbers of the rs-type as followed: TMP_ESP_5_138707912 (rs373778813), TMP_ESP_5_138718289 (rs367973515), TMP_ESP_5_138718909 (rs376403265), TMP_ESP_5_138718915 (rs370823646). Moreover, as databases had been curated more dramatic changes could be sees. For example, information about SNV rs143475469 was available in the databases in 2015 and in early 2016; however, information about this SNV is no longer available. Even though this SNV is not in the database, their assay is still available in ThermoFisher company website (http://www.thermofisher.com/order/genome-database/details/pcr-sequencing-primers/Hs00552977_CE).

Figure 3: The location of selected SNVs in the SLC23A1 protein. According to most models SLC23A1 has 12 trans-membranes domains as depicted. SNVs were numbered from 1 to 17 as the following: 1) rs141167635, 2) rs139127139, 3) COSM117494, 4) rs34054084, 5) COSM204877, 6) COSM308150, 7) COSM217333, 8) TMP_ESP_5_138707912, 9) COSM20486, 10) rs6596474, 11) TMP_ESP_5_ 138718915, 12) TMP _ESP_5_138718909, 13) TMP_ESP_5_138718289, 14) rs200503816, 15) rs115102991, 16) rs138079930, 17) rs143475469. SNV types are colored differently; green indicates nonsense, red missense, and orange frameshift variations.



3.2 Site directed mutagenesis

Mutagenesis of genetically encoded proteins is an elegant tool to study the relationship between the protein structure and its function (Ling & Robinson, 1997). The site directed mutagenesis is an *in vitro* method for creating an intentional change in the DNA sequence, which leads to a change in the produced protein from that sequence. In this study PCR based approaches were employed and primers were designed to introduce a specific change in the original sequence of *SLC23A1* ORF. These changes

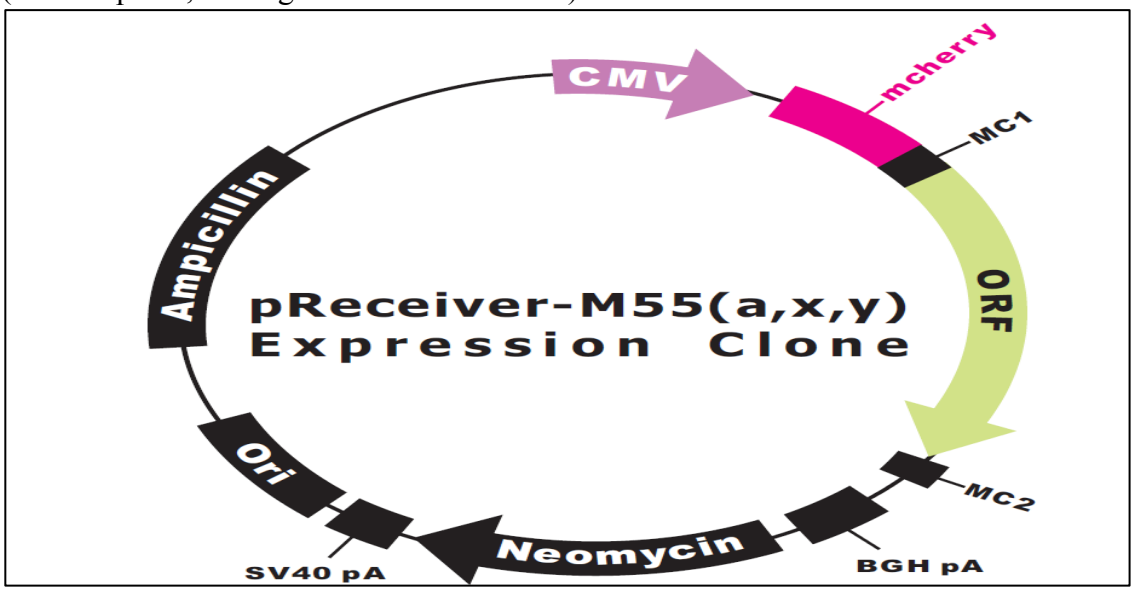
corresponded to known human SNV sequences, which includes addition, deletion or substitution of a nucleotide.

The PCR Template

The functional ancestral *SLC23A1* transcript variant 1 (NM_005847.4; NCBI, 2013) was obtained in the plasmid pReceiver-M55-*SLC23A1* (GeneCopoeia, catalogue # EX-U1235-M55). The plasmid encodes a N-terminal tag of the red fluorescent protein mCherry (Figure 4).

The size of the plasmid without an insert is 6087 nucleotides while the *SLC23A1* ORF is 1797 nucleotides; collectively the size of the whole pReceiver-M55-*SLC23A1* plasmid is 7884 nucleotides. The sequence of the plasmid and the *SLC23A1* ORF can be found in the appendices A and B.

Figure 4: Map of the plasmid pReceiver-M55, which carried the human SLC23A1 ORF (GeneCopoeia, catalogue # EX-U1235-M55)



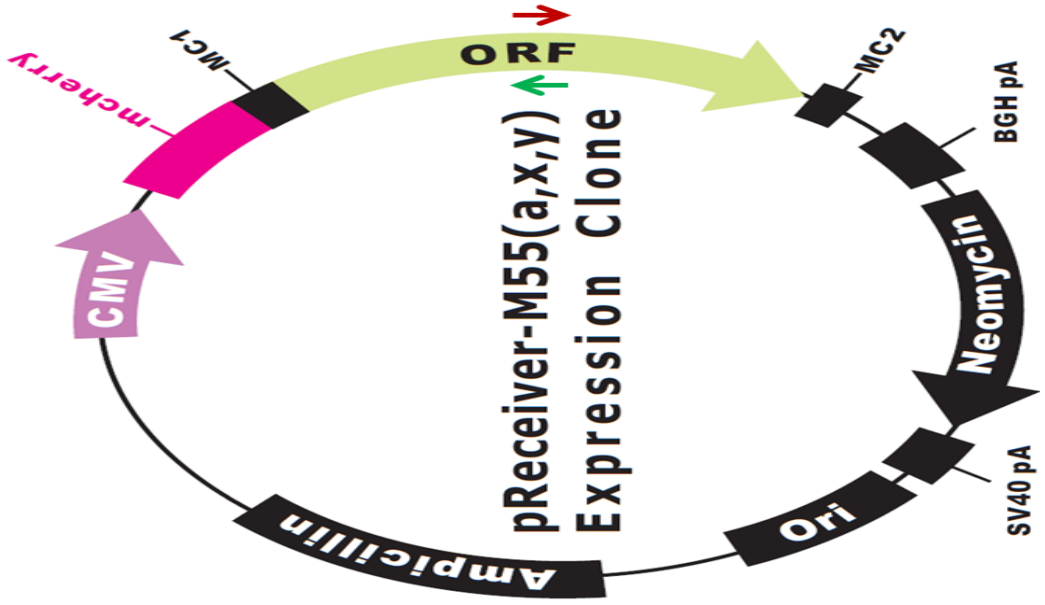
Three PCR based approaches of site directed mutagenesis were followed.

Amplification of the Whole Plasmid

Overlapping complementary primers were designed to attempt the amplification of the entire plasmid. If successful, the linear amplicon with overlapping ends could be ligated to represent the plasmid carrying the site directed mutation. Only one PCR step would be required to introduce the desired changes into the *SLC23A1* ORF. The primers were designed for the PCR be complementary to one another with overlapping primer pairs, one forward and one reverse, each of which was carrying a change corresponding to the sequence of SNVs being introduced into the *SLC23A1* ORF. A scheme is depicted in

Figure 5

Figure 5: Depiction of the approach to introduce mutations into the pReceiver-M55-SLC23A1 plasmid. Overlapping forward (red) and reverse (green) primers introducing the mismatched base were designed to amplify the entire plasmid.



Amplification of the 5' and 3' portions of the ORF and subsequent ligation and tagging for Gateway® subcloning

In order to produce mutated *SLC23A1* ORFs, two separate PCR reactions, each amplifying overlapping 5' and the 3' amplicons of the ORF with primers containing the base-change were conducted (Figure 6). Successfully amplified PCR products resulted from those PCR reactions were diluted at different concentrations ratios 1:100, 1:300 and 1:500. Moreover, a PCR purification step and gel extraction was applied to some PCR products using Quiagen PCR purification and gel extraction kits.

In a second PCR, the self-priming overlapping amplicons were meant to be utilized to amplify the entire *SLC23A1* ORF using primers at the respective 5' and 3' ends, which were flanked with the Gateway® AttB1 and AttB2, respectively. The Gateway® flanking sites will enable subsequent subcloning into any Gateway® DONR plasmid, and later transfer into a huge variety of expression plasmids. The pReceiver-M55-*SLC23A1* plasmid was utilized as the template.

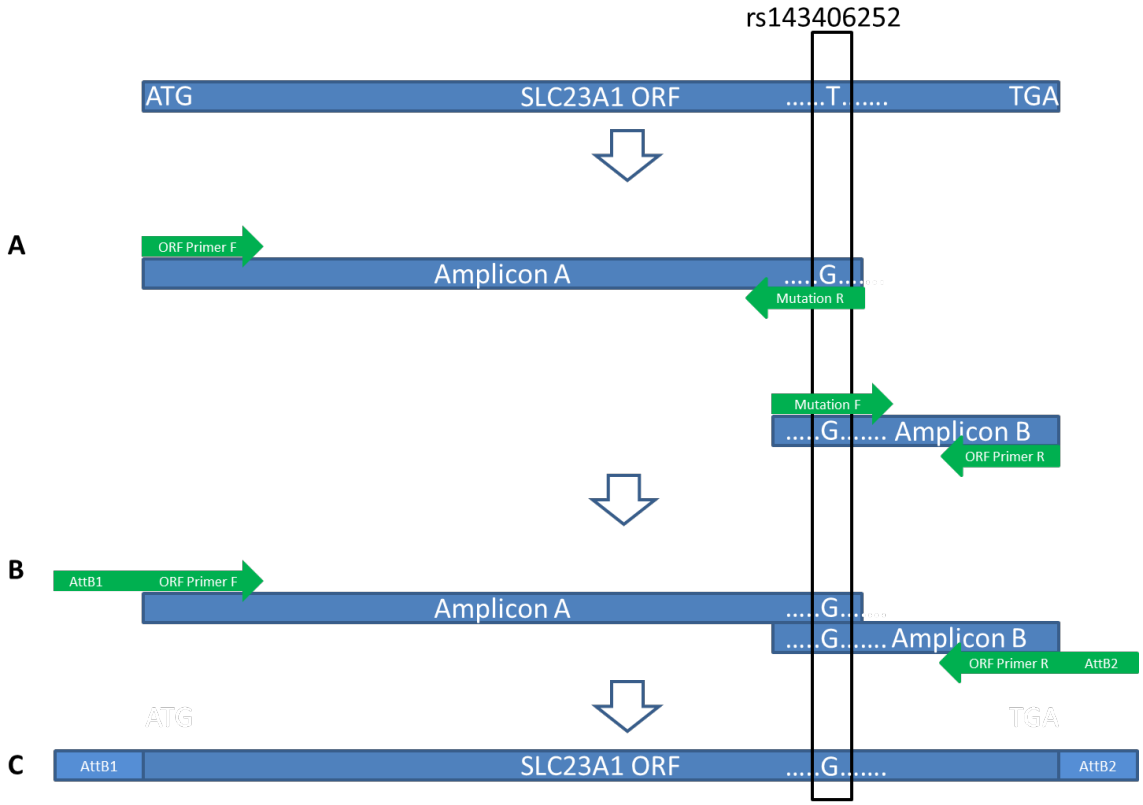
The following terminal *SLC23A1*-ORF primers tagged with Gateway®-AttB sequences were used (*SLC23A1* gene specific sequences are underlined):

5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGGGCCCAGGAGGACC TCGA3'

5'GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGACCTTGGTGCACACAGA TG3'.

Figure 6: Scheme of the site directed mutagenesis approach to introduce a SNV into the middle regions of the SLC23A1 ORF. The example illustrates SNV rs143406252. In the graph the SLC23A1 template and amplicons are depicted in blue, the primers in green. The variation to be introduced is introduced through two primers with identical sequence,

one in the sense (Mutation F, forward) and one in the antisense (Mutation R, reverse) orientation. (A) In two separate PCR reactions two amplicons of the respective 5' and 3' section of the SLC23A1 ORF will be produced by using the Mutation Forward and Mutation Reverse primers paired with, primers framing the ORF (ORF Primer F and ORF Primer R). (B) The two amplicons will be combined in a single PRC reaction to self-prime each other; this reaction also contains the ORF Primer pair, however now tagged with the Gateway® recognition sites AttB1 and AttB2 on the forward and reverse primer, respectively. (C) The final amplicon will contain the mutation and the SLC23A1 ORF flanked by the Gateway® recognition sites AttB1 and AttB2.



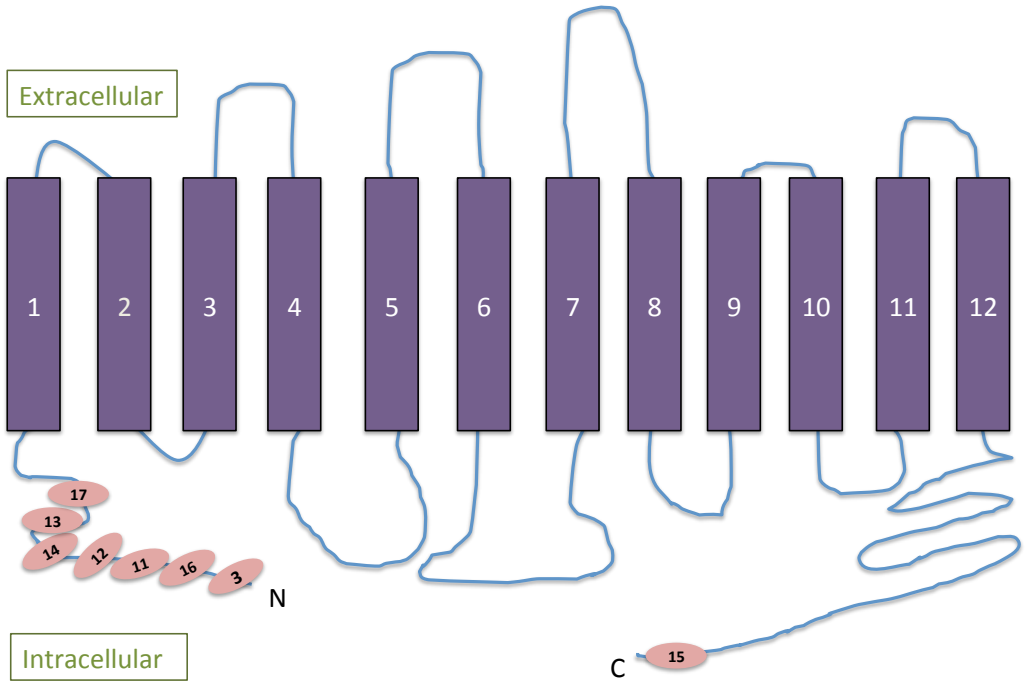
Introduction of selected SNV proximal to the 5' and 3' of the *SLC23A1* ORF through one amplicon

The SNVs chosen for this approach are located very proximal to the 5' and 3' end of the *SLC23A1* ORF (Table 3, Figure 7).

Table 3: SNVs very proximal to the 5' and 3' end of the *SLC23A1* ORF chosen to be introduced via a single amplicon.

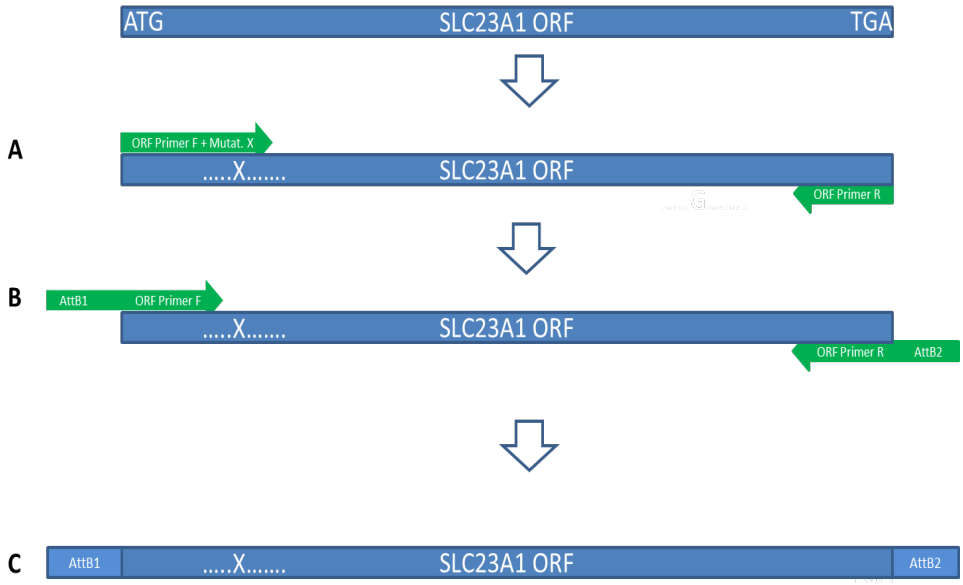
SNV-ID	Type of SNV	Location on the chromosome	Exon #	Amino Acid #	Amino acid change	Allele Change	Predicted impact by Sift/ Polyphen
TMP_ESP_5_138718915	Missense	5:138718915	1	10	R/V	G/A	Deleterious/ probably damaging
TMP_ESP_5_138718909	Missense	5:138718909	1	12	Q/E	C/G	Tolerated/ Benign
TMP_ESP_5_138718289	Missense	5:138718289	2	14	E/D	A/T	Tolerated/ Benign
rs200503816	Missense	5:139382605	2	13	H/Y	G/A	Tolerated/ Benign
COSM117494	Missense	5:138718925	1	6	D/E	G/C	Tolerated/ Benign
rs115102991	Missense	5:138718272	14	598	V/L	C/G	Deleterious/ probably damaging
rs138079930	Missense	5:138718277	1	8	E/K	C/G	Tolerated/ Benign
rs143475469	Missense	5:138718921	2	18	D/E	C/G	Tolerated/ Benign

Figure 7: location of the chosen SNVs on the SLC23A1 trans-membranes. SNV number (3) was used again with different primer combination for this approach.



Therefore, modifications of previously used primers to amplify the *SLC23A1* ORF could be used (Figure). They were engineered to contain the desired mutations. Consequently the mutated ORF can be tagged with the Gateway® recognition sites AttB1 and AttB2 in another round of PCR (Liu & Naismith, 2008).

Figure 8: Scheme of the site directed mutagenesis approach to introduce SNVs into the 5' and 3'regions of the SLC23A1 ORF. In the graph the SLC23A1 template and amplicons are depicted in blue, the primers in green. A) In a first round of PCR the mutation is introduced. The mutation is introduced through a forward or reverse primer, depending on the location in the 5' or 3' respectively. B) In a following round of PCR the SLC23A1 ORF is tagged with the Gateway® recognition sites AttB1 and AttB2 on the forward and reverse primer. C) The final amplicon will contain the mutation and the SLC23A1 ORF flanked by the Gateway® recognition sites AttB1 and AttB2.



3.3 Primers/oligonucleotides

IDT DNA (Integrated DNA Technology, Coralville, Iowa, USA) synthesised the primers utilized in this project. The sequence of the primers is listed in Table and

Table 4. Sequencer 5.1 software was used to design the primers of both the SLC23A1's ORF and the 17 nonsynonymous genetic variations. The primers were designed according to the sequence listed in the ENSEMBL or NCBI databases (ENSEMBL, 2013; http://www.ncbi.nlm.nih.gov/, 2013).

Table 4: Forward and Reverse Primers in the 5' and 3' of the SLC23A1 ORF used to amplify the full ORF and to include the AttB1 and AttR1 Gateway® recognition sites.

Forward primer	**Reverse primers**
SLC23A1AttB F1	**SLC23A1AttB R1**
5'GGGGACAAGTTTGTACAAAAAAGCAGGC TTCATGAGGGCCCAGGAGGACCTCGA3'	5'GGGGACCACTTTGTACAAGAAAGCTGGGTATCAGA CCTTGGTGCACACAGATG3'

Table 4: Forward and Reverse Primers used to introduce the site directed changes into the SLC23A1 ORF. For each SNV (first column) the overlapping forward and reverse primers are listed (second column). The size of the amplicon resulting from the first round of PCR (mutation primer combined with primer at the respective 5' and 3' ORF end) is listed in the third column.

SNV ID	**Primer Sequence (5' → 3')**	**Amplicon Size (bp)**
COSM117494	Forward Primer AGGAGGAGCTCGAGGGCCGG	1810
	Reverse Primer CCGGCCCTCGAGCTCCTCCT	51
COSM204877	Forward Primer TGAGGCGATGTGTGTGGGCC	1620
	Reverse Primer GGCCCACACACATCGCCTCA	241
COSM308150	Forward Primer CCACGACCTGCACATGGTTA	1601
	Reverse Primer TAACCATGTGCAGGTCGTGG	260
COSM20486	Forward Primer GTCCAGAGTGCAATCATGGT	1620

SNV ID	Primer Sequence (5' → 3')	Amplicon Size (bp)
	Reverse Primer ACCATGATTGCACTCTGGAC	241
rs6596474	Forward Primer CCCTCTTCTCCTCGCTCCCT	567
	Reverse Primer AGGGAGCGAGGAGAAGAGGG	1294
rs141167635	Forward Primer GGCGACTGAGCTGGCTCCCA	1212
	Reverse Primer TGGGAGCCAGCTCAGTCGCC	649
rs139127139	Forward Primer TCAGTAGGGCCTGCCCACGG	896
	Reverse Primer CCGTGGGCAGGCCCTACTGA	965
TMP_ESP_5_138707912	Forward Primer TCTGATACAGTAGAAAGCTG	248
	Reverse Primer CAGCTTTCTACTGTATCAGA	1613
RS 34054084	Forward Primer ATCATGACCGTGGTGGCTGC	1036
	Reverse Primer GCAGCCACCACGGTCATGAT	825
COSM217333	Forward Primer GCAAGTTCACGGCCGCTCTT	574
	Reverse Primer AAGAGCGGCCGTGAACTTGC	1287

SNV ID	Primer Sequence (5' → 3')	Amplicon Size (bp)
COSM117494 F1 GTW	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAA AATGAGGGCCCAGGAGGAGCTCGAGG	1831
rs115102991 R1 AttB2	GGGGACCACTTTGTACAAGAAAGCTGGGTATCA GAGCTTGGTGCACACAG	1827
rs138079930 F1 GTW	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAA AATGAGGGCCCAGGAGGACCTCAAGG	1831
rs143475469 F1	ATGAGGGCCCAGGAGGACCTCGAGGGCCGGACA CAGCATGAAACCACCAGGGAGCCCTCG	1797
rs200503816 F1	ATGAGGGCCCAGGAGGACCTCGAGGGCCGGACA CAGTATGAAA	1797
TMP_ESP_5_138718289 F1	ATGAGGGCCCAGGAGGACCTCGAGGGCCGGACA CAGCATGATACCACCAGGGAC	1797
TMP_ESP_5_138718909 F1	ATGAGGGCCCAGGAGGACCTCGAGGGCCGGACA GAGCATGAAA	1797
TMP_ESP_5_138718915 F1	ATGAGGGCCCAGGAGGACCTCGAGGGCTGGACA CAG	1797

3.4 PCR protocols

Three different polymerase enzymes were used for the site directed mutagenesis with respectively different PCR reaction set up and different thermals cycling protocols. Thermal gradients were applied to optimize primer performance.

3.4.1 Phusion™ polymerase high fidelity Polymerase

Phusion™ high fidelity polymerase from New England Biolabs (Ipswich, MA, USA product ID M0530),was used according to the manufacturers' recommendations. A typical 10µl PCR reaction contained the following components:

Component	**10µl Reaction**	**Final Concentration**
Nuclease-free water	4.7µl	____
5XPhusion HF Buffer	2 µl	1X
10mM dNTPs	0.2µl	200µM
10µM Forward Primer	1µl	1µM
10µM Reverse Primer	1µl	1µM
Template DNA, SLC23A1 pRceiver M-55	1µl	1ng
Phusion DNA Polymerase 2 units/µl	0.1 µl	0.2 units/10µl PCR

Typically, the PCR reaction setup was doubled six times in a master mix prepared and aliquoted to assess performance in thermal-gradient conditions.
If applicable, positive controls amplifying the *SLC23A1* ORF with the primers listed in table 3 were included. These primers are proven to reliably yield an amplicon. If applicable, negative controls were included which contained all PCR components except the template.

Thermo-cycling was performed on a Veriti ™ 96-well thermal cycler (Applied Biosystems, Foster City, CA, USA) following the manufacturers recommendations. The thermo-cycling conditions were as follows:

Step	Temperature	Time
Initial Denaturation	98 °C	30 seconds
35 Cycles of A) Denaturation B) Annealing C) Extension	 98 °C 47°C, 52°C, 57°C, 62°C, 67°C, 72°C 72°C	 10 seconds 20 seconds 90 seconds
Final Extension	72°C	5 minutes
Hold	4°C	Forever

3.4.2 Taq DNA Polymerase

Taq™ DNA polymerase from New England Biolabs (Ipswich, MA, USA, product ID M0320) was used according to the manufacturers' recommendations.

A typical 10µl PCR reaction contained the following components:

Component	10 µl reaction	Final Concentration
10X Standard Taq (Mg-free) Reaction Buffer	1.25 µl	1x
25 mM MgCl2	0.75 µl	1.5mM
10 mM dNTPs	0.25 µl	200µM
10 µM Forward Primer	0.25 µl	0.2µM
10 µM Reverse Primer	0.25 µl	0.2µM
Nuclease-free water	6.9 µl	____
Template DNA ,SLC23A1-pRceiver M-55	0.25 µl	1ng
Taq DNA Polymerase enzyme	0.062 µl	0.25 units/ 10µl

Typically, the PCR reaction setup was doubled six times in a master mix prepared and aliquoted to assess performance in thermal-gradient conditions.

If applicable, positive controls amplifying the *SLC23A1* ORF with the primers listed in table 3 were included. These primers are proven to reliably yield an amplicon. If applicable, negative controls were included which contained all PCR components except the template.

Thermo-cycling was performed on a Veriti ™ 96-well thermal cycler (Applied Biosystems, Foster City, CA, USA) following the manufacturers recommendations. The thermo-cycling conditions were as follows:

STEP	TEMP	TIME
Initial Denaturation	95°C	30 seconds
30 Cycles of A) Denaturation B) Annealing C) Extension	 95°C 45°C, 50°C, 55°C, 60°C, 64°C, 68°C 68°C	 30 seconds 30 seconds 90 seconds
Final Extension	68°C	5 minutes
Hold	4°C	Forever

3.4.3 Q5™ Hot Start High-Fidelity DNA Polymerase

Q5™ high fidelity polymerase from New England Biolabs (Ipswich, MA, USA, product ID M0491) was used according to the manufacturers' recommendations.

PCR mixture (10µl) contains contained the following components

Component	**10 µl Reaction**	**Final Concentration**
Nuclease-free water	6.2 µl	____
5X Q5 Reaction Buffer	2 µl	1X
10 mM dNTPs	0.2 µl	200 µM
10 µM Forward Primer	0.5µl	0.5 µM
10 µM Reverse Primer	0.5 µl	0.5 µM

Template DNA, SLC23A1-pRceiver M-55	0.5	1ng
Phusion DNA Polymerase 2 units/µl	0.1 µl	0.02 U/µl

If applicable, positive controls amplifying the *SLC23A1* ORF with the primers listed in Table 4 were included. These primers are proven to reliably yield an amplicon. If applicable, negative controls were included which contained all PCR components except the template.

Thermo-cycling was performed on a Veriti^{TM} 96-well thermal cycler (Applied Biosystems, Foster City, CA, USA) following the manufacturers recommendations. The thermo-cycling conditions were as follows:

Step	**Temperature**	**Time**
Initial Denaturation	98 °C	30 seconds
35 Cycles of A) Denaturation B) Annealing C) Extension	 98 °C 47°C, 52°C, 57°C, 62°C, 67°C, 72°C 72°C	 10 seconds 20 seconds 90 seconds
Final Extension	72°C	5 minutes
Hold	4°C	Forever

3.5 Visualizing the amplicons on a DNA Gel and Imaging

Amplicon sizes were assessed by gel electrophoresis using 1.0% TAE agarose gels, where DNA was stained using 10mg/ml ethidium bromide solution (Bio-Rad Laboratories, Richmond, CA) included in the gel. Individual PCR samples were loaded on the gel after addition of 1µl Gel Loading Dye (blue and purple (6X) New England Biolabs, Ipswich, MA, USA) to 5µl of PCR reaction solution.

Imaging was performed in a Gel Doc™ XR+ Gel Documentation System (Bio-Rad Laboratories, Richmond, CA) and Red™ from Protein Simple (previously knowing as Cell Bioscience, San Jose, CA, USA). Amplicons were compared to the known DNA standard 1 kilobases (kb) ladder from New England Biolabs (Ipswich, MA, USA, product ID N3232S). The scale is ranging from 0.5–10.0 kilobases. It was diluted according to the manufacture's measures; the loading mixture was as the following: nuclease-free water 4µl, 6X Blue or purple Loading Dye 1µl and DNA Ladder 1µl to reach the total volume of 6µl. It is important to note that 5µl of 1kb was added in each used well in the gel.

4. Results

4.1. Objective A: Introducing SNVs into pReceiver-M55-*SLC23A1* through amplification of the entire plasmid.

Ten SNVs were chosen to be introduced into the pReceiver-SLC23A1-M55 plasmid (

Figure 5). This represents the quickest approach to yield a mutated expression plasmid tagged with the red fluorescent protein mCherry. The following SNV were selected for this approach:

1. RS 14116735 is nonsense SNV that changes the nucleotide base G to A causing an alteration in amino acid number 206 arginine to a stop codon (R 206 stop).
2. RS 139127139 is nonsense SNV where the nucleotide changed from C to T, which signal for a stop codon at acid number 311 tryptophan (W 311 stop).
3. COSM 117494 is missense SNV that substitutes the nucleotide G with C, which leads to change in the amino acid from aspartic acid at position 6 to glutamic acid (D 6 E). It is located on exon 1 on the SLC23A1 gene.
4. Rs34054084 is a Frameshift SNV located on exon 8 on the SLC23A1 gene, an extra nucleotide C is inserted causing a change at amino acid number 264.
5. COSM 204877 is a missense SNV where the amino acid at position 70 changed from leucine to methionin (L 70 M) as a result of substituting nucleotide G with C. It is located on exon 3 on the SLC23A1 gene.
6. COSM30850 is a missense SNV that changes the amino acid number 76 from glutamine to leucine (Q 76 L). It is located on exon 3 on the SLC23A1 gene. The nucleotide changed from T to A.

7. COSM217333 is a Frameshift SNV, located on exon 11 on the SLC23A1 gene, at amino acid number 423. An extra nucleotide C is added.
8. Temp138707912 (rs373778813) is nonsense SNV, causing a gain stop codon, which signal for a stop codon at the amino acid number 527 tryptophan (W 527 stop). It is located on the exon 14 on SLC23A1 gene. The nucleotide changed from C to T.
9. COSM20486 is a missense SNV changes the amino acid glycine on position 158 to serine (G 158 S). It is located on exon 6 on the SLC23A1 gene. The nucleotide changed from C to T.
10. Rs6596474 is a missense SNV that changes the amino acid serine on position 421to alanine (S 421A). It is located on exon 11 on the SLC23A1 gene. The nucleotide changed from G to C.

Upon successful amplification, the linear 7884 nucleotides amplicon could be ligated to a full length plasmid carrying the desired mutation. This expression plasmid would thereafter be ready for transfection into mammalian cells for further biochemical and imaging studies (Eck, Kwon, Chen, Mian, & Levine, 2013).

Initial attempts to amplify the entire pReceiver-*SLC23A1*-M55 plasmid with Phusion™ polymerase did not yield the desired amplicons. See as an example (Figure 8), additional information in the appendix (

Figure19 to Figure 25). All PCRs were repeated at least three times. In all cases of failed amplifications of the entire plasmid, the presence of the template and the validity of the reaction setup and PCR conditions were confirmed through positive amplifications of the *SLC23A1* ORF with the validated primers described in Table and

Table 4.

Similar, attempts to amplify the entire pReceiver-M55-*SLC23A1* plasmid with Q5™ polymerase did not yield the desired amplicons. See as an example (Figure 9), additional information in the appendix (Figure 26,

Figure 75). All PCRs were repeated at least three times. In all cases of failed amplifications of the entire plasmid, the presence of the template and the validity of the reaction setup and PCR conditions were confirmed through positive amplifications of the *SLC23A1* ORF with the validated primers described in Table and

Table 4.

Figure 8: Example for a typical agarose gel visualizing amplicons for PCR reactions performed with PhusionTM polymerase for attempted mutations of the entire pReceiver-SLC23A1-M55 plasmid. Here, four primer combinations were tested with the indicated temperature gradients: 1) RS 14116735 F1+R1; 2) RS 139127139 F1+R1; 3) COSM 117494 F1+R1; and 4) RS 34054084 F1+R1. The lanes are flanked by the NEB 1 kb DNA ladder. Expected amplicon size: 7884 nucleotides.

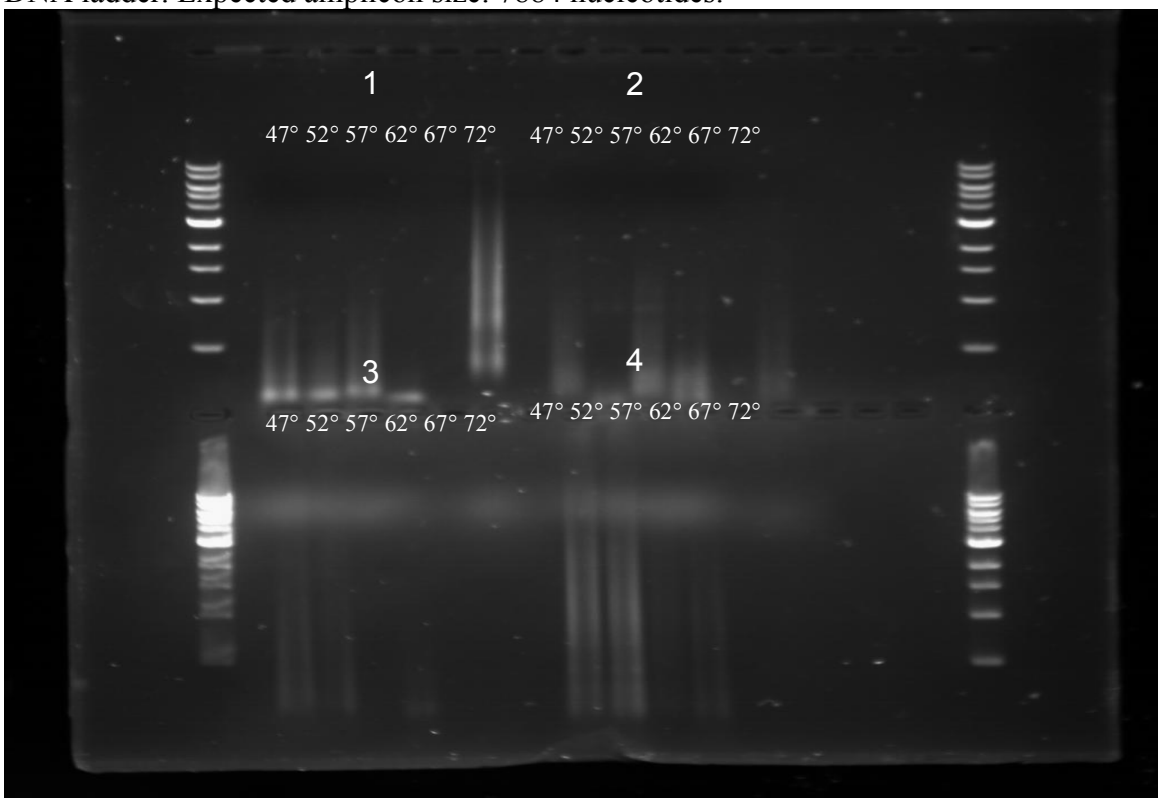
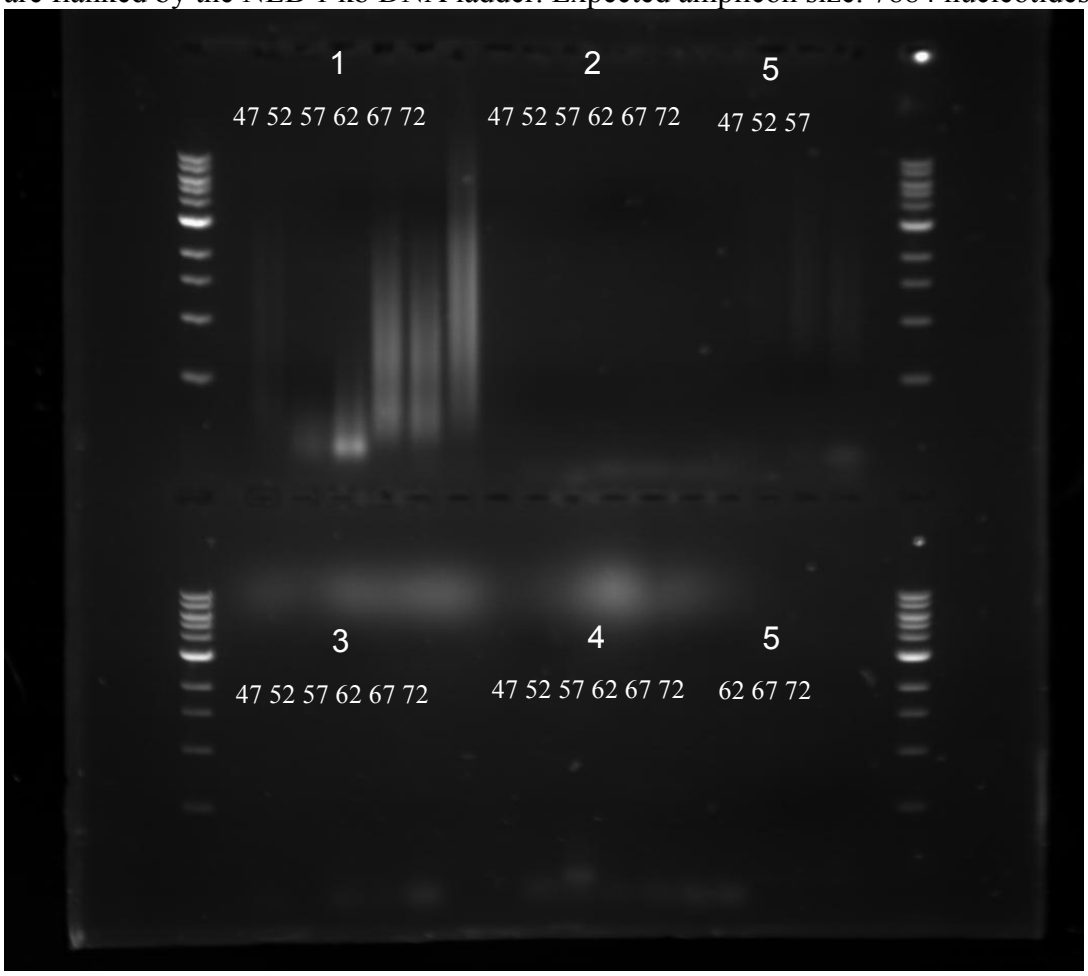


Figure 9: Example for a typical agarose gel visualizing amplicons for PCR reactions performed with Q5 polymerase for attempted mutations of the entire pReceiver-SLC23A1-M55 plasmid. Here, five primer combinations were tested with the indicated temperature gradients: 1) RS 14116735 F1+R1; 2) RS 139127139 F1+R1; 3) COSM117494 F1+R1; 4) RS 34054084 F1+R1; 5) COSM 204877 F1+R1. The lanes are flanked by the NEB 1 kb DNA ladder. Expected amplicon size: 7884 nucleotides.

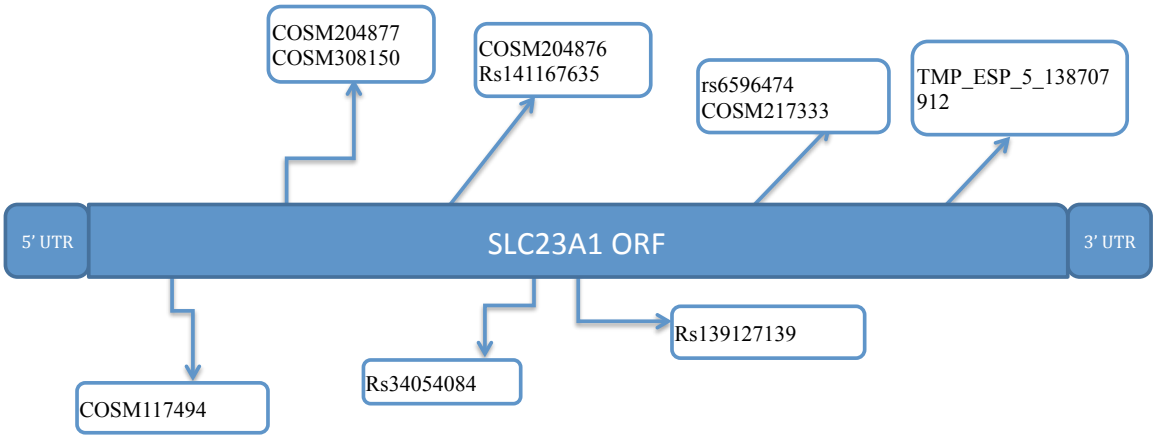


In general, the direct introduction of the SNVs into the pReceiver-*SLC23A1*-M55 plasmid would be the most efficient site directed mutagenesis approach. However, all attempt for ten mutations failed and therefore the strategy was changed the latter approaches, described below.

4.2. Objective B: Introducing SNVs into the middle parts of the *SLC23A1* ORF through amplification of two 5' and the 3' amplicons carrying the variation, followed by ligation of the amplicons and tagging with Gateway® flanking sites ("two step" PCR approach).

The aim of this objective was, to introduce the same ten SNVs listed for Objective A into the *SLC23A1* ORF, using a two-step PCR approach. First the mutations are introduced into two separate 5' and 3' amplicons of the *SLC23A1* ORF. Upon successful amplification, these two amplicons ought to be ligated in a second PCR and in the same reaction the flanking Gateway® tags are added. Thereafter the amplicon is ready for further subcloning using the Gateway® technology. SNVs rs14116735, rs139127139, COSM117494, rs34054084, COSM 204877, COSM 30850, COSM 217333, TMP_ESP_5_138707912, COSM 20486 and rs6596474 were chosen for this approach, their location in the transcript is depicted in Figure 10.

Figure 10: Location of the SNVs selected for the "two step" PCR approach within the SLC23A1 transcript.



All desired 5' and 3' amplicons were produced, as listed in Table . Images for supporting DNA gels can be found in the appendix as referred to in Table .

However, none of the final ligations could be achieved.

Table 6: Summary of 5' and 3' of amplicons containing individual variations, and results of the final ligations. Although 5' and the 3' amplicons for all SNVs were produced, final ligations failed.

Variant (SNV ID)	**5' amplicon size**	**3' amplicon size**	**Final ligation**
rs141167635	649nt (appendix Figure 26, Figure 27	1212nt (appendix Figure 28, Figure 29)	No amplicon (appendix Figure 30, Figure 31)
rs139127139	965 nt (appendix Figure 32, Figure 33)	896 nt (appendix Figure 34, Figure 35)	No amplicon (appendix Figure 36, Figure 37)
COSM117494	51 nt (appendix Figure 38, Figure 39Figure 40.)	1810 nt (appendix Figure 41, Figure 42)	No amplicon (appendix Figure 43, Figure 44.)
rs34054084	825 nt (appendix Figure 45)	1036 nt (appendix Figure 46, Figure 47)	No amplicon (appendix Figure 48, Figure 49.)
COSM204877	241 nt (appendix Figure 50, Figure 51)	1620 nt (appendix Figure 52, Figure 53)	No amplicon (appendix Figure 54,

Variant (SNV ID)	5' amplicon size	3' amplicon size	Final ligation
			Figure 55.)
COSM308150	260 nt (appendix Figure 56, Figure 57)	1601 nt (appendix Figure 58)	No amplicon (appendix Figure 59, Figure 60)
COSM217333	1287 nt (appendix Figure 61)	574 nt (appendix Figure 62)	No amplicon (appendix Figure 63)
TMP_ESP_5_138707912	1613nt (appendix Figure 64)	248 nt (appendix Figure 65)	No amplicon (Figure 66, Figure 67)
COSM20486	505 nt (appendix Figure 68)	1356 nt (appendix Figure 69)	No amplicon (appendix Figure 70, Figure 71)
rs6596474	1294 nt (appendix Figure 72)	567 nt (appendix Figure 73)	No amplicon (appendix Figure 74, Figure 75)

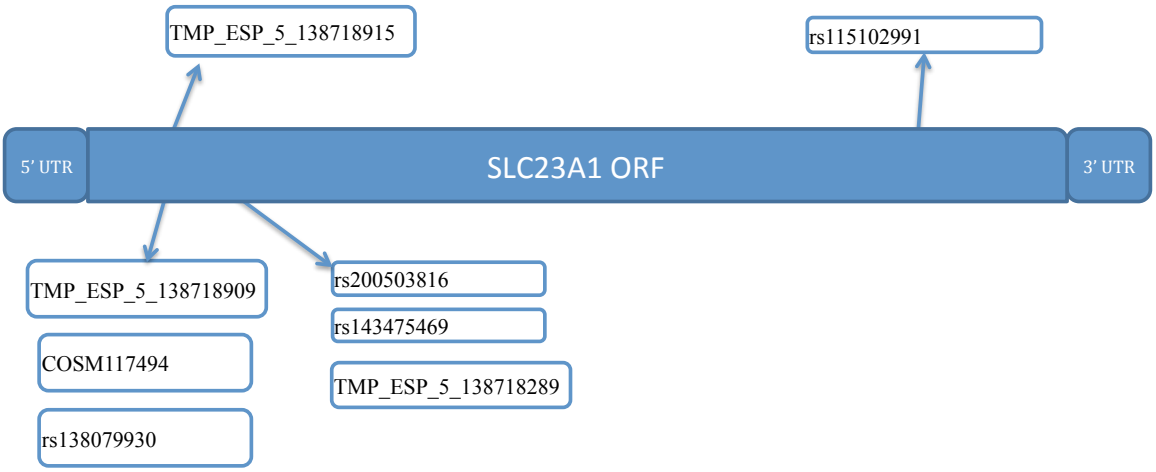
4.3. Objective C: Introducing SNVs in the 5' and the 3' through amplification of the entire *SLC23A1* ORF and tagging with Gateway® flanking sites.

Seven SNVs are located in proximity to the 5' start of the human *SLC23A1* ORF (Figure 11). These variations could therefore be introduced using extensions of the forward primer validated to reliably amplify the *SLC23A1*:

1. TMP ESP 5 138718915
2. TMP ESP 5 138718909
3. TMP ESP 5 138718289
4. COSM117494
5. rs200503816
6. rs138079930
7. rs143475469

Similar, SNV rs115102991is located proximal to the 3' termination site of the *SLC23A1* ORF (Figure 12), and could therefore be amplified when integrated into the antisense reverse primer, which is validated to yield reliable amplicons.

Figure 11: Location of SNVs located proximal to the 5' start and the 3' termination site of SLC23A1 ORF.

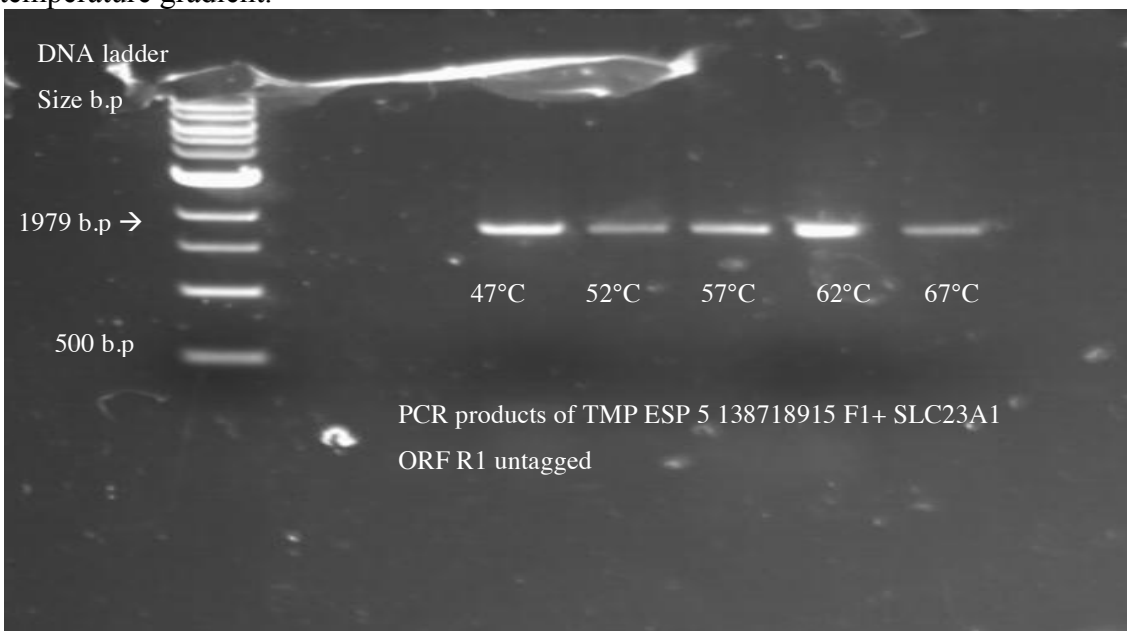


All amplicons introducing these SNVs were achieved and the detailed are described in the following paragraphs for each variation. However, amplicons for SNVs (rs138079930 and rs143475469) could not be achieved (see appendix Figure 76, Figure 79).

4.3.1. TMP_ESP_5_138718915

SNV TMP ESP 5 138718915 changes the arginine on positon 10 to tryptophan (R10W). The base G changed to A, and the amplicon introducing this change was achieved with the primer combination TMP_ESP_5_138718915 F1+ SLC23A1 ORF R1 using Phusion™ high fidelity polymerase (Figure 12).

Figure 12: Agarose gel image depicting an amplicon for the SLC23A1 ORF containing the SNV TMP ESP 5 138718915 was produced using primer combination TMP ESP 5 138718915 F1+ SLC23A1 ORF R1 and PhusionTM high fidelity polymerase over a temperature gradient.

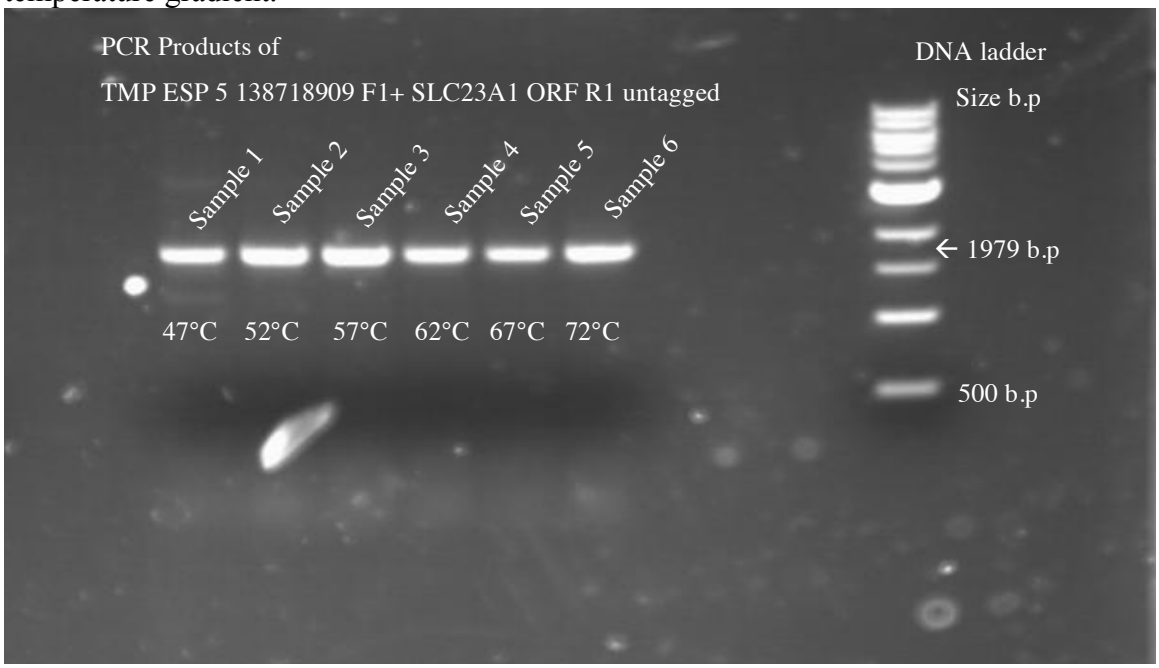


4.3.2 TMP ESP 5 138718909

This is missense SNV that changes glutamine on position 12 to glutamic acid (Q12E). The base C changed to G and this changed was achieved with the primer combination TMP ESP 5 138718909F1 + SLC23A1 ORF R1 and Phusion™ high fidelity polymerase (Figure 13)

Figure 13: Agarose gel image depicting an amplicon for the SLC23A1 ORF containing the SNV TMP ESP 5 138718909 using the primer combination TMP_ESP_5_138718909

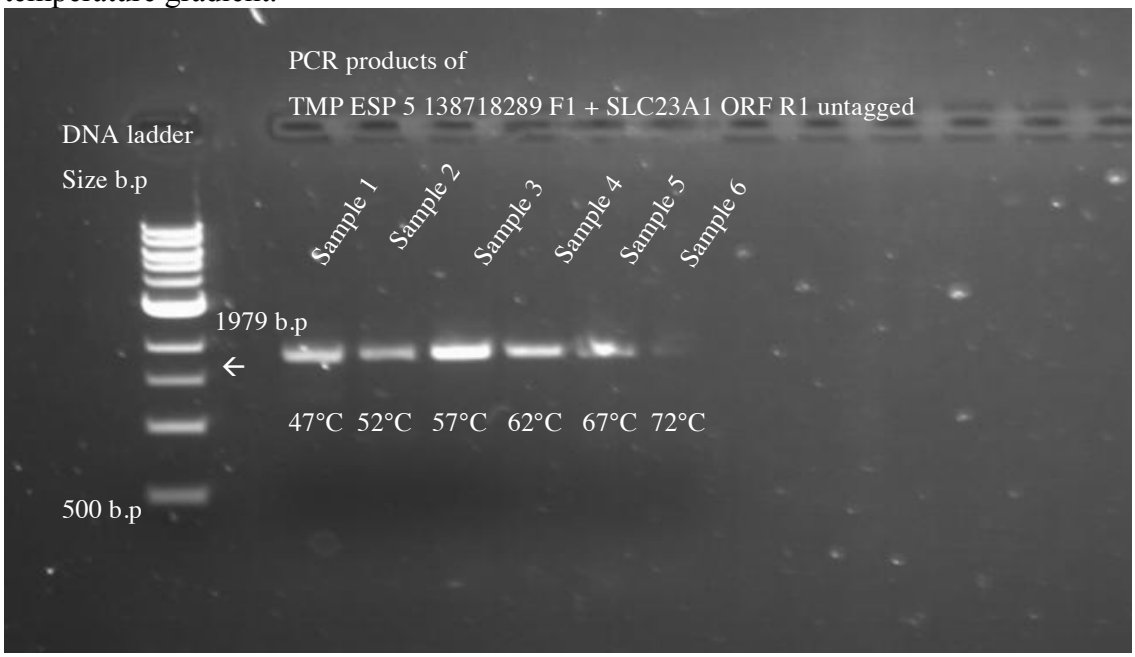
F1+ SLC23A1 ORF R1 untagged and PhusionTM high fidelity polymerase over a temperature gradient.



4.3.3 TMP_ESP_5_138718289

This is missense SNV changes the amino acid glutamic acid on position 14 to aspartic acid (E14D). The base A changed to T and this change was achieved using the primers combination TMP_ESP_5_138718289F1 + SLC23A1 ORF R1 untagged and Phusion™ high fidelity polymerase (Figure 14).

Figure 14: Agarose gel image depicting an amplicon for the SLC23A1 ORF containing the SNV TMP ESP 5 138718289 using the primer combination TMP ESP 5 138718289 F1 + SLC23A1 ORF R1 untagged and PhusionTM high fidelity polymerase over a temperature gradient.

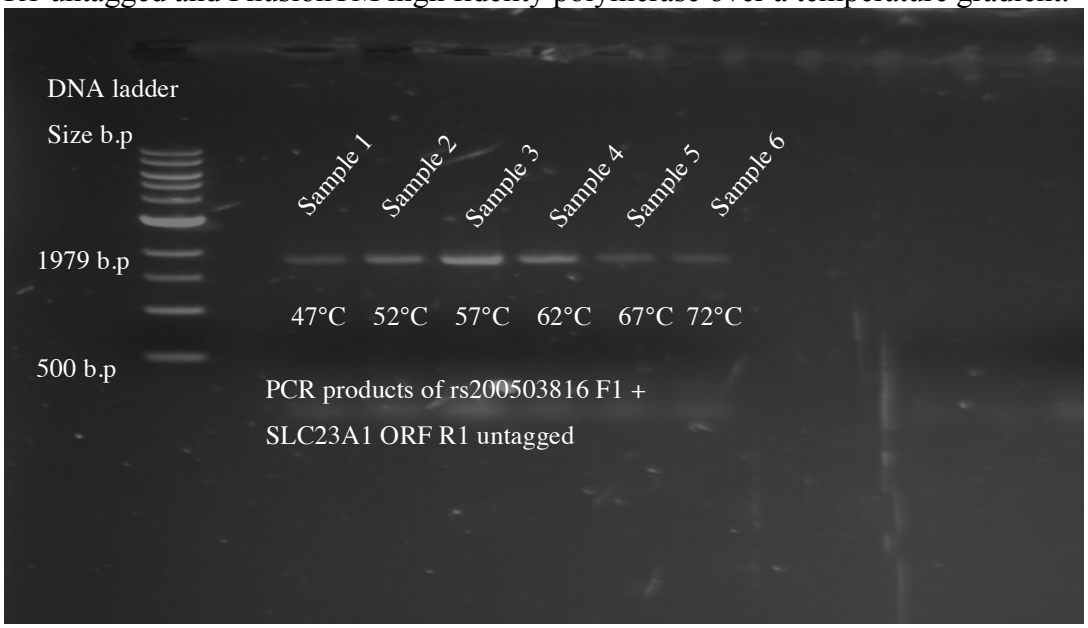


4.3.4 RS200503816

This is a missense SNV that changes the amino acid histidine to tyrosine on position 13 (H13Y). The base changes from G to A, and this change was reached using the primers combination rs200503816F1 + SLC23A1 ORF R1 and Phusion^{TM} high fidelity polymerase (

Figure 15: Agarose gel image depicting an amplicon for the SLC23A1 ORF containing the SNV rs200503816 using the primer combination rs200503816 F1 + SLC23A1 ORF R1 untagged and PhusionTM high fidelity polymerase over a temperature gradient.
).

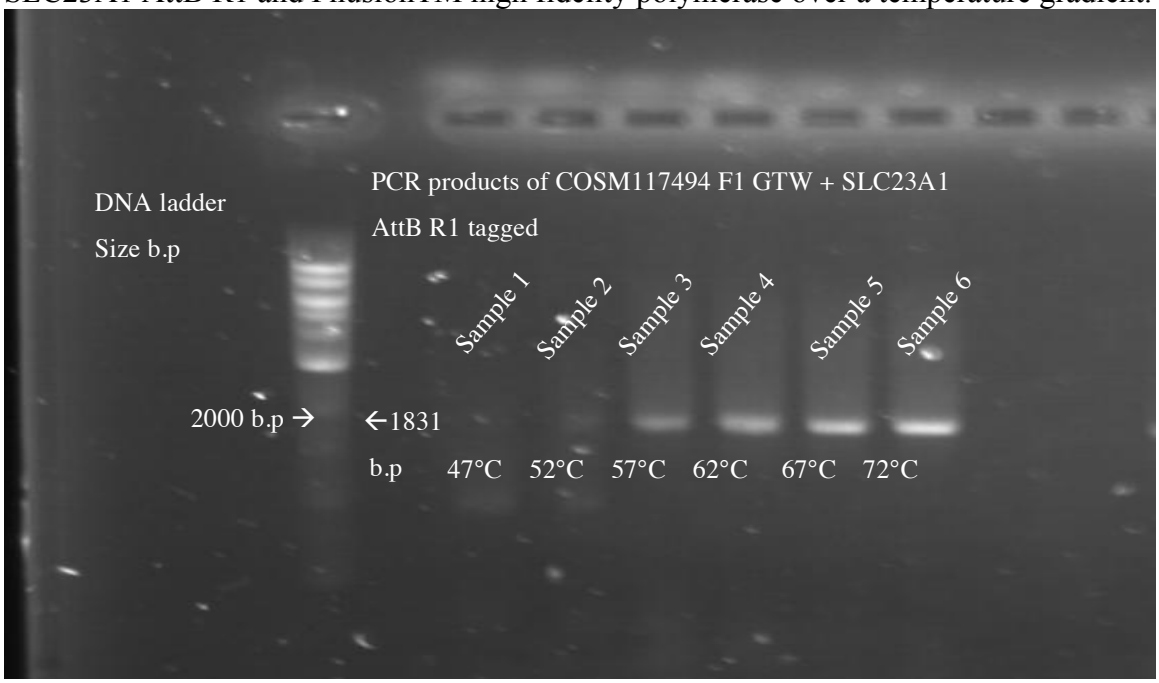
Figure 15: Agarose gel image depicting an amplicon for the SLC23A1 ORF containing the SNV rs200503816 using the primer combination rs200503816 F1 + SLC23A1 ORF R1 untagged and PhusionTM high fidelity polymerase over a temperature gradient.



4.3.5. COSM117494

This is a missense SNV that changes the amino acid aspartic acid on position 6 to glutamic acid (D6E). The base G substituted with base C and this changed was achieved using the primers combination COSM117494 F1GTW + SLC23A1 ORF R1 untagged and Phusion^{TM} high fidelity polymerase (Figure 16).

Figure 16: Agarose gel image depicting an amplicon for the SLC23A1 ORF containing the SNV COSM117494 using the primer combination COSM117494 F1 GTW + SLC23A1 AttB R1 and PhusionTM high fidelity polymerase over a temperature gradient.



4.3.6. Rs115102991

This is a missense SNV that changes the amino acid from valine on position 598 to leucine (V598L). The base C changed to G and this change was achieved using the primers combination rs115102991 R1AttB2 + SLC23A1 ORF R1 untagged and Phusion^{TM} high fidelity polymerase reaction (Figure 17).

Figure 17: Agarose gel image depicting an amplicon for the SLC23A1 ORF containing the SNV rs115102991 using the primer combination rs115102991 R1 AttB2 + SLC23A1 AttB F1 and PhusionTM high fidelity polymerase over a temperature gradient.

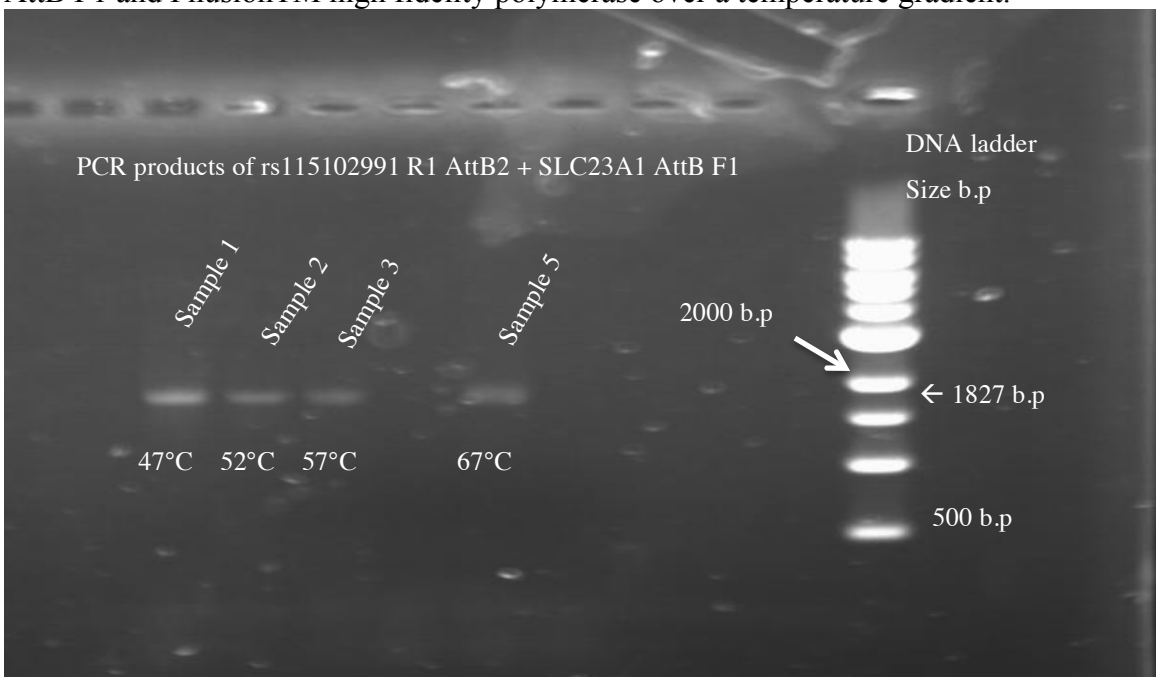
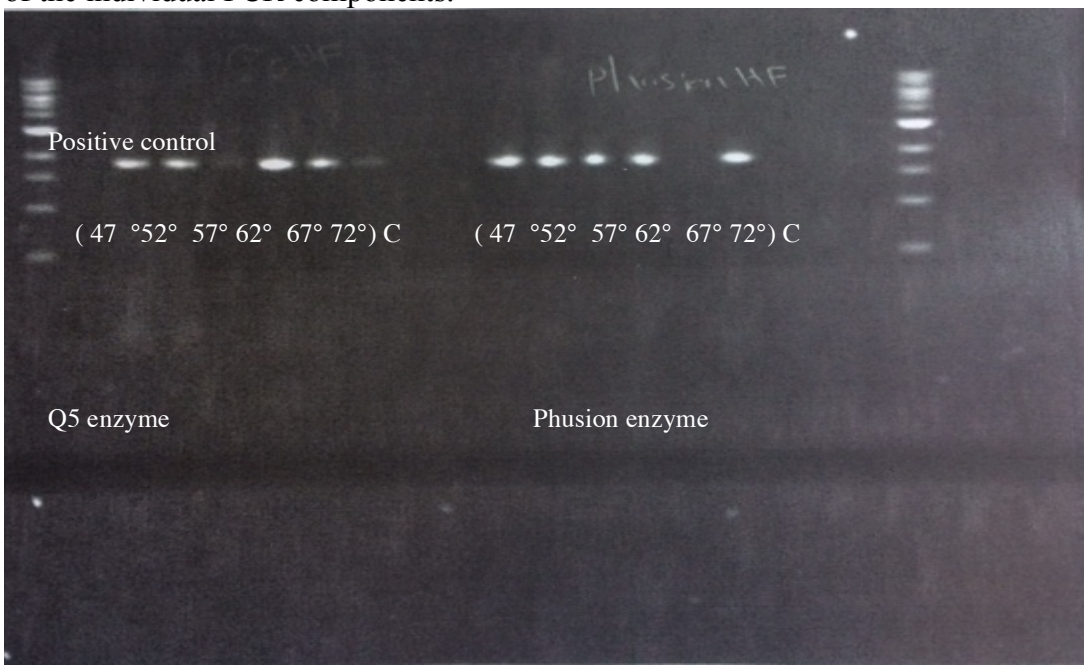


Figure 18: Example for a gel image visualizing applicable positive controls for amplifications of the SLC23A1 ORF performed with PhusionTM high fidelity polymerase and Q5TM HF polymerase. Identical positive controls were performed when amplifications to introduce SNVs repeatedly failed, in order to validate the performance of the individual PCR components.



5- Discussion

Rationale underlying this pilot project.

Due to the successful completion of the human genome project in 2003, we currently experience a surge in genetic knowledge (Lander, 2011). Since the early 2000s a wealth of genetic information had been discovered through various sequencing projects (Duncan, Pollastri, & Smoller, 2014) (Munroe, Barnes, & Caulfield, 2013; Ombrello, Sikora, & Kastner, 2014) (Gonzaga-Jauregui, Lupski, & Gibbs, 2012). With the development of genome wide association studies many genetic variation has been associated to various diseases (Abu-Amero, Kondkar, & Chalam, 2015; Christoffersen & Tybjaerg-Hansen, 2015; Jimenez-Jimenez, Alonso-Navarro, Garcia-Martin, & Agundez, 2016; Shen & Jia, 2016). However, associations in epidemiological studies are only indicating a possible causation for a disease. In order to validate that any genetic associations are causative, further studies are warranted. These include confirmations in independent cohorts, controlled randomized clinical trials, or biological models.

To date various diseases had been associated to variations in ascorbic acid transporters genes, and specifically the *SLC23A1* gene seems to be relevant (AmirShaghaghi, Kloss, & Eck, 2016). This notion is derived from the fact that the genes global elimination is not lethal for the *Slc23a1*$^{-/-}$ mouse, but the *slc23a1*$^{-/-}$ mice show high ascorbic acid losses in the urine, corresponding low systemic levels, accompanied by high lethality in the offspring litters (Corpe et al., 2010a). Moreover, the SNP rs35817838, which changes the base adenine at position 772 to guanine, encoding a change of the amino acid methionine 258 to valine (M258V), severely decreases SLC23A1 activity when expressed in the Xenopus laevis oocyte system (Corpe et al.,

2010a). SNP rs33972313 is in linkage with SNP rs35817838, which is associated with lower circulating ascorbic acid levels (Timpson et al., 2010), increased susceptibility to Crohn's Disease (Amir Shaghaghi et al., 2014), gastric cancer (Duell et al., 2013), ischemic heart disease and all-cause mortality (Kobylecki, Afzal, & Nordestgaard, 2015). The existence of disease association creates the rationale to conduct this pilot study. The initial objectives are to introduce variations into an *SLC23A1* expression plasmid. In follow up studies, it can be determined how these impact on SLC23A1's function. This can be done using the *Xenopus laevis* oocyte expression system, a valid model to determine the impact of individual genetic variations in *SLC23A1* on the transport capacity (Corpe et al., 2010a).

Ultimately, detrimental variations are likely involved in disease causing processes and the suggested in vitro studies enable the prioritization of variations to be studied in individuals. Since most putative detrimental variations currently found in genetic databases are very rare, this will narrow the range of SNVs which needs to be screened in order to identify carriers of potential disease causing mutations.

Approaches used to introduce DNA variations.

In order to produce mutated plasmids carrying the *SLC23A1* ORF a PCR based approach using the high fidelity polymerases Phusion™ and Q5™ was chosen. Despite the use of these highly efficient polymerases, the objectives to introduction selected variations into the *SLC23A1* protein coding region were only partially achieved. Out of the three approaches chosen two yielded results. Reasons for the successes and failures are discussed in the following paragraphs.

PCR based introduction of site directed mutagenesis is a general and simple method

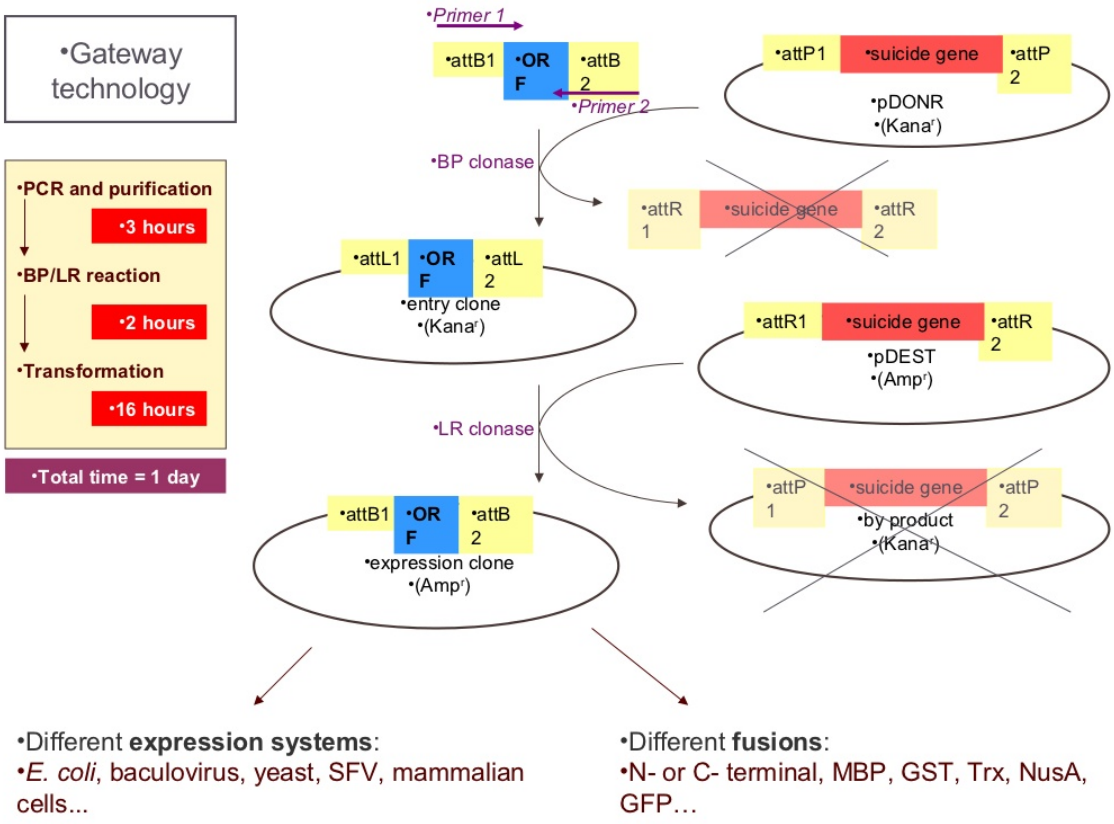
for specific sequence changes in a plasmid. This approach is successfully used since 1989 (Hemsley, Arnheim, Toney, Cortopassi, & Galas, 1989) and forms the basis for commercially available kits, such as the QuikChange™ Kit (Stratagene). Generally, after overlapping amplicons are created, they are fused through scarless DNA assembly methods. The approach chosen in this project deviated from previous protocols, such as the QuikChange™ protocol, in that the two higher fidelity polymerases (Phusion™ and Q5™) over PfuTurbo™ DNA polymerase were used; however, the primers were still complementary overlapping. Complementary primers were used in the intention to linearly amplify the target plasmid to produce linear double-stranded DNA molecules with short homologous ends. The annealing of these homologous overhangs is the basis to form circular plasmids when transformed into *Escherichia coli*, utilizing an endogenous plasmid repair mechanism (QuikChange™ Kit Manual, Stratagene).

All attempts to amplify the entire pReceiver-M55-*SLC23A1* plasmid with ten mutation-carrying primer pairs failed. The amplicon size of 7884 nucleotides is well within the capacity of the Phusion™ and Q5™ polymerases ("NewEnglandBiolabs: PCR Using Q5® High-Fidelity DNA Polymerase. Manual," ; "NewEnglandBiolabs: Phusion™ High-Fidelity DNA Polymerase. Manual,") and successful amplifications of similar sized products are reported (Xia, Cui, Li, & Wu, 2014). Therefore, the failure to produce the linear amplicon could be related to the formation of primer-dimers. It was reported that primers over 28 nucleotides failed to yield an amplicon of a plasmid when Phusion™ and Q5™ were utilized, but that overlapping primers of 20 nucleotides did yield products (Xia et al., 2014). It could be speculated, that Phusion™ and Q5™ require higher annealing temperatures, which may promote the formation of the perfectly matching primer dimers

over the formation of the primer-template duplex containing mismatches. However, it also could be due to the intrinsic nature of the enzymes, and an enzyme with a higher affinity to the substrate (primer-template duplex) could have a better chance to stabilize the duplex before its dissociation. However, at least some of the shorter length primers utilized here should have yielded an amplicon. However, primer-dimer formation dependents on the specific sequences of individual primers, which makes prediction of success or failure for individual oligonucleotides in PCR reactions impossible.

The use of partially overlapping primers instead of complementary primers had been suggested in order to avoid excessive primer dimer formation and to improve amplicon yield (Xia et al., 2014). In future attempts, the redesign of primers will be one option to improve amplification success.

The amplification of two pieces of the *SLC23A1* ORF and subsequent assembly to the full length ORF is a complementary method to the amplification of the entire plasmid. The advantage of the two piece amplification method is that the ends of the full length amplicons can be tagged with sequences enabling further sub-cloning. In this project the introduction of the Gateway® tagging sites AttB1 and AttB2 were chosen to enable further sub-cloning into an entry plasmid and thereafter into destination expression plasmids. In contrast to the "around the plasmid" amplification, this approach also eliminates potential carry over of un-mutated template plasmid, since the two Gateway® cloning steps utilize selection based on the two different antibiotics Kanamycin and Ampicillin. This eliminates possible plasmid carryover of the un-mutated template plasmid through differential antibiotics selection (Gateway® Technology manual - Thermo Fisher Scientific, https://tools.thermofisher.com/content/sfs/manuals/):



The "two step" PCR based introduction seemed to be specifically feasible for this project, since the primer pair utilized to amplify the entire human *SLC23A1* ORF had been validated to yield strong amplicons with Phusion™ and Q5™ polymerases. Therefore, amplifications with the respective *SLC23A1* ORF primers in combination with the primers introducing the mutations should have been feasible. In fact, all of the twenty amplicons of these 5' and 3' pieces were successful (Table).

For the subsequent seamless ligation into the full length mutated *SLC23A1* ORF, a PCR based approach utilizing the same polymerases used to produce the 5' and 3' pieces was chosen. In this "second step" ligation the Gateway® tags AttB1 and AttB2 were added to the respective forward and reverse primers. It had been validated that this

addition to the *SLC23A1* ORF primers does not impact on the PCR amplicon. These primers could therefore be used in the second PCR round to enable subcloning using the Gateway® system, for which a multitude of expression vectors exist (Atanassov et al., 2009).

However, the second PCR step ligating the 5' and 3' pieces into the full length mutated *SLC23A1* ORF was not achieved. The failure of the final assembly comes as a surprise, since this and similar techniques are widely used (Smith, 2003), as demonstrated by the existence of a Wikipedia entry (Polymerase cycling assembly, https://en.wikipedia.org/wiki/Polymerase-_cycling_assembly) and the use in commercial gene assembly methodology, such as the gBlocks® Gene Fragments assembly (Integrated DNA Technologies [IDT], https://www.idtdna.com/pages/decoded/decoded-articles/core-concepts/decoded/2012/09/21/assembly-pcr-for-novel-gene-synthesis). In spite of the use of temperature gradients in reactions using three different polymerases (Phusion, Q5, Taq) no amplicons were produced. The integrity of individual reaction settings was confirmed using the validated control amplifying the full length *SLC23A1* ORF. Therefore, the lack of amplifications are not related to technical failures. It can only be speculated that a formation of dimeric DNA components in the preformed reactions or secondary structures caused the inhibition of the PCR reaction.

To circumvent PCR related problems, future experimental designs could consider the use of alternative scarless assembly methods, such as the Gibson assembly (Gibson, 2011), which can joint DNA fragments in an isothermal reaction at at 50 °C using a dedicated exonuclease and a ligase to seal the single stranded nicks. Moreover, a variety of other methods, such as SLIC (sequence and ligase independent cloning, Li 2007),

CPEC (or circular polymerase extension cloning, Quan 2009), SLiCE (Seamless Ligation Cloning Extract, Zhang 2012) could be considered to improve results on final ligations.

For the variations in proximity to the 5' or 3' end of the *SLC23A1* ORF, modified primers containing the desired mutations and amplifying the entire amplicon in one piece were used. Five variations in the 5' and one in the 3'prime of the *SLC23A1* ORF were introduced using this method. These are now ready for sub-cloning with the Gateway system, since the respective tags have been added to the amplicons. The success of this PCR based approach finally demonstrated that PCR based methodology can be used in site directed mutagenesis of the *SLC23A1* plasmid.

Impacts of *SLC23A1* variations on its biology.

To date only a limited number of variations, all of them SNPs, have been tested if they altered the function of the SLC23A1 protein. Out of four SNPs tested, the SNP rs35817838-G allele was shown to severely reduce SLC23A1 function (Corpe et al., 2010b). This proofs the principal that variations in SLC23A1 can significantly impact on its transporter function and that these variations are found in the general population or specific subpopulations. However, how detrimental variations impact on the health status of carriers still needs to be investigated. In general, common SNPs, found at population frequencies of more than 5%, are predicted to have lower impacts on a proteins biology. In contrast, rare variations with low population penetrance are predicted to exhibit high impacts on the protein functions.

Beyond the common SNPs discovered about 15 years ago, which had been tested for biological impact, recently many rare variations were deposited into the databases. The

presented project was a pilot study evaluating how to introduce selected rare variations into *SLC23A1*. However, even before *in vitro* studies following up on this project will proof detrimental impacts, it is likely that variations abrogating SLC23A1 functions exist. At the beginning of this project, six SNV predicted to have extremely high impact on the transporters functions were found in the databases. Specifically, four SNVs introducing a premature stop codon (Table 6), and two SNVs introducing a frameshift and premature termination (Table 7) exist. Except SNV TMP_ESP_5_138707912, these variations all shorten the SLC23A1 protein significantly, implying the abrogation of its transporter function. These SNV have extremely low allelic frequencies, e.g. 0.00003942 for rs141167635 (p.Arg210Ter), 0.000008 for TMP_ESP_5_138707912/rs373778813, and 0.000008 for rs34054084. These frequencies are also seen for most other SNV in the *SLC23A1* gene, e.g. the missense mutations currently found in the database.

Table 5: SNV in SLC23A1 causing a gain of Stop codon (as of September 2013).

ID	Chr: bp	Alleles	Amino Acid #	AA change
rs11242462	5:138720108	C/T	45	W/*
rs141167635	5:138715928	G/A	206	R/*
rs139127139	5:138715037	C/T	311	W/*
TMP_ESP_5_138707912	5:138707912	C/T	527	W/*

Table 6: SNV in SLC23A1 causing a Frameshift (as of September 2013).

ID	Chr: bp	Alleles	Class	Type	Amino acid #
rs34054084	5:138715501 -138715500	-/C	insertion	Frameshift variant	265
COSM217333	5:138713966 -138713965	-/C	insertion	Frameshift variant	423

These low allelic frequencies are comparable to mutations causing rare genetic disease. This mean, that a carriers of a non-functional copy likely would have another functional *SLC23A1* copy, compensating for the loss of one allele, and therefore not resulting in a disease phenotype. This phenomenon is seen in many rare genetic diseases, where carriers are unaffected, however homozygote or compound heterozygote carriers of detrimental mutations will develop a sever phenotype. It could be speculated that detrimental genetic variations in *SLC23A1* could be the unrecognized cause for a rare genetic disorder.

In order to elucidate potential disease mechanism the knowledge of the biological impact of individual variations is the basis for further studies. Follow up studies for this project will determine the variations impacting on the transporters functions, enabling prioritization for genetic screening in existing populations. It is anticipated that not all of the known variations, specifically not all missense variations will impact on the functions. The functional studies of the individual variations will specifically enable the prioritization of all existing SNVs for population screening.

Besides germline mutations, few somatic mutations in SLC23A1 are found in the databases. These are spontaneous mutations not found in most of the body's cells, but in cancers. This opens up questions about the role of *SLC23A1* mutations in cancer development and treatment. Low ascorbic acid levels are associated with the development of cancers, and genetic variations in the ascorbic acid transporters are associated with certain cancers (Skibola et al., 2008). Therefore, loss of function mutations in *SLC23A1* could very well contribute to lowered intracellular ascorbic acid levels, which over time could contribute to cancer development. How somatic variations

contribute to cancer development will be one of the key areas in the area of ascorbic acid Nutrigenomics.

Moreover, the efficacy of cancer therapies could be influenced by intracellular ascorbic acid levels. Higher intracellular levels of the antioxidant ascorbic acid could inhibit the efficacy of antineoplastic therapies (Mut-Salud et al., 2016). Paradoxically, loss of function somatic mutations in *SLC23A1* could contribute to higher efficacy of these therapies. If this turns out to be the case, future precision therapies for certain cancer types could be based on SLC23A1 genotypes.

6- Future Objectives

This pilot study forms the basis to optimize site directed mutagenesis of the *SLC23A1* ORF in an expression plasmid. Once this is achieved, mutations can be functionally tested in vitro, in the *Xenopus laevis* oocyte system or in mammalian cell culture. This will elucidate impacts on the transport efficiency and intracellular targeting. Based on these result, carriers of potentially detrimental genotypes can be identified for vitamin C pharmacokinetics studies, in order to assess an effect on systemic levels.

The long-term goal is to evaluate personalized dietary intakes for vitamin C in order to optimize individuals' health in the framework of precision medicine.

References

Abu-Amero, K., Kondkar, A. A., & Chalam, K. V. (2015). An Updated Review on the Genetics of Primary Open Angle Glaucoma. *Int J Mol Sci, 16*(12), 28886-28911. doi:10.3390/ijms161226135

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat Methods, 7*(4), 248-249. doi:10.1038/nmeth0410-248

Amir Shaghaghi, M., Bernstein, C. N., Serrano Leon, A., El-Gabalawy, H., & Eck, P. (2014). Polymorphisms in the sodium-dependent ascorbate transporter gene SLC23A1 are associated with susceptibility to Crohn disease. *Am J Clin Nutr, 99*(2), 378-383. doi:10.3945/ajcn.113.068015

AmirShaghaghi, M., Kloss, O., & Eck, P. (2016). Genetic Variation in Human Vitamin C Transporter Genes in Common Complex Diseases. *Advances in nutrition (Bethesda, Md.), 7*, 1-12. doi:doi:10.3945/an.115.009225

Becker, K. G., Barnes, K. C., Bright, T. J., & Wang, S. A. (2004). The genetic association database. *Nat Genet, 36*(5), 431-432. doi:10.1038/ng0504-431

Cahill, L. E., & El-Sohemy, A. (2009). Vitamin C transporter gene polymorphisms, dietary vitamin C and serum ascorbic acid. *J Nutrigenet Nutrigenomics, 2*(6), 292-301. doi:000314597 [pii]
10.1159/000314597

Christoffersen, M., & Tybjaerg-Hansen, A. (2015). Novel genes in LDL metabolism--a comprehensive overview. *Curr Opin Lipidol, 26*(3), 179-187. doi:10.1097/MOL.0000000000000175

Corpe, C. P., Tu, H., Eck, P., Wang, J., Faulhaber-Walter, R., Schnermann, J., . . . Levine, M. (2010a). Vitamin C transporter Slc23a1 links renal reabsorption, vitamin C tissue accumulation, and perinatal survival in mice. *J Clin Invest, 120*(4), 1069-1083. doi:10.1172/jci39191

Corpe, C. P., Tu, H. B., Eck, P., Wang, J., Faulhaber-Walter, R., Schnermann, J., . . . Levine, M. (2010b). Vitamin C transporter Slc23a1 links renal reabsorption, vitamin C tissue accumulation, and perinatal survival in mice. *Journal of Clinical Investigation, 120*(4), 1069-1083. doi:10.1172/Jci39191

Duell, E. J., Lujan-Barroso, L., Llivina, C., Muñoz, X., Jenab, M., Boutron-Ruault, M.-C., . . . González, C. A. (2013). Vitamin C transporter gene (SLC23A1 and SLC23A2) polymorphisms, plasma vitamin C levels, and gastric cancer risk in the EPIC cohort. *Genes & Nutrition, 8*(6), 549-560. doi:10.1007/s12263-013-0346-6

Duncan, L. E., Pollastri, A. R., & Smoller, J. W. (2014). Mind the gap: why many geneticists and psychological scientists have discrepant views about gene-environment interaction (GxE) research. *Am Psychol, 69*(3), 249-268. doi:10.1037/a0036320

Eck, Kwon, Chen, Mian, & Levine. (2013). The human sodium-dependent ascorbic acid transporters SLC23A1 and SLC23A2 do not mediate ascorbic acid release in the proximal renal epithelial cell. *Physiological Reports, 1*(6), e00136. doi:10.1002/phy2.136

Eck, P., Erichsen, H. C., Taylor, J. G., Yeager, M., Hughes, A. L., Levine, M., & Chanock, S. (2004). Comparison of the genomic structure and variation in the two human sodium-dependent vitamin C transporters, SLC23A1 and SLC23A2. *Hum Genet, 115*(4), 285-294. doi:10.1007/s00439-004-1167-x

Erichsen, H. C., Eck, P., Levine, M., & Chanock, S. (2001). Characterization of the genomic structure of the human vitamin C transporter SVCT1 (SLC23A2). *J Nutr, 131*(10), 2623-2627.

Erichsen, H. C., Peters, U., Eck, P., Welch, R., Schoen, R. E., Yeager, M., . . . Chanock, S. (2008). Genetic Variation in Sodium-Dependent Vitamin C Transporters SLC23A1 and SLC23A2 and Risk of Advanced Colorectal Adenoma. *Nutrition and cancer, 60*(5), 652-659. doi:10.1080/01635580802033110

Genetics Home Reference. (2014). Retrieved 2014 http://ghr.nlm.nih.gov/handbook/mutationsanddisorders/genemutation

Gibson, D. G. (2011). Chapter fifteen - Enzymatic Assembly of Overlapping DNA Fragments. In V. Christopher (Ed.), *Methods in Enzymology* (Vol. Volume 498, pp. 349-361): Academic Press.

Gonzaga-Jauregui, C., Lupski, J. R., & Gibbs, R. A. (2012). Human Genome Sequencing in Health and Disease. *Annual review of medicine, 63*, 35-61. doi:10.1146/annurev-med-051010-162644

Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., & Galas, D. J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Research, 17*(16), 6545-6551.

Hiroyasu Tsukaguchi1, Taro Tokui1,4,5, Bryan Mackenzie1,4, Urs V. Berger1, Xing-Zhen Chen1, Yangxi Wang1, Richard F. Brubaker3 & Matthias A. Hediger1. (1999). A family of mammalian Na+-dependent L-ascorbic acid transporters. *Nature, 399*(6731), 70-75. doi:10.1038/19986

Initial sequencing and analysis of the human genome. (2001). *Nature, 409*(6822), 860-921. doi:http://www.nature.com/nature/journal/v409/n6822/suppinfo/409860a0_S1.html

Institute of Medicine Panel on Dietary, A., & Related, C. (2000) *Dietary Reference Intakes for Vitamin C, Vitamin E, Selenium, and Carotenoids*. Washington (DC): National Academies Press (US)

Ishii, T. M., Zerr, P., Xia, X. M., Bond, C. T., Maylie, J., & Adelman, J. P. (1998). Site-directed mutagenesis. *Methods Enzymol, 293*. doi:10.1016/s0076-6879(98)93007-5

Jimenez-Jimenez, F. J., Alonso-Navarro, H., Garcia-Martin, E., & Agundez, J. A. (2016). Advances in understanding genomic markers and pharmacogenetics of Parkinson's disease. *Expert Opin Drug Metab Toxicol*. doi:10.1517/17425255.2016.1158250

Kobylecki, C. J., Afzal, S., & Nordestgaard, B. G. (2015). Genetically Low Antioxidant Protection and Risk of Cardiovascular Disease and Heart Failure in Diabetic Subjects. *EBioMedicine, 2*(12), 2010-2015. doi:10.1016/j.ebiom.2015.11.026

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protocols, 4*(8), 1073-1081.

Lander1, L. M. L., Bruce Birren1, Chad Nusbaum1, Michael C. Zody1, Jennifer Baldwin1, Keri Devon1, Ken Dewar1, Michael Doyle1, William FitzHugh1, Roel Funke1, Diane Gage1, Katrina Harris1, Andrew Heaford1, John Howland1, Lisa Kann1, Jessica Lehoczky1, Rosie LeVine1, Paul McEwan1, Kevin McKernan1, James Meldrim1, Jill P. Mesirov1, Cher Miranda1, William Morris1, Jerome Naylor1, Christina Raymond1, Mark Rosetti1, Ralph Santos1, Andrew Sheridan1, Carrie Sougnez1, Nicole Stange-Thomann1, Nikola Stojanovic1, Aravind Subramanian1 & Dudley Wyman1 for Whitehead Institute for Biomedical Research, Center for Genome Research:, Jane Rogers2, John Sulston2, Rachael Ainscough2, Stephan Beck2, David Bentley2, John Burton2, Christopher Clee2, Nigel Carter2, Alan Coulson2, Rebecca Deadman2, Panos Deloukas2, Andrew Dunham2, Ian Dunham2, Richard Durbin2, Lisa French2, Darren Grafham2, Simon Gregory2, Tim Hubbard2, Sean Humphray2, Adrienne Hunt2, Matthew Jones2, Christine Lloyd2, Amanda McMurray2, Lucy Matthews2, Simon Mercer2, Sarah Milne2, James C. Mullikin2, Andrew Mungall2, Robert Plumb2, Mark Ross2, Ratna Shownkeen2 & Sarah Sims2 for The Sanger Centre:, Robert H. Waterston3, Richard K. Wilson3, LaDeana W. Hillier3, John D. McPherson3, Marco A. Marra3, Elaine R. Mardis3, Lucinda A. Fulton3, Asif T. Chinwalla3, Kymberlie H. Pepin3, Warren R. Gish3, Stephanie L. Chissoe3, Michael C. Wendl3, Kim D. Delehaunty3, Tracie L. Miner3, Andrew Delehaunty3, Jason B. Kramer3, Lisa L. Cook3, Robert S. Fulton3, Douglas L. Johnson3, Patrick J. Minx3 & Sandra W. Clifton3 for Washington University Genome Sequencing Center, Trevor Hawkins4, Elbert Branscomb4, Paul Predki4, Paul Richardson4, Sarah Wenning4, Tom Slezak4, Norman Doggett4, Jan-Fang Cheng4, Anne Olsen4, Susan Lucas4, Christopher Elkin4, Edward Uberbacher4 & Marvin Frazier4 for US DOE Joint Genome Institute:, Richard A. Gibb. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921. doi:http://www.nature.com/nature/journal/v409/n6822/suppinfo/409860a0_S1.html

Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature, 470*(7333), 187-197. doi:http://www.nature.com/nature/journal/v470/n7333/abs/10.1038-nature09792-unlocked.html - supplementary-information

Levine, M., Padayatty, S. J., & Espey, M. G. (2011). Vitamin C: a concentration-function approach yields pharmacology and therapeutic discoveries. *Advances in nutrition (Bethesda, Md.), 2*(2), 78-88.

Ling, M. M., & Robinson, B. H. (1997). Approaches to DNA mutagenesis: an overview. *Anal Biochem, 254*(2), 157-178. doi:10.1006/abio.1997.2428

Liu, H., & Naismith, J. (2008). An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology, 8*(1), 91.

Monsen, E. R. (2000). Dietary reference intakes for the antioxidant nutrients: vitamin C, vitamin E, selenium, and carotenoids. *J Am Diet Assoc, 100*(6), 637-640. doi:10.1016/s0002-8223(00)00189-9

Munroe, P. B., Barnes, M. R., & Caulfield, M. J. (2013). Advances in blood pressure genomics. *Circ Res, 112*(10), 1365-1379. doi:10.1161/circresaha.112.300387

Mut-Salud, N., Álvarez, P. J., Garrido, J. M., Carrasco, E., Aránega, A., & Rodríguez-Serrano, F. (2016). Antioxidant Intake and Antitumor Therapy: Toward Nutritional Recommendations for Optimal Results. *Oxidative Medicine and Cellular Longevity, 2016*, 6719534. doi:10.1155/2016/6719534

NewEnglandBiolabs: PCR Using Q5® High-Fidelity DNA Polymerase. Manual.

NewEnglandBiolabs: Phusion™ High-Fidelity DNA Polymerase. Manual.

Ng, P. C., & Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research, 31*(13), 3812-3814.

Nishikimi, M., Fukuyama, R., Minoshima, S., Shimizu, N., & Yagi, K. (1994). Cloning and chromosomal mapping of the human nonfunctional gene for L-gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *J Biol Chem, 269*(18), 13685-13688.

Ombrello, M. J., Sikora, K. A., & Kastner, D. L. (2014). Genetics, genomics and their relevance to pathology and therapy. *Best practice & research. Clinical rheumatology, 28*(2), 175-189. doi:10.1016/j.berh.2014.05.001

Padayatty, S. J., & Levine, M. (2001). New insights into the physiology and pharmacology of vitamin C. *CMAJ, 164*(3), 353-355.

Savini, I., Rossi, A., Pierro, C., Avigliano, L., & Catani, M. V. (2008). SVCT1 and SVCT2: key proteins for vitamin C uptake. *Amino Acids, 34*(3), 347-355. doi:10.1007/s00726-007-0555-7

Shen, L., & Jia, J. (2016). An Overview of Genome-Wide Association Studies in Alzheimer's Disease. *Neurosci Bull*. doi:10.1007/s12264-016-0011-3

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res, 29*(1), 308-311.

Sinha, R., Block, G., & Taylor, P. R. (1993). Problems with estimating vitamin C intakes. *Am J Clin Nutr, 57*(4), 547-550.

Skibola, C. F., Bracci, P. M., Halperin, E., Nieters, A., Hubbard, A., Paynter, R. A., . . . Smith, M. T. (2008). Polymorphisms in the Estrogen Receptor 1 and Vitamin C and Matrix Metalloproteinase Gene Families Are Associated with Susceptibility to Lymphoma. *PLoS ONE, 3*(7), e2816. doi:10.1371/journal.pone.0002816

Szeto, Y. T., Tomlinson, B., & Benzie, I. F. F. (2002). Total antioxidant and ascorbic acid content of fresh fruits and vegetables: implications for dietary planning and food preservation. *British Journal of Nutrition, 87*(01), 55-59.

Timpson, N. J., Forouhi, N. G., Brion, M. J., Harbord, R. M., Cook, D. G., Johnson, P., . . . Davey Smith, G. (2010). Genetic variation at the SLC23A1 locus is associated with circulating concentrations of L-ascorbic acid (vitamin C): evidence from 5 independent studies with >15,000 participants. *Am J Clin Nutr, 92*(2), 375-382. doi:10.3945/ajcn.2010.29438

Tsukaguchi, H., Tokui, T., Mackenzie, B., Berger, U. V., Chen, X.-Z., Wang, Y., . . . Hediger, M. A. (1999). A family of mammalian Na+-dependent L-ascorbic acid transporters. *Nature, 399*(6731), 70-75. doi:http://www.nature.com/nature/journal/v399/n6731/suppinfo/399070a0_S1.html

Wang, H., Dutta, B., Huang, W., Devoe, L. D., Leibach, F. H., Ganapathy, V., & Prasad, P. D. (1999). Human Na(+)-dependent vitamin C transporter 1 (hSVCT1): primary structure, functional characteristics and evidence for a non-functional splice variant. *Biochim Biophys Acta, 1461*(1), 1-9.

Xia, M.-W., Cui, C.-B., Li, C.-W., & Wu, C.-J. (2014). Three New and Eleven Known Unusual C25 Steroids: Activated Production of Silent Metabolites in a Marine-Derived Fungus by Chemical Mutagenesis Strategy using Diethyl Sulphate. *Marine Drugs, 12*(3). doi:10.3390/md12031545

Appendices

Appendix A: SLC23A1 ORF sequence

ATGAGGGCCCAGGAGGACCTCGAGGGCCGGACACAGCATGAAACCACCAGG
GACCCCTCGACCCCGCTACCCACAGAGCCTAAGTTTGACATGTTGTACAAGA
TCGAGGACGTGCCACCTTGGTACCTGTGCATCCTGCTGGGCTTCCAGCACTAC
CTGACATGCTTCAGTGGTACCATCGCCGTGCCCTTCCTGCTGGCTGAGGCGCT
GTGTGTGGGCCACGACCAGCACATGGTTAGTCAGCTCATCGGCACCATCTTC
ACGTGCGTGGGCATCACCACTCTCATCCAGACCACCGTGGGCATCCGGCTGC
CGCTGTTCCAGGCCAGTGCCTTTGCATTTCTGGTTCCAGCCAAAGCCATACTG
GCTCTGGAGAGATGGAAATGCCCCCCGGAAGAGGAGATCTACGGTAACTGG
AGTCTGCCCCTGAACACCTCTCATATTTGGCACCCACGGATACGGGAGGTCC
AGGGTGCAATCATGGTGTCCAGCGTGGTGGAGGTGGTGATTGGCCTGCTGGG
GCTGCCTGGGGCCCTGCTCAACTACATTGGGCCTCTCACAGTCACCCCCACTG
TCTCCCTCATTGGCCTTTCTGTCTTCCAAGCTGCTGGCGACCGAGCTGGCTCC
CACTGGGGCATCTCAGCTTGCTCCATTCTCCTGATCATCCTCTTCTCCCAGTAC
CTGCGCAACCTCACCTTCCTGCTGCCTGTCTACCGCTGGGGCAAGGGCCTCAC
TCTCCTCCGCATCCAGATCTTCAAAATGTTTCCTATCATGCTGGCCATCATGA
CCGTGTGGCTGCTCTGCTATGTCCTGACCTTGACAGACGTGCTGCCCACAGAC
CCAAAAGCCTATGGCTTCCAGGCACGAACCGATGCCCGTGGTGACATCATGG
CTATTGCACCCTGGATCCGCATCCCCTACCCCTGTCAGTGGGGCCTGCCCACG
GTGACTGCGGCTGCTGTCCTGGGAATGTTCAGCGCCACTCTGGCAGGCATCAT
TGAGTCCATCGGAGATTACTACGCCTGTGCCCGCCTGGCTGGTGCACCACCCC
CTCCAGTACATGCTATCAACAGGGGCATCTTCACCGAAGGCATTTGCTGCATC
ATCGCGGGGCTATTGGGCACGGGCAACGGGTCCACCTCGTCCAGTCCCAACA
TTGGCGTCCTGGGAATTACCAAGGTGGGCAGCCGGCGCGTGGTGCAGTATGG
TGCGGCTATCATGCTGGTCCTGGGCACCATCGGCAAGTTCACGGCCCTCTTCG
CCTCGCTCCCTGACCCCATCCTGGGGGGCATGTTCTGCACTCTCTTTGGCATG
ATTACAGCTGTGGGGCTGTCCAACCTGCAATTTGTGGACATGAACTCCTCTCG

CAACCTCTTCGTGCTGGGATTTTCCATGTTCTTCGGGCTCACGCTGCCCAATT
ACCTGGAGTCCAACCCTGGCGCCATCAATACAGGCATTCTTGAAGTGGATCA
GATTCTGATTGTGCTGCTGACCACGGAGATGTTTGTGGGCGGGTGCCTTGCTT
TCATACTTGACAACACAGTGCCAGGGAGCCCAGAGGAGCGTGGTCTGATACA
GTGGAAAGCTGGGGCTCATGCCAACAGTGACATGTCTTCCAGCCTCAAGAGC
TACGATTTCCCCATTGGGATGGGCATAGTAAAAAGAATTACCTTTCTGAAATA
CATTCCTATCTGCCCAGTCTTCAAAGGATTTTCTTCAAGTTCAAAAGATCAGA
TTGCAATTCCAGAAGACACTCCAGAAAATACAGAAACTGCATCTGTGTGCAC
CAAGGTCTGA

Appendix B: pReceiver-M55 plasmid sequence

AACCCAGCTTTCTTGTACAAAGTGGTTCGATCTAGAATGGCTAGCCTCGAGTT
TAAACCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCC
GTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAA
TGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG
GGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG
CTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGG
CTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGT
GTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCG
CTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTC
AAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCAC
CTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGC
CCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGT
GGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTT
TGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGA
TTTAACAAAAATTTAACGCGAATTAATGCTGTGGAATGTGTGTCAGTTAGGGT
GTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCT
CAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGA
AGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAA
CTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCAT
GGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGA
GCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAA
AAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGAT
GAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGC
CGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGC
TGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTT
TGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCG
CGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACG
TTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCA
GGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTG

ATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCA
CCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTT
GTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAA
CTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGA
CCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCT
GGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAG
CGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCG
CTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCT
ATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGCGAAATGAC
CGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCC
TTCTATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGAT
CCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTA
TTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAA
TAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGT
ATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAT
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC
ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC
GTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTT
CGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCA
CAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA
AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG
CCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTT
CGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTG
TAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCG
ACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACAC
GACTTATCGCCACTGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTA
TGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACT

AGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAA
AAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGG
TTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAA
GATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACG
TTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTT
TAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTG
GTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTC
TATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATA
CGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCAC
GCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGA
GCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTT
GCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGT
TGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCAT
TCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGC
AAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGG
CCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATT
CTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGG
GATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAAC
GCTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCG
ATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAG
CGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAAT
AAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATT
GAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATT
TAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC
CTGACGTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGT
ACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGT
GTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAA
GGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCT
GCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTT
ATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTC

CGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACC
CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGG
GACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGG
CAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGAC
GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCC
TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT
TTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCC
AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAAC
GGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGG
TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGA
GAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGA
CCCAAGCTGGCTAGTTAAGCTTACCATGGTGAGCAAGGGCGAGGAGGATAAC
ATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCG
TGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACG
AGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGA
AGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTC
AAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCC
AGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGG
CACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGG
GAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAG
ATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTC
AAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAAC
GTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGG
AACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGC
TGTACAAG

Appendix C: Gel images of the first cite directed mutagenesis approach (Introduction of SNVs through amplification of the whole plasmid)

Gel images include the one PCR step for these primers combinations rs14116735, rs139127139, COSM117494, rs34054084, COSM 204877, COSM 30850, COSM 217333, TMP_ESP_5_138707912, COSM 20486 and rs6596474. The method used for this approach was described in details in the method section.

Figure19: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 7884 nucleotides for each of the primer combinations 1-(rs14116735 F1+R1), 2- (rs139127139 F1+R1), 3- (COSM117494 F1+R1) and 4-(rs34054084 F1+R1), in a PCR reaction containing Phusion HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1k.b DNA ladder on the sides of the image.

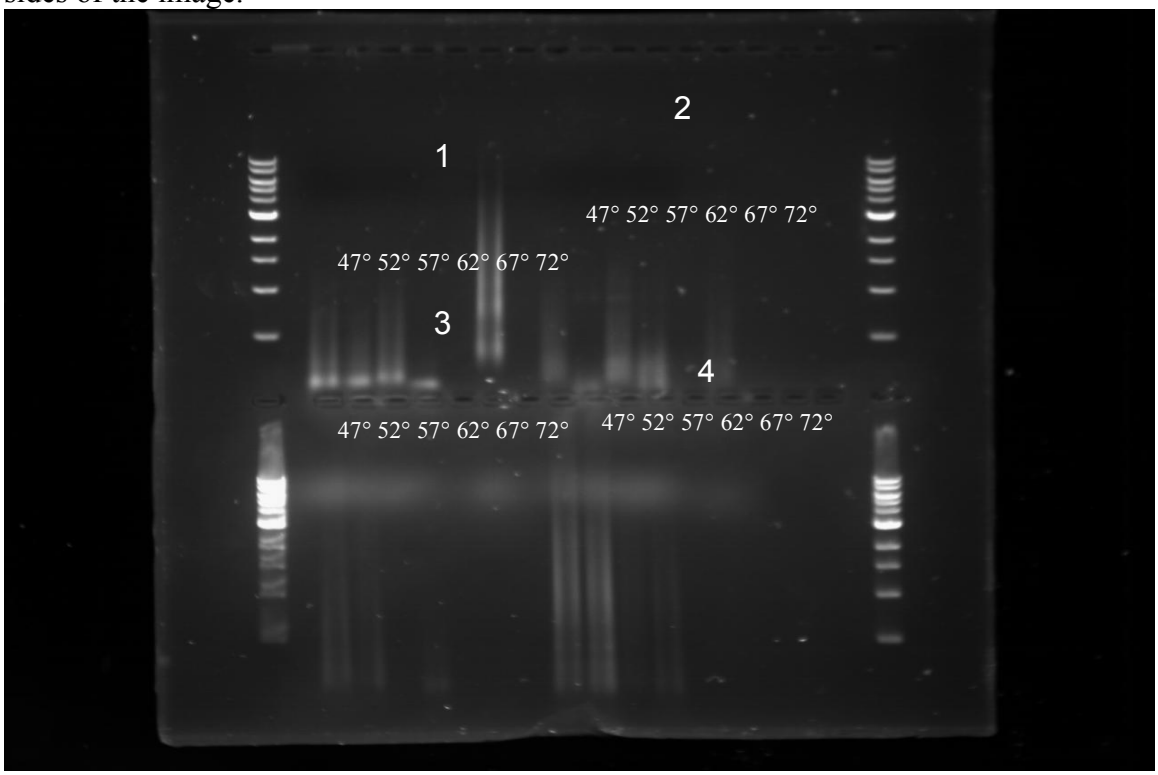


Figure 20: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 7884 nucleotides for each of the primer combinations 5- (COSM204877 F1+R1), 6- (COSM30850 F1+R1), 7- (COSM217333 F1+R1) and 8- (TMP_ESP_5_138707912 F1+R1) in a PCR reaction containing Phusion HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1k.b DNA ladder on the sides of the image.

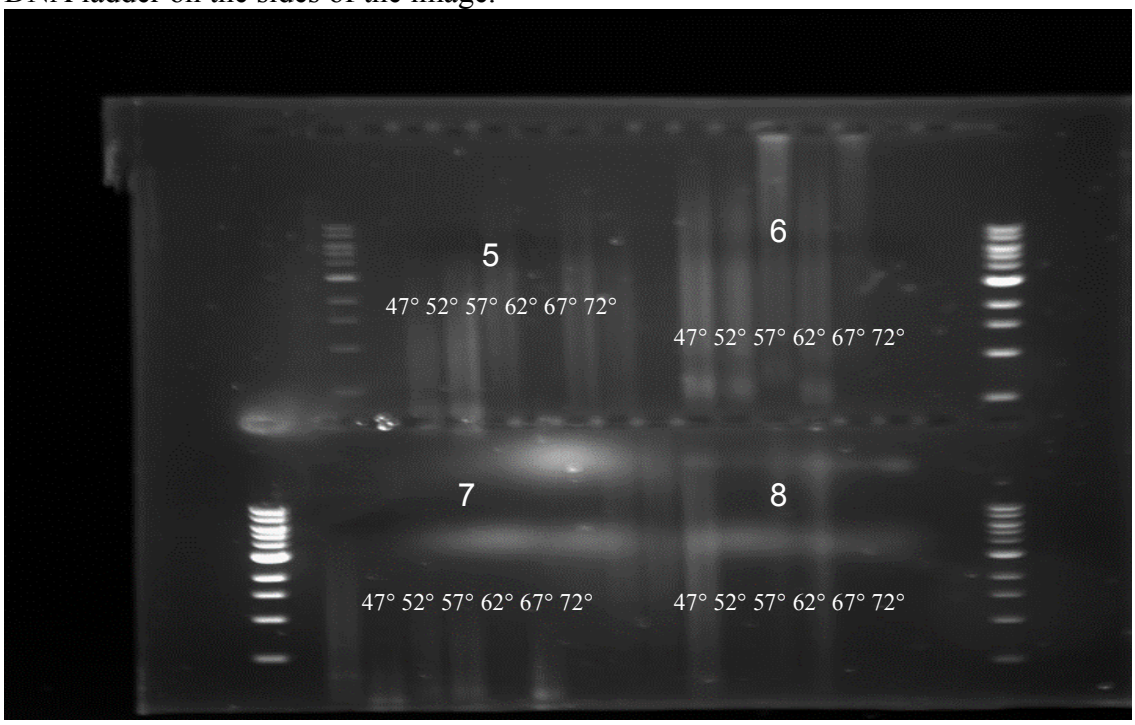
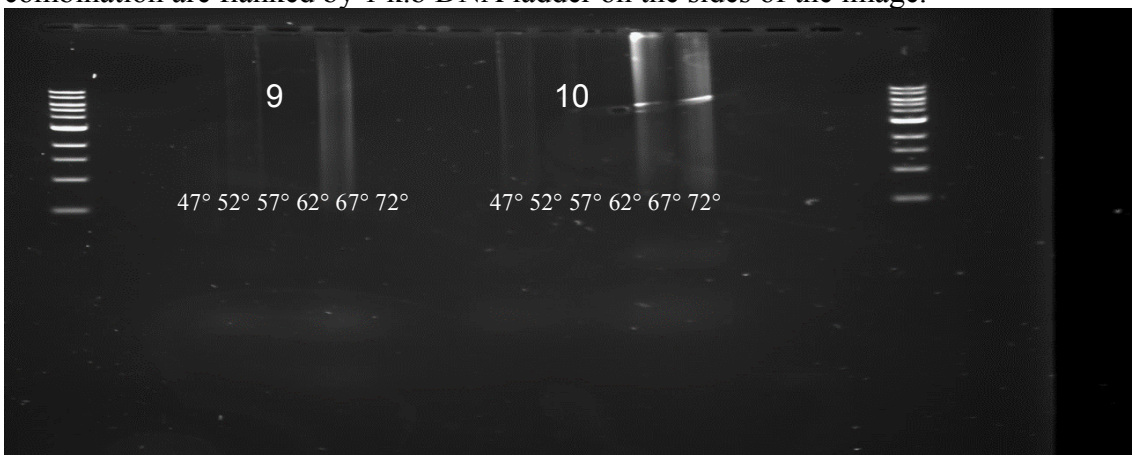


Figure 21: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 7884 nucleotides for each of the primer combinations 9- (COSM20486 F1+R1) and 10- (RS 6596474 F1+R1) in a PCR reaction containing Phusion HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 k.b DNA ladder on the sides of the image.



Since the Phusion did not work successfully, the Q5 HF enzyme was used.

Following the same protocol and PCR conditions. The amplification was not successful

as shown in the figures below:

Figure 22: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 7884 nucleotides for each of the primer combinations 1(rs14116735 F1+R1), 2- (rs139127139 F1+R1), 3-(COSM117494 F1+R1), 4-(rs34054084 F1+R1), 5- (COSM204877 F1+R1) in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 k.b DNA ladder on the sides of the image.

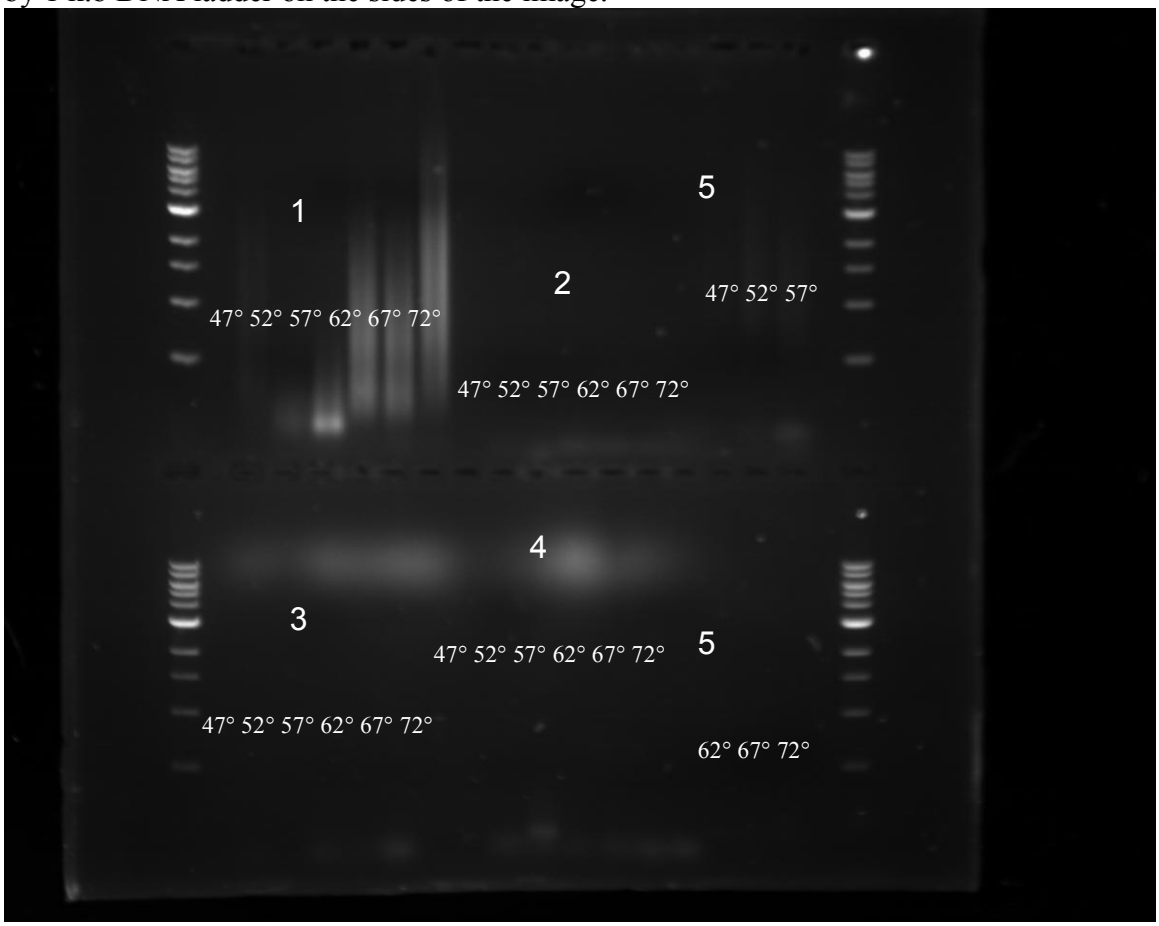
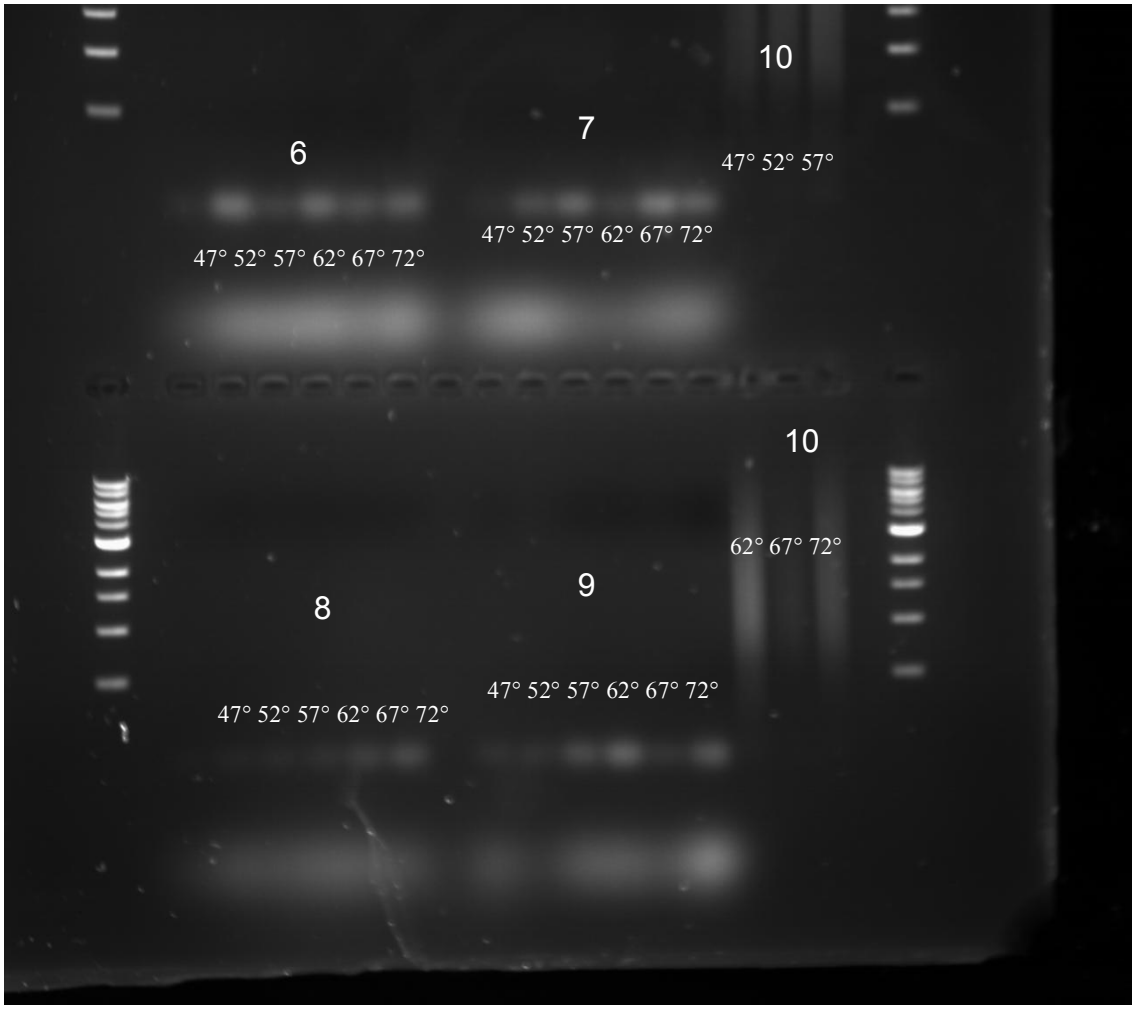


Figure 23: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 7884 nucleotides for each of the primer combinations COSM30850 F1+R1, COSM217333 F1+R1 and TMP_ESP_5_138707912 F1+R1, COSM20486 F1+R1 and RS 6596474 F1+R1 in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 k.b DNA ladder on the sides of the image.



Both enzymes Q5 and Phusion did not work successfully with all ten primer combinations. However, the Q5 enzyme was used for another trial primer combination rs14116735 F1+R1, rs139127139 F1+R1, COSM117494 F1+R1, rs34054084 F1+R1, COSM204877 F1+R1, COSM30850 F1+R1, COSM217333 F1+R1, TMP_ESP_5_138707912 F1+R1, COSM20486 F1+R1 and rs6596474 F1+R1, following

the same protocol and PCR conditions. The amplification was not successful as shown in the figures below:

Figure 24: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 7884 nucleotides for each of the primer combinations RS 14116735 F1+R1, RS 139127139 F1+R1, COSM117494 F1+R1, RS 34054084 F1+R1 and COSM204877 F1+R1 in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 k.b DNA ladder on the sides of the image.

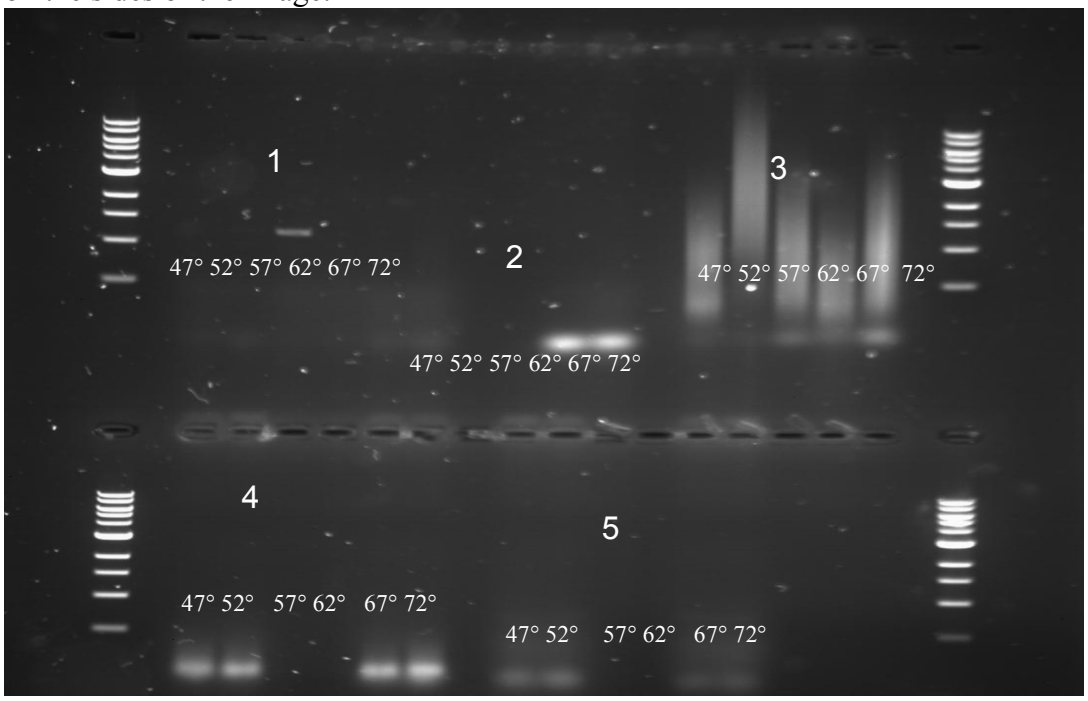
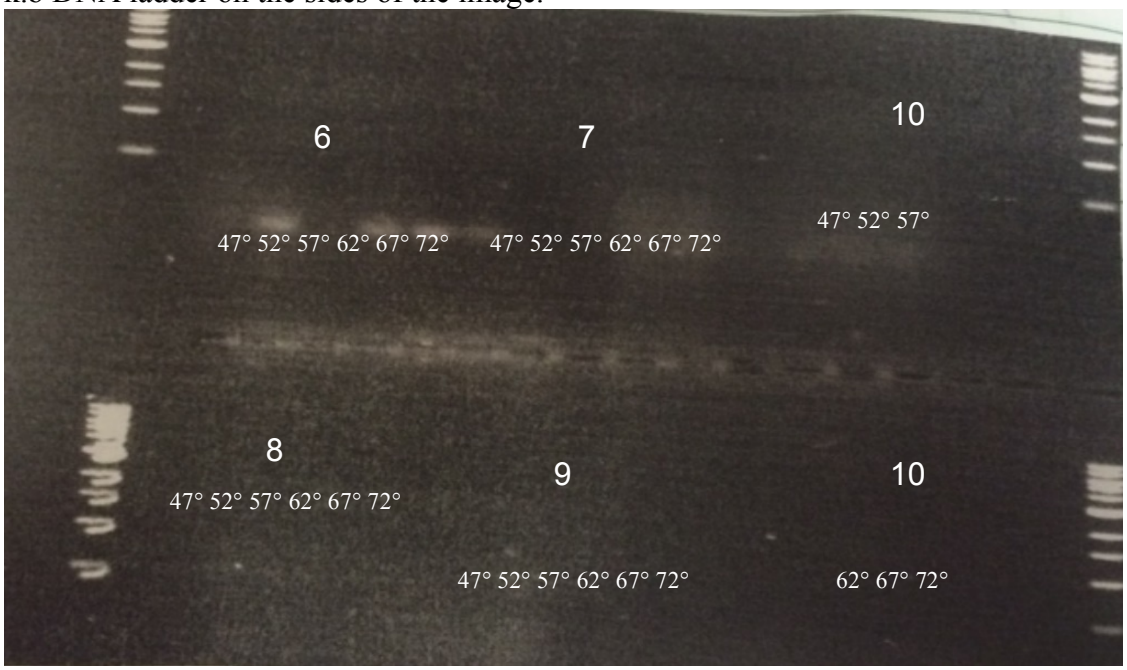


Figure 25: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 7884 nucleotides for each of the primer combinations COSM30850 F1+R1, COSM217333 F1+R1 and TMP_ESP_5_138707912 F1+R1, COSM20486 F1+R1 and RS 6596474 F1+R1 in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 k.b DNA ladder on the sides of the image.



Appendix D: The gel images for the second approach of PCR cite directed mutagenesis (Introduction of SNVs through amplification of two distinct amplicons for the 5' and 3' pertaining the respective SNV, with subsequent introduction of flanking Gateway® tags).

The same SNV targeted in the previous appendix C were introduced into the SLC23A1 ORF using the second approach described earlier in the method section. However, these gel images show the three PCR steps. In the following sections successes and failures for each SNV are described and visualized in the figures. Three polymerases, Phusion HF, Q5 HF and regular Taq polymerase, were utilized. Gel images include the three PCR steps for these primers combinations of rs14116735, rs139127139, COSM117494, rs34054084, COSM 204877, COSM 30850, COSM 217333, TMP_ESP_5_138707912, COSM 20486 and rs6596474.

Each of the following gel-images is showing individual lanes representing the PCR reaction of a gradient setup. The results for the amplification of individual SNV are described in the following sections.

Rs141167635

Rs141167635 is nonsense SNV, causing the gain of a stop codon, which alters the amino acid arginine 206 to a stop codon. It is located on the exon 6 on SLC23A1 gene.

The 649 nucleotides 5' amplicon introducing SNV rs141167635 was formed with the primers combination, RS141167635 R1+SLC23A1AttB1 F1 in a Phusion HF containing reaction (Figure 26 and Figure 27).

Figure 26: Gel image of the gradient PCR for the 649 nucleotides amplicon for the primer combination rs14116735 R1+ SLC23A1AttB1 F1. Amplification across all temperatures, with strong signals at all temperatures, except for 62°C, was achieved.

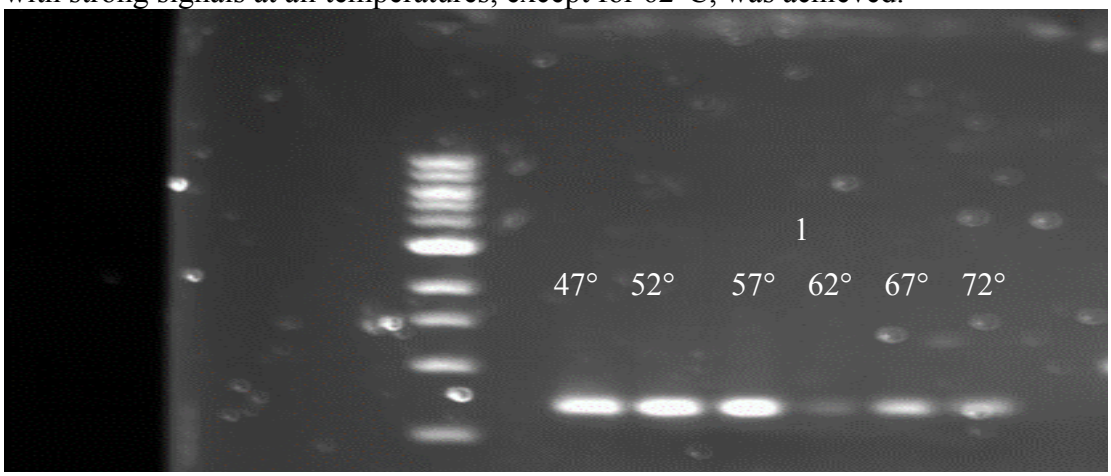
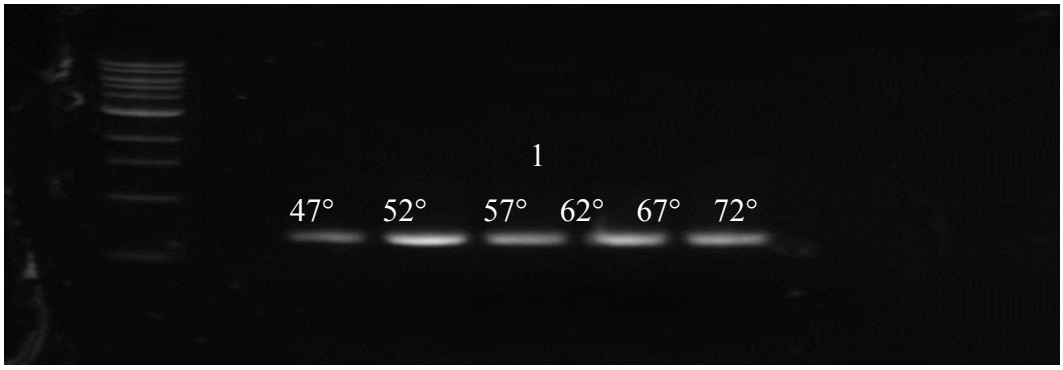


Figure 27: Gel image of the gradient PCR for the 649nucleotide amplicon for the primer combination rs14116735 R1+ SLC23A1AttB1 F1 using Phusion HF enzyme, amplification across all temperatures, except for 72°C, with the strongest signal at 50°C was achieved.



The 1212 nucleotide 3' amplicon introducing SNV rs141167635 was formed with the primers combination, rs141167635 F1 + SLC23A1AttB1 R1 in a Phusion HF containing reaction (Figure 28 and Figure 29).

Figure 28: Gel image of the gradient PCR for the 1212 nucleotide amplicon for the primer combination rs14116735 F1+ SLC23A1AttB1 R1 using Phusin HF enzyme, successful amplification across all temperatures, with the strong signal at all temperatures was achieved.

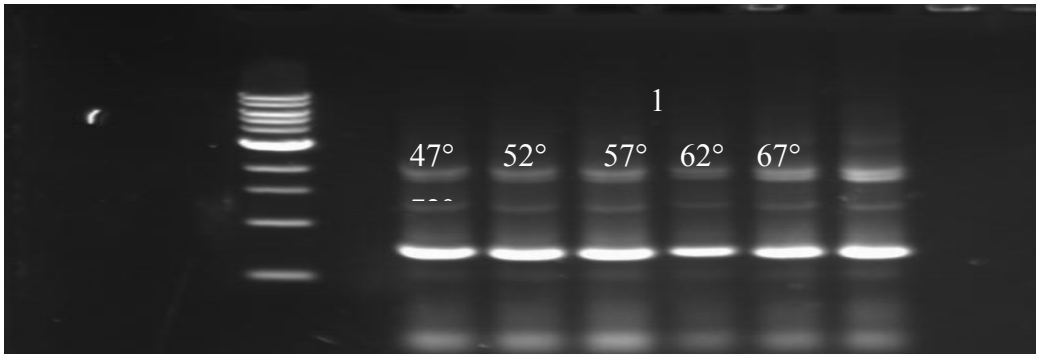
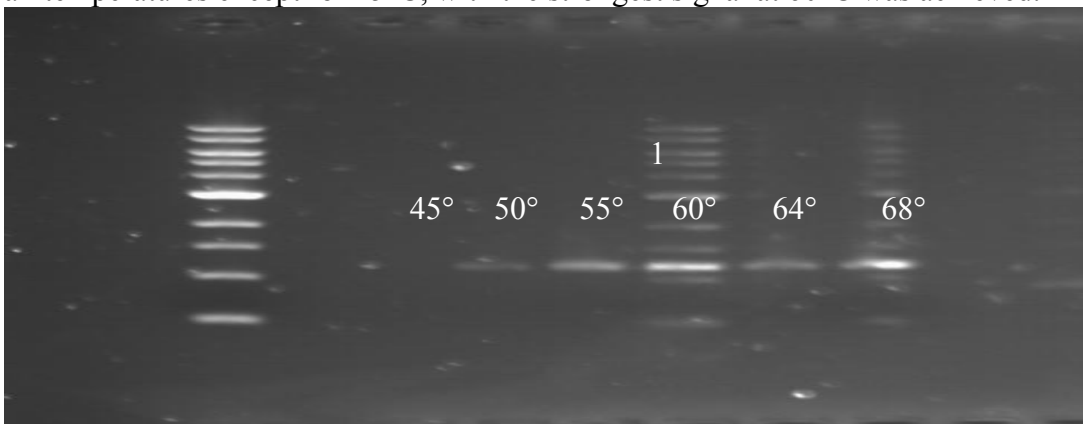


Figure 29: Gel image of the gradient PCR for the 1212 nucleotide amplicon for the primer combination rs14116735 F1+ SLC23A1AttB1 R1, successful amplification across all temperatures except for 45°C, with the strongest signal at 60°C was achieved.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. Although some attempts failed (Figure 30 and Figure 31) an amplicon of the expected 1861 nucleotides was produced at 67°C in a reaction containing Q5 HF polymerase

(Figure 31). However, the product did not show the correct size after the PCR purification.

Figure 30: Agarose gel to visualize PCR amplicons of the entire pReciver – M55 plasmid, the expected size is 1797 nucleotides for each of the primer combinations (rs14116735 R1 +SLC23A1AttB1 F1) + (rs14116735 F1 +SLC23A1AttB1 R1) in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 k.b DNA ladder on the sides of the image.

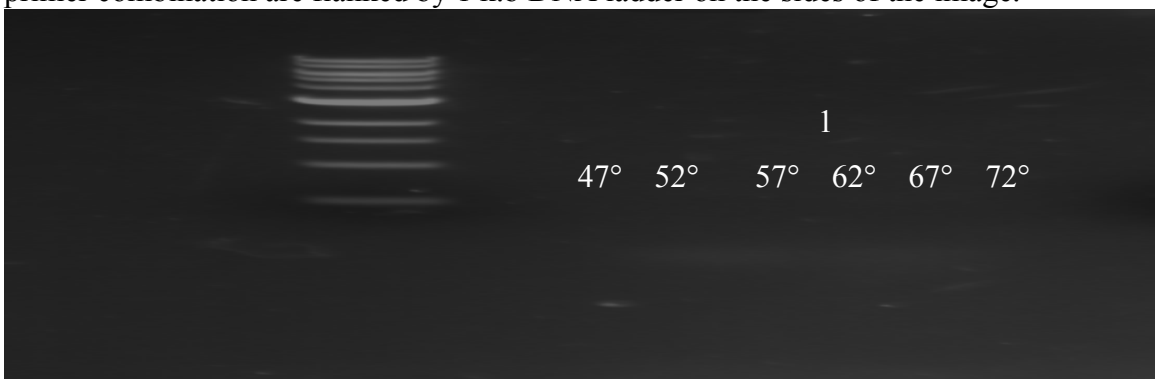
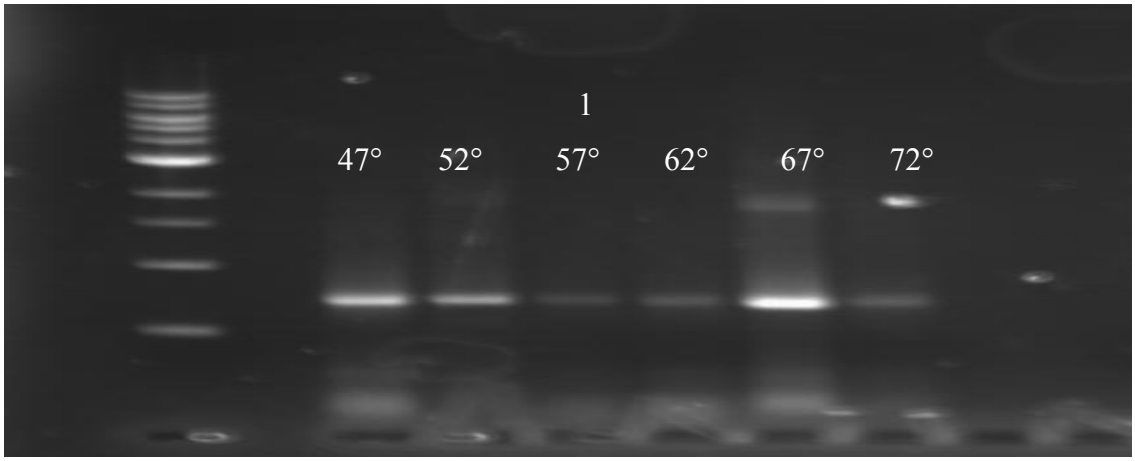


Figure 31: Gel image of the gradient PCR for the 1861 nucleotide amplicon for the ligation of the primer combination (rs14116735 R1 +SLC23A1AttB1 F1) and (rs14116735 F1 +SLC23A1AttB1 R1), using Q5 HF enzyme. It shows only one faint band at 67°C.



Rs139127139

This is nonsense SNV, causing a gain stop codon, which signal for a stop codon at the amino acid tryptophan (W). It is located on the exon 9 on SLC23A1 gene, at amino acid number 311.

The 965 nucleotide 5' amplicon introducing SNV rs139127139 was formed with the primers combination, RS 139127139 R1 + SLC23A1AttB1 F1 in a regular Taq (Figure 32) and Phusion HF containing reaction (Figure 33)

Figure 32: Gel image of the gradient PCR for the 965 nucleotide amplicon for the primer combination RS 139127139 R1 + SLC23A1AttB1 F1 using Taq enzyme, successful amplification across all temperatures with the strongest signal at 45°C and 50°C was achieved.

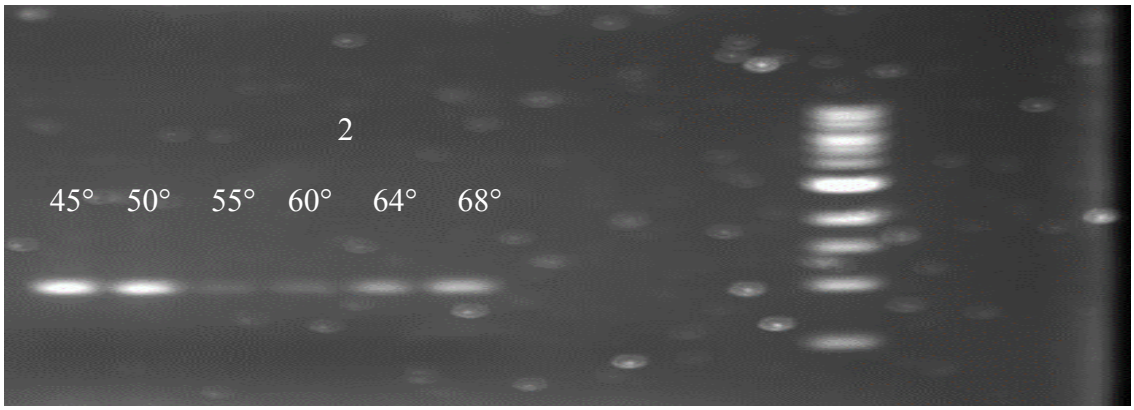
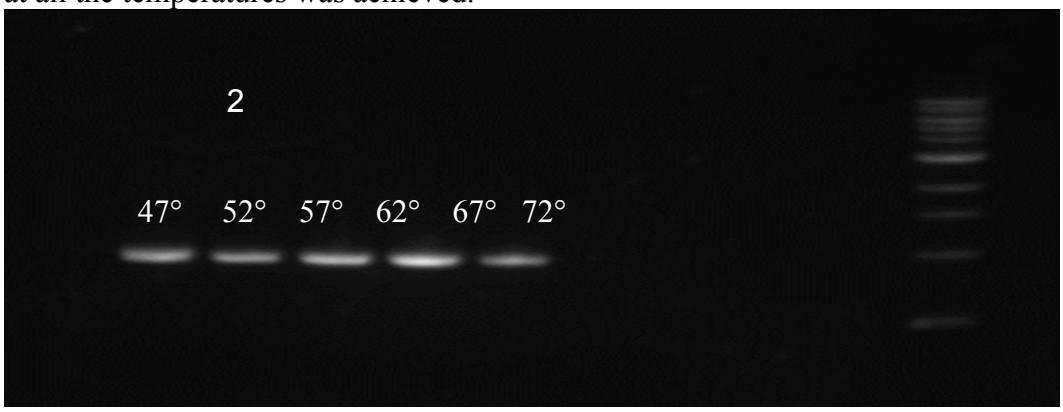


Figure 33: Gel image of the gradient PCR for the 965 nucleotide amplicon for the primer combination RS 139127139 R1 + SLC23A1AttB1 F1 using Phusion HF enzyme, successful amplification across all temperatures, except for 72°C with the strong signals at all the temperatures was achieved.



The 896 nucleotide 3' amplicon introducing SNV rs141167635 was formed with the primers combination, RS 139127139 F1 + SLC23A1AttB1 R1 in a regular Taq polymerases (

Figure 34) and Phusion HF containing reaction (Figure 35).

Figure 34: Gel image of the gradient PCR for the 896 nucleotide amplicon for the primer combination RS 139127139 F1+ SLC23A1AttB1 R1 using Taq enzyme, successful amplification across all temperatures was achieved.

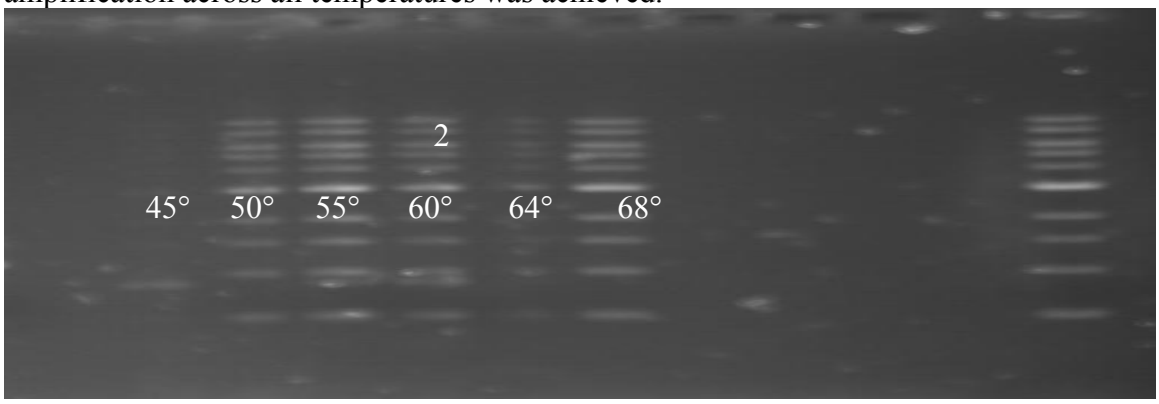
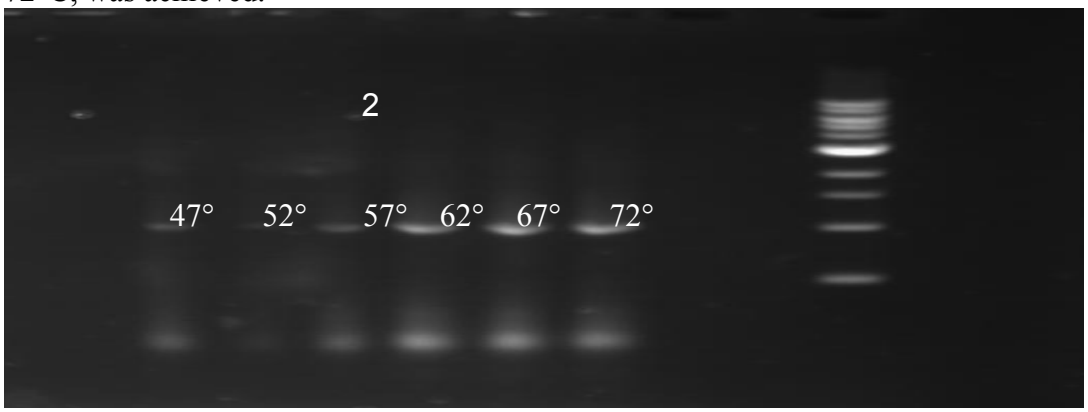


Figure 35: Gel image of the gradient PCR for the 896 nucleotide amplicon for the primer combination RS 139127139 F1 + SLC23A1AttB1 R1 using Phusin HF enzyme, successful amplification across all temperatures with strong signals at 62°C, 67°C and 72°C, was achieved.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was either not obtained (Figure 36) or it was at the incorrect size (Figure 37).

Figure 36: Agarose gel to visualize PCR amplicon resulted of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (Rs139127139 R1 + SLC23A1AttB1 F1) and (RS139127139F1 +SLC23A1AttB1 R1), in a PCR reaction containing regular Taq enzyme. None of the reactions yielded an amplicon at the correct size. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

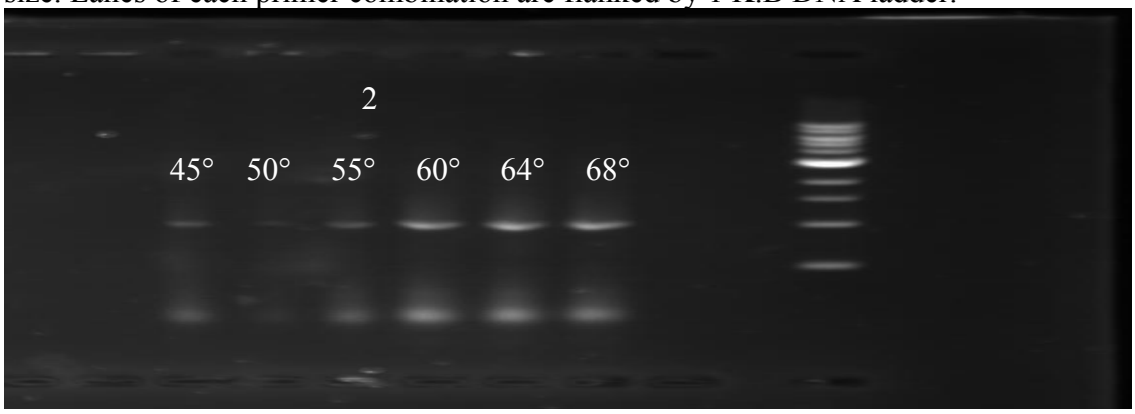
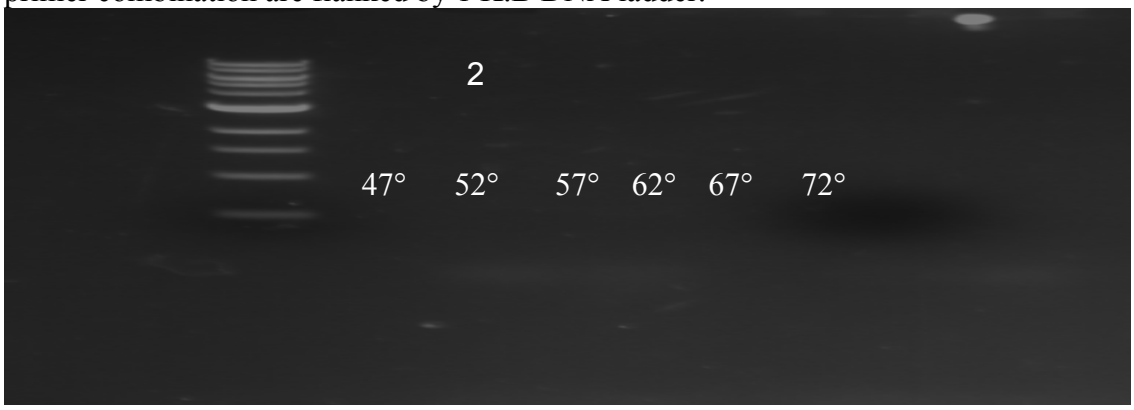


Figure 37: Agarose gel to visualize PCR amplicon resulting of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (Rs139127139 R1 + SLC23A1AttB1 F1) and (RS139127139F1 +SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF enzyme. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.



COSM117494

This is missense SNV that changes the amino acid from aspartic acid (D) to glutamic acid (E). It is located on exon 1 on the SLC23A1 gene, at amino acid number 6.

The 51 nucleotide 5' amplicon introducing SNV COSM117494 was attempted with the primers combination, COSM117494 R1 + SLC23A1AttB1 F1 in a regular Taq

and Phusion HF containing reaction, the attempt was not successful as gel image did not show the correct amplicon size (Figure 38 and Figure 39). Therefore, another attempt was done in Phusion HF containg reaction. In this trial the 51 nucleotides amplicon was obtained successfully (Figure 40).

Figure 38: Agarose gel to visualize PCR amplicons of, the primer combination COSM117494 R1 +SLC23A1AttB1 F1 in a PCR reaction containing regular Taq. The expected size is 51 nucleotides, however, none of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

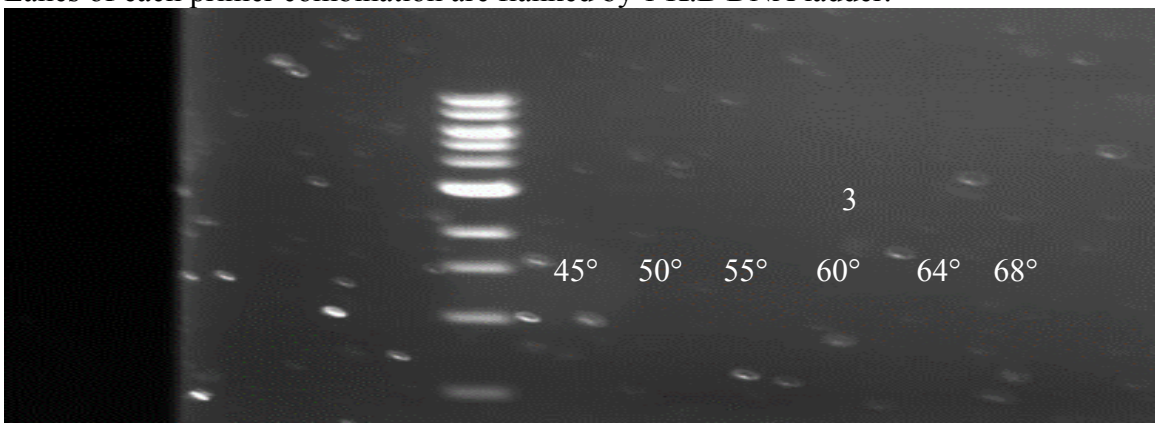
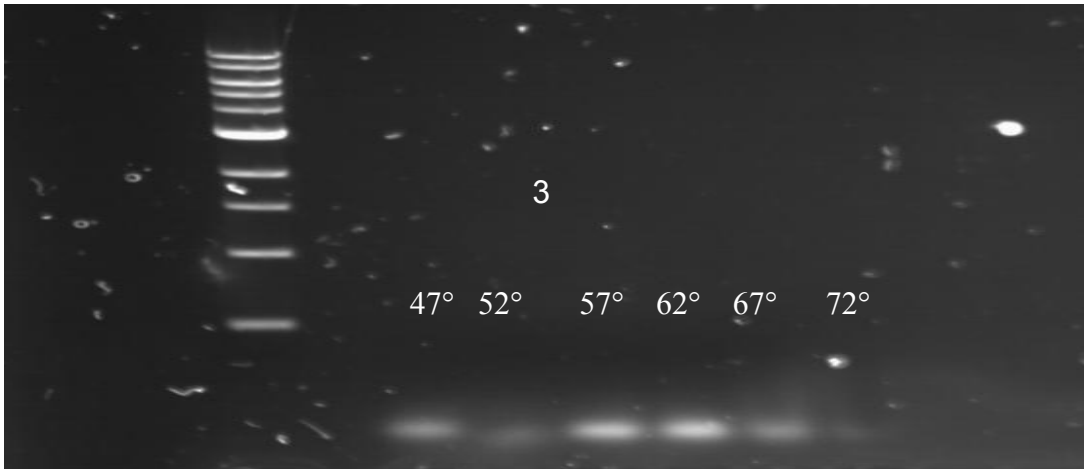


Figure 39: Agarose gel to visualize PCR amplicons of, the primer combination COSM117494 R1 +SLC23A1AttB1 F1 in a PCR reaction containing Phusion HF. The expected size is 51 nucleotides, however, none of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

Figure 40: Gel image of the gradient PCR for the 51 nucleotide amplicon for the primer combination COSM117494 R1 + SLC23A1AttB1 F1 using Phusion HF enzyme, successful amplification across all temperatures with strong signals at 57°C and 62°C was achieved.



The 1810 nucleotide 3' amplicon introducing SNV COSM117494 was attempted with the primers combination, COSM117494 F1 + SLC23A1AttB1 R1 in a regular Taq polymerases and Phusion HF containing reaction. The attempt was not successful in a regular Taq containing reaction (Figure 41) but the 1810 nucleotides amplicon was successfully obtained in a Phusion HF reaction (Figure 42).

Figure 41: Agarose gel to visualize PCR amplicons for the primer combination COSM117494 F1 + SLC23A1AttB1 R1 in a PCR reaction containing regular Taq. The expected size is 1861.nrcloetides, however, none of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder

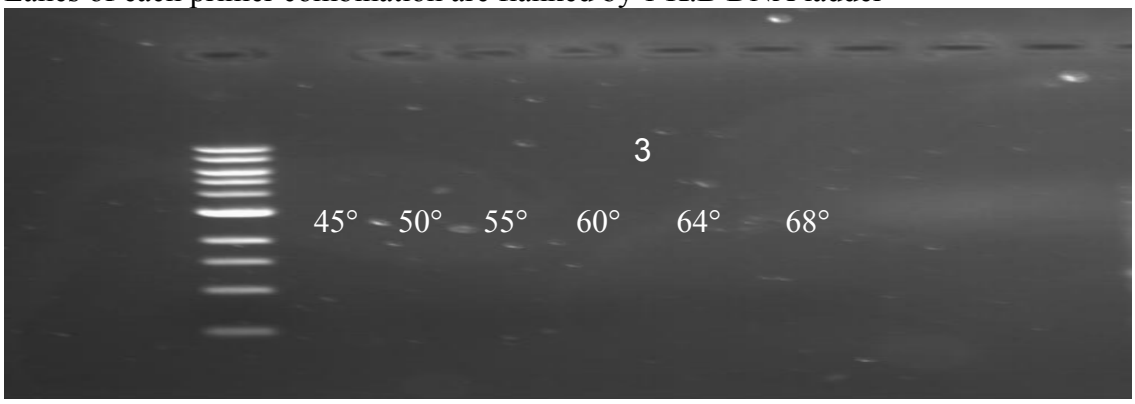
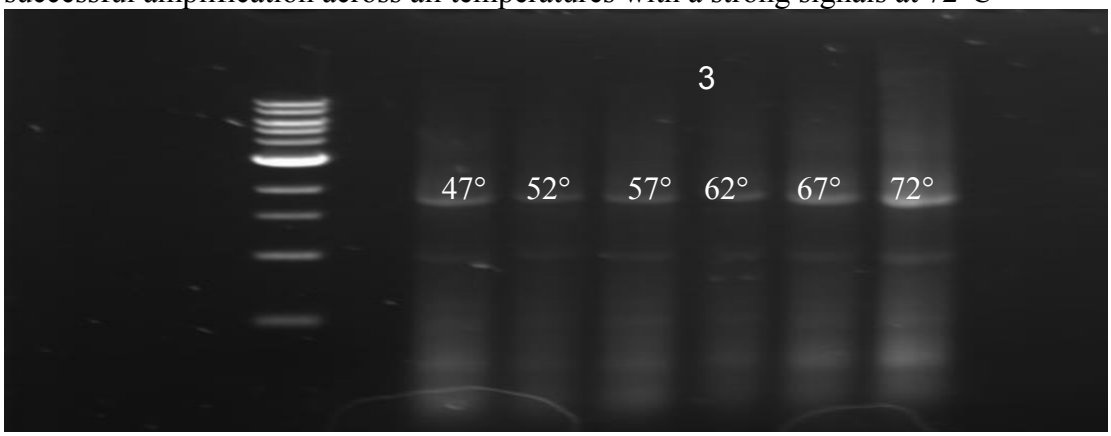


Figure 42: Gel image of the gradient PCR for the 1810 nucleotide amplicon for the primer combination COSM117494 F1 + SLC23A1AttB1 R1 using Phusion HF enzyme, successful amplification across all temperatures with a strong signals at 72°C



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and a regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was either not obtained or it was at the incorrect size (Figure 43 and Figure 44).

Figure 43: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combination (COSM117494 R1 +SLC23A1AttB1 F1) and (COSM117494 F1 +SLC23A1AttB1 R1in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

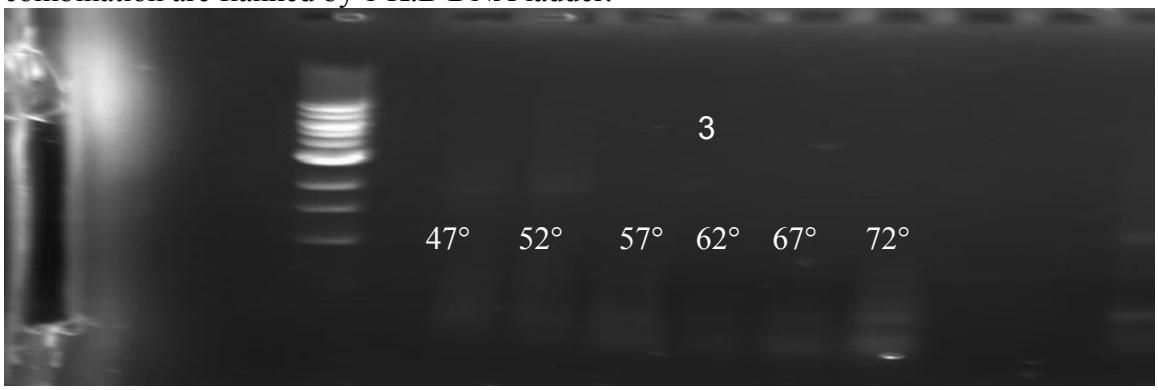
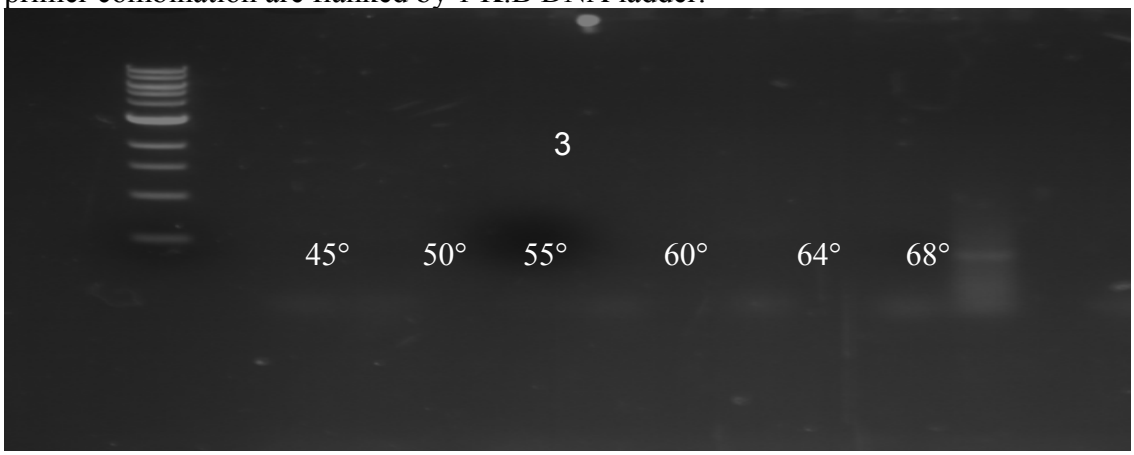


Figure 44: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combination (COSM117494 R1 +SLC23A1AttB1 F1) and (COSM117494 F1 +SLC23A1AttB1 R1 in a PCR reaction containing a regular Taq. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

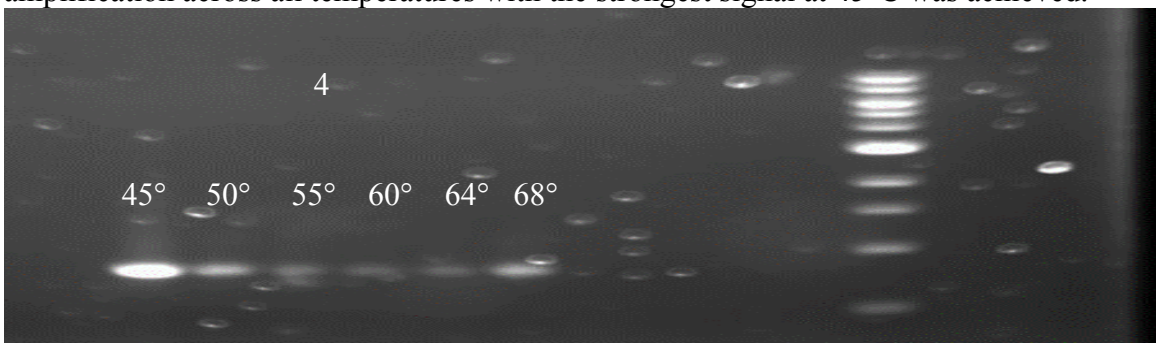


Rs34054084

This is a Frameshift SNV, located on exon 8 on the SLC23A1 gene, at amino acid number 264.

The 825 nucleotide 5' amplicon introducing SNV rs34054084 was formed with the primers combination, rs34054084 R1 + SLC23A1AttB1 F1, in a regular Taq containing reaction (Figure 45).

Figure 45: Gel image of the gradient PCR for the 825 nucleotide amplicon for primer combination rs34054084 R1+ SLC23A1AttB1 F1 using Taq enzyme, successful amplification across all temperatures with the strongest signal at 45°C was achieved.



The1036 nucleotide 3' amplicon introducing SNV rs141167635 was formed with the primers combination, rs34054084 F1 + SLC23A1AttB1 R1 first trial was in a regular Taq polymerases, and the amplicon was obtained successfully only at three temperatures, 45°C, 50°C, 55°C, 60°C and 64°C (Figure 46). However, in Phusion HF containing reaction, the amplicon was obtained at all temperatures (

Figure 47).

Figure 46: Gel image of the gradient PCR for the 1036 nucleotide amplicon for primer combination rs34054084 F1+ SLC23A1AttB1 R1 using Taq enzyme, successful amplification at 45°C, 50°C, 55°C, 60°C and 64°C was achieved.

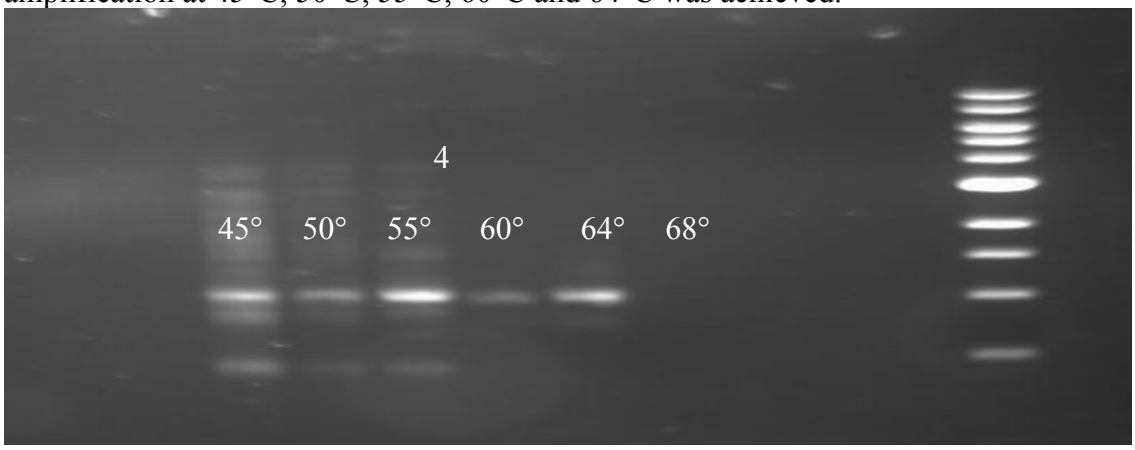
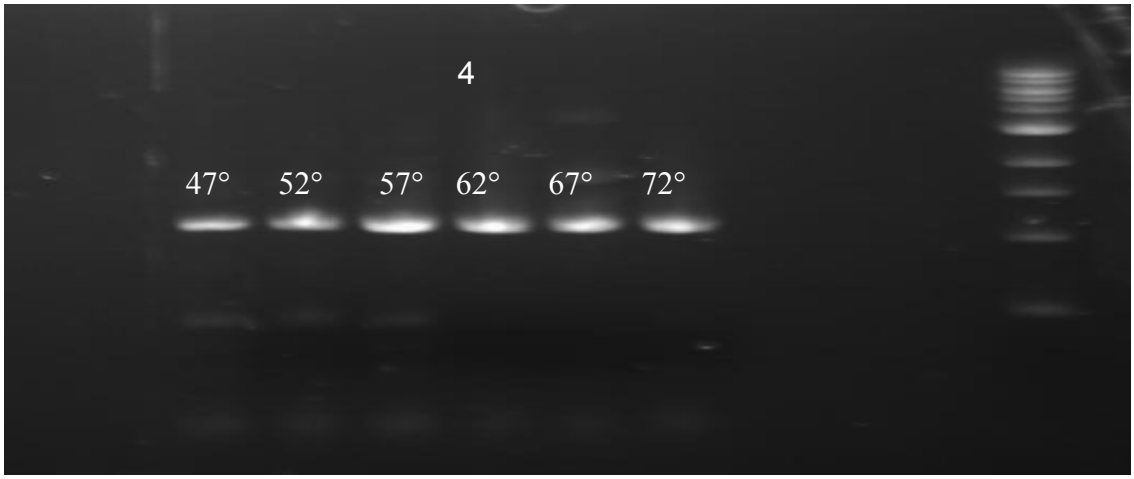


Figure 47: Gel image of the gradient PCR for the 1036 nucleotide amplicon for the primer combination rs34054084 F1 + SLC23A1AttB1 R1 using Phusion HF enzyme, successful amplification across all temperatures with a strong signals at 67°C was achieved.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was either not obtained (Figure 48) or it was at the incorrect size (

Figure 49).

Figure 48: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (RS 34054084 R1 + SLC23A1AttB1 F1) and (rs34054084 F1 +SLC23A1AttB1 R1), in a PCR reaction containing regular Taq. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

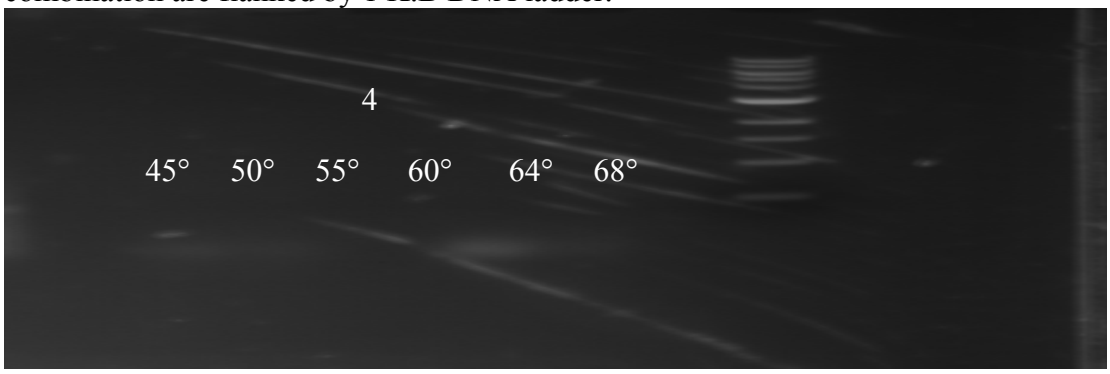
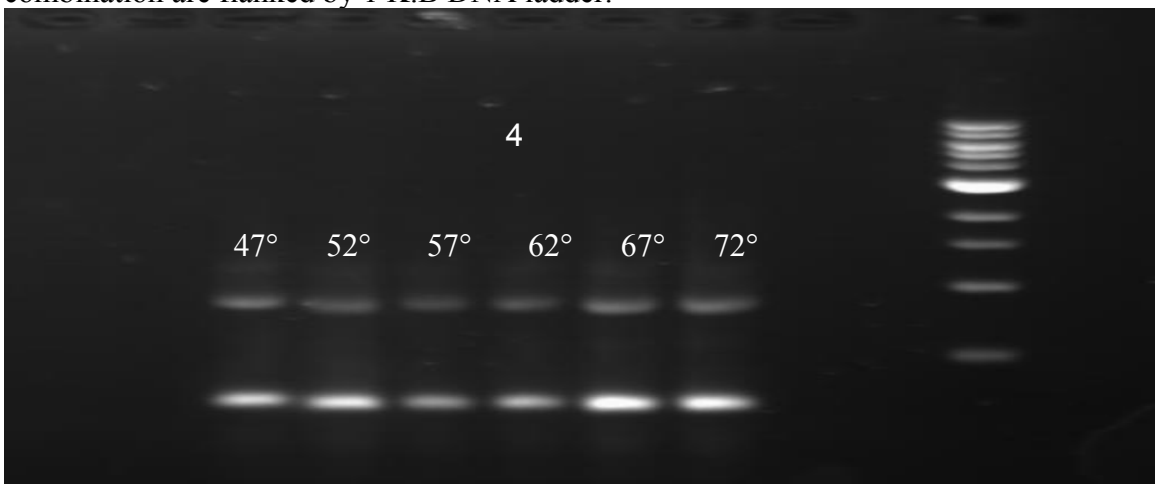


Figure 49: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (rs34054084 R1 + SLC23A1AttB1 F1) and (rs34054084 F1 +SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.



COSM204877

This is a missense SNV that changes the amino acid from leucine (L) to methionine (M). It is located on exon 3 on the SLC23A1 gene, at amino acid number 70.

The 241 nucleotide 5' amplicon introducing SNV COSM204877 was formed with the primers combination, COSM204877 R1 + SLC23A1AttB1 F1, in a regular Taq (Figure 50) and phusion HF containing reaction (Figure 51)

Figure 50: Gel image of the gradient PCR for first the 241 nucleotide amplicon for the primer combination COSM204877 R1+ SLC23A1AttB1 F1 using Taq enzyme, successful amplification across all temperatures, with the strongest signal at 55°C was achieved.

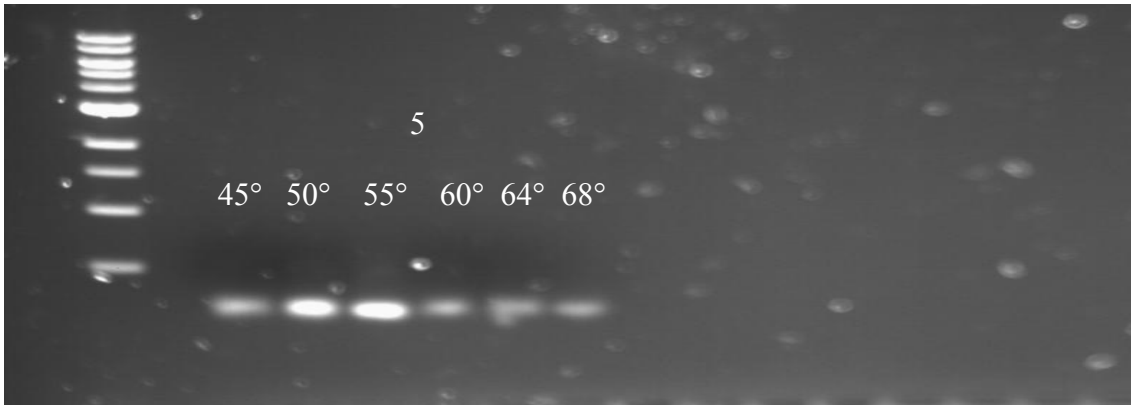
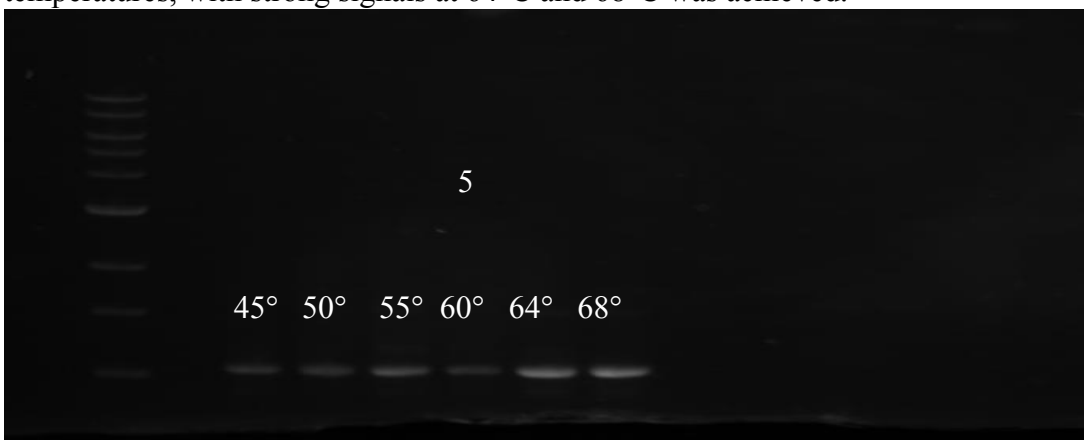


Figure 51: Gel image of the gradient PCR for the 241 nucleotide amplicon for the primer combination COSM204877 R1+ SLC23A1AttB1 F1, successful amplification across all temperatures, with strong signals at 64°C and 68°C was achieved.



The1620 nucleotide 3' amplicon introducing SNV COSM204877 was formed with the primers combination, COSM204877 F1 + SLC23A1AttB1 R1 in a regular Taq polymerases (Figure 52) and in Phusion HF containing reaction (Figure 53).

Figure 52: Gel image of the gradient PCR for the 1620 nucleotide amplicon for the primer combination COSM204877 F1+ SLC23A1ATTB1 R1 using Taq enzyme, successful amplification across all temperatures, with the strongest signal at 60°C was achieved. The DNA ladder is not shown in the image due to the gel plate condition used in the trial.

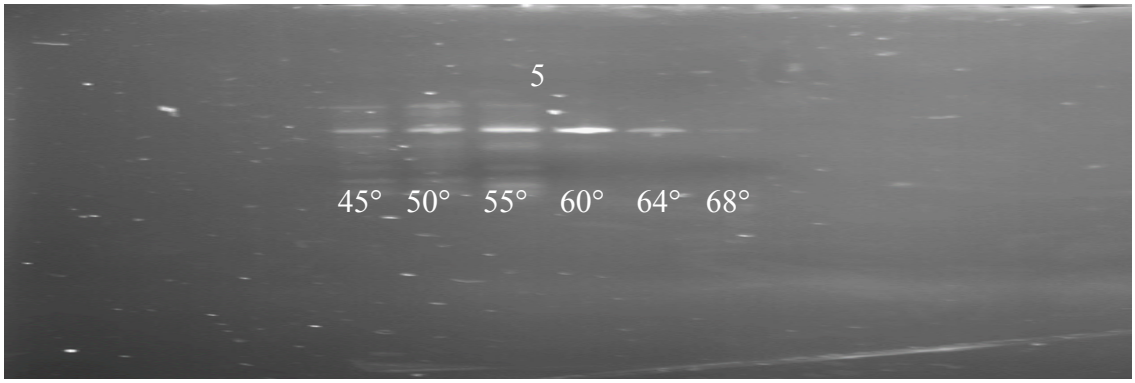
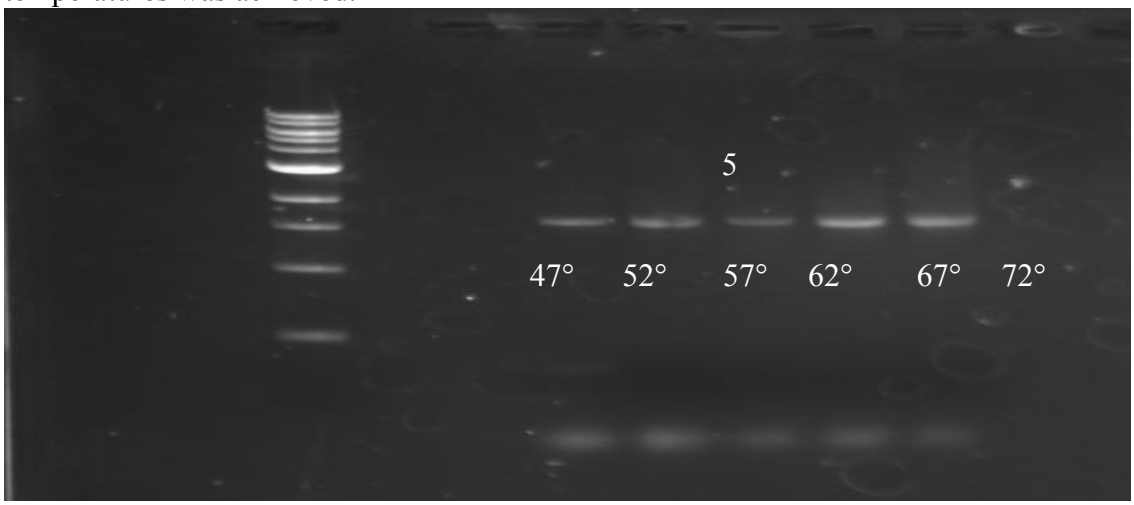


Figure 53: Gel image of the gradient PCR for the 1620 nucleotide amplicon for the primer combination COSM204877 F1+ SLC23A1ATTB1 R1 using Phusion HF enzyme, successful amplification across all temperatures except for 72°C, with the strongest at all temperatures was achieved.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases.

The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was either it was at the incorrect size (Figure 54) or not obtained (

Figure 55).

Figure 54: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (COSM204877 R1 +SLC23A1AttB1 F1) and (COSM204877 F1 +SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon at the correct size. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

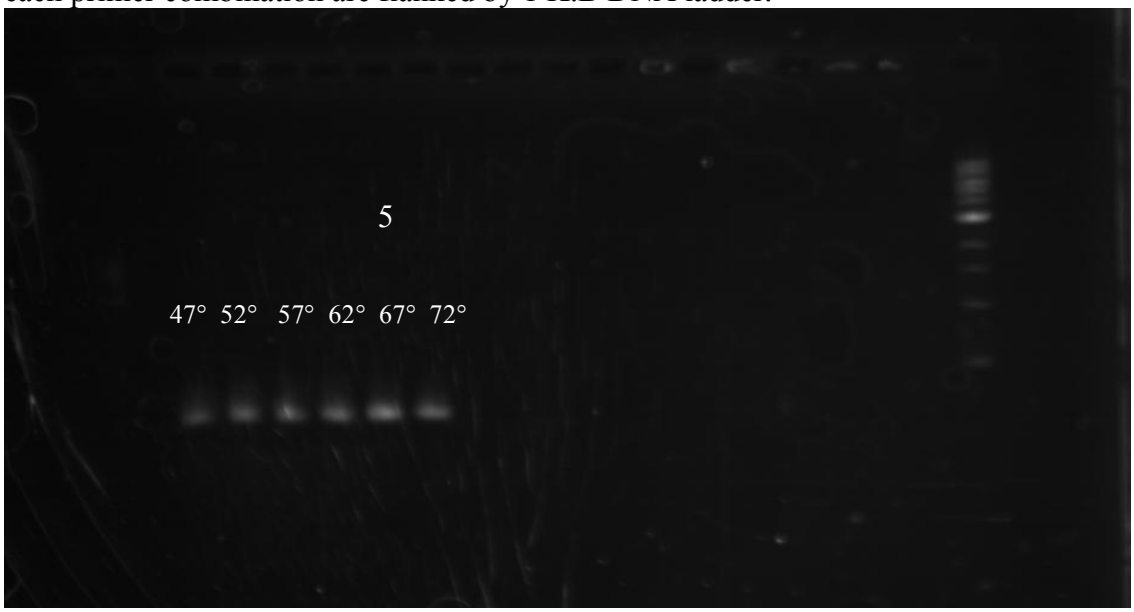
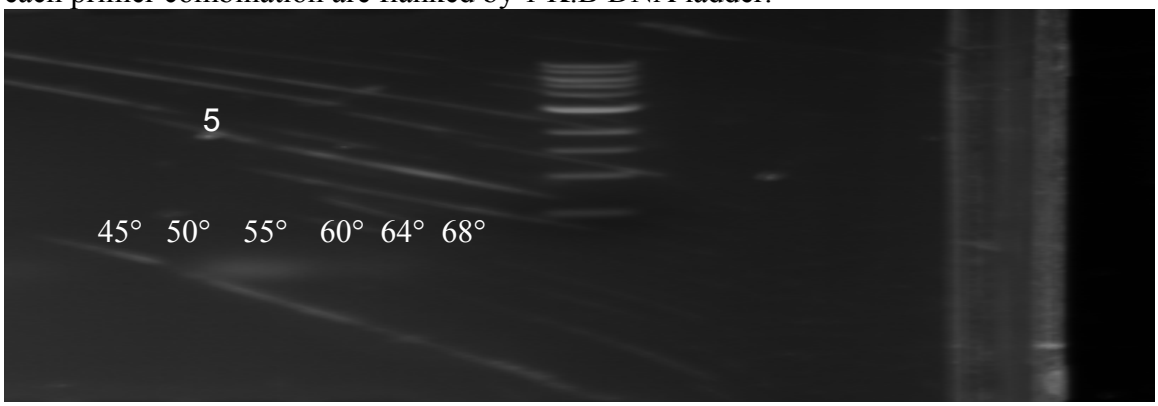


Figure 55:Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (COSM204877 R1 +SLC23A1AttB1 F1) and (COSM204877 F1 +SLC23A1AttB1 R1), in a PCR reaction containing regular Taq polymerase. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.



COSM30850

This is a missense SNV that changes the amino acid from glutamine (Q) to leucine (L). It is located on exon 3 on the SLC23A1 gene, at amino acid number 76.

The 260 nucleotide 5' amplicon introducing SNV COSM30850 was formed with the primers combination, COSM30850 R1 + SLC23A1AttB1 F1, in a regular Taq and phusion HF containing reaction (Figure 56 and Figure 57).

Figure 56: Gel image of the gradient PCR for the 260 nucleotide amplicon for the primer combination COSM30850 R1+ SLC23A1AttB1 F1 using regular Taq polymerase, successful amplification across all temperatures, with strong signals at all temperatures was achieved.

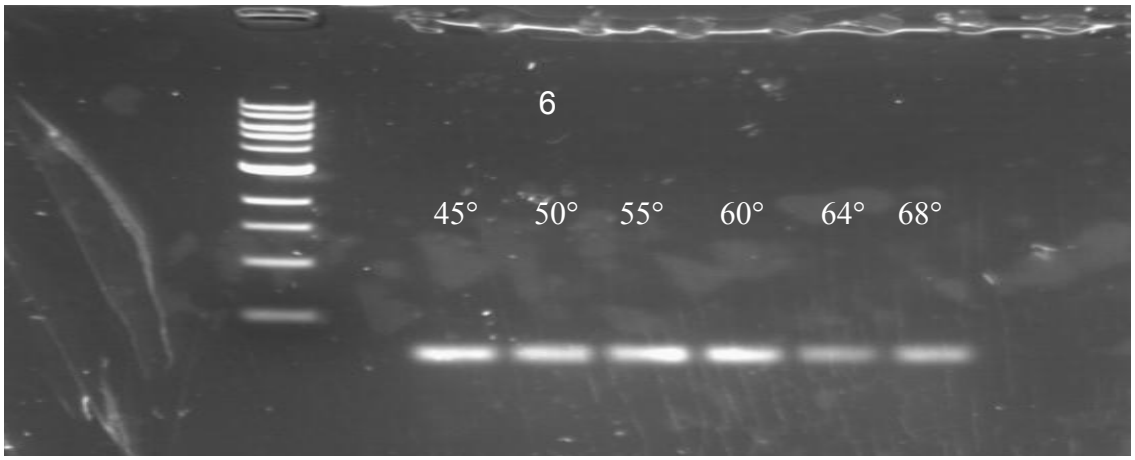
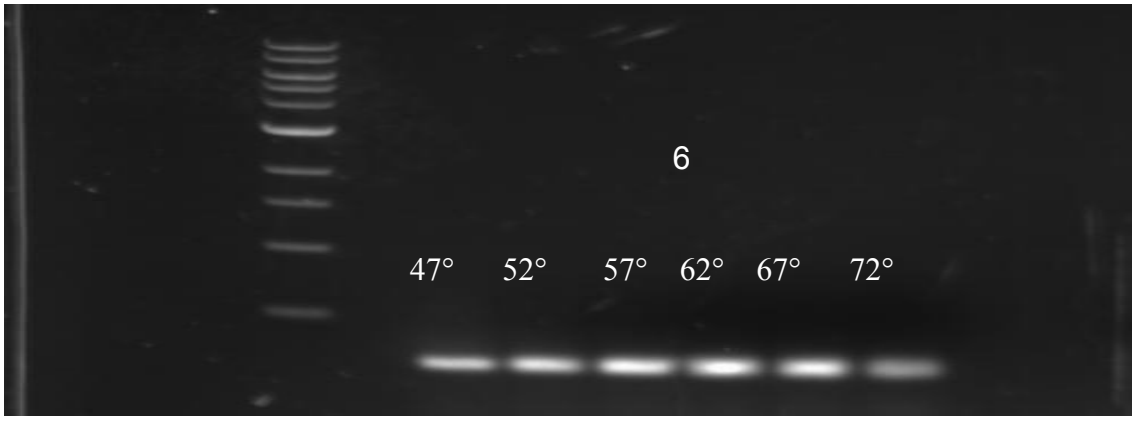
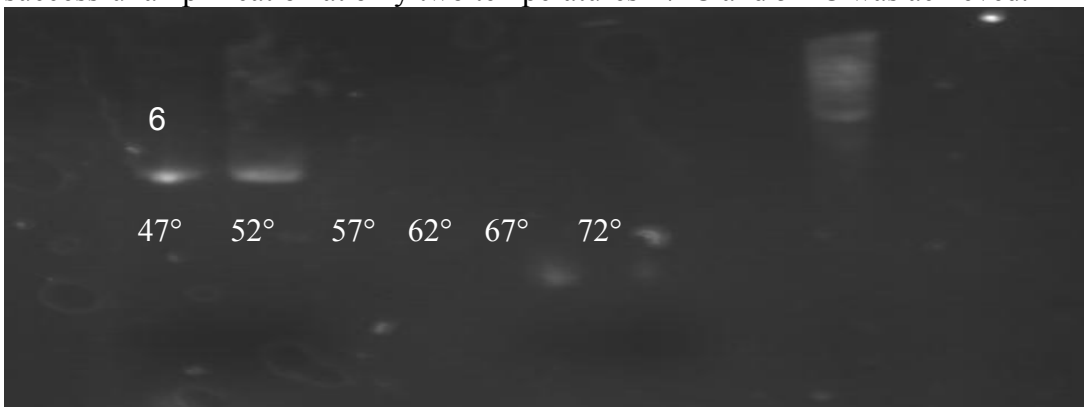


Figure 57: Gel image of the gradient PCR for the 260 nucleotide amplicon for the primer combination COSM30850 R1+ SLC23A1AttB1 F1using Phusion HF enzyme, successful amplification across all temperatures, with the strong signal at all temperatures was achieved.



The1601 nucleotide 3' amplicon introducing SNV COSM30850 was formed with the primers combination, COSM30850 F1 + SLC23A1AttB1 R1 in Phusion HF containing reaction, and the amplicon was successfully amplified only at two temperatures 47°C and 52°C (Figure 58).

Figure 58: Gel image of the gradient PCR for the 1601 nucleotide amplicon for the primer combination COSM30850 F1+ SLC23A1AttB1 R1using Phusion HF enzyme, successful amplification at only two temperatures 47°C and 52°C was achieved.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was not obtained (Figure 59 and Figure 60).

Figure 59: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (Cosm30850 R1 +SLC23A1AttB1 F1) and (COSM30850 F1 +SLC23A1AttB1 R1), in a PCR reaction containing regular Taq polymerase. None of the reactions yielded an amplicon at the correct size. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

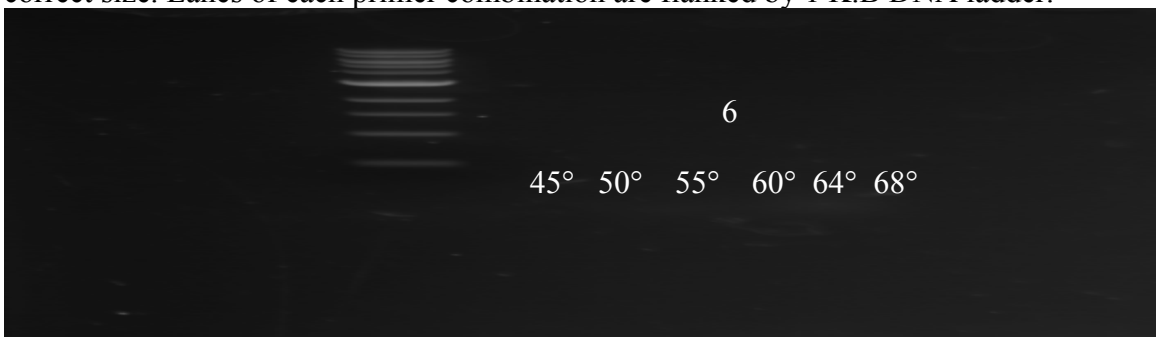
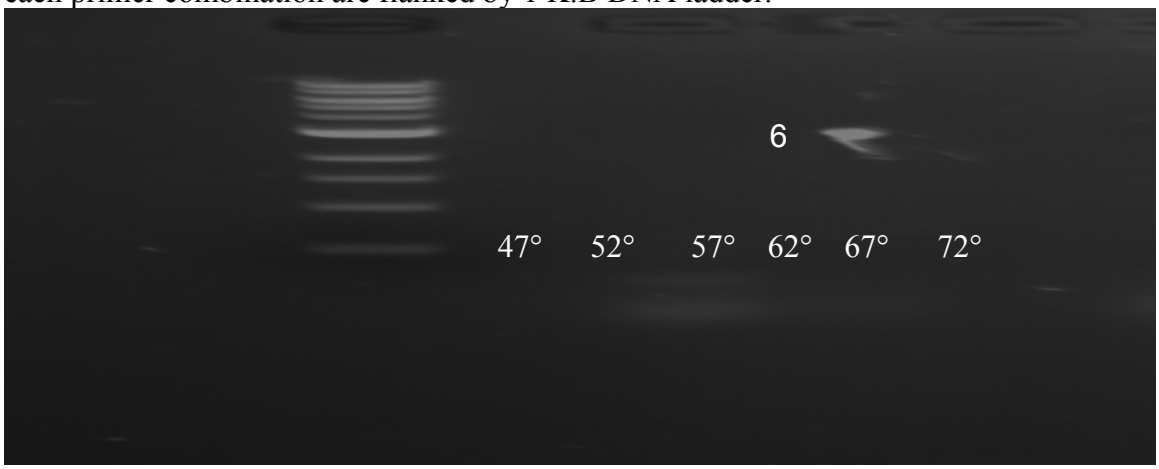


Figure 60: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (Cosm30850 R1 +SLC23A1AttB1 F1) and (COSM30850 F1 +SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon at the correct size. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

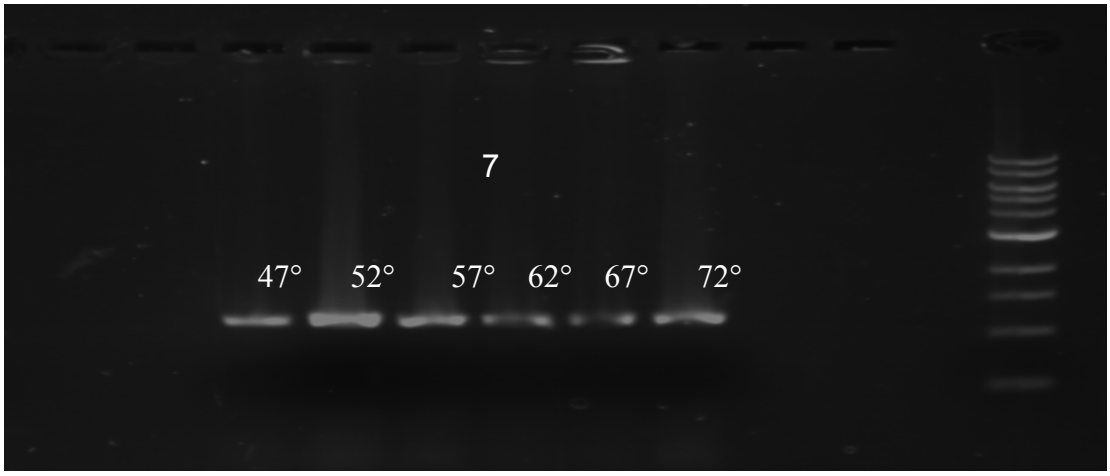


COSM217333

This is a Frameshift SNV, located on exon 11 on the SLC23A1 gene, at amino acid number 423.

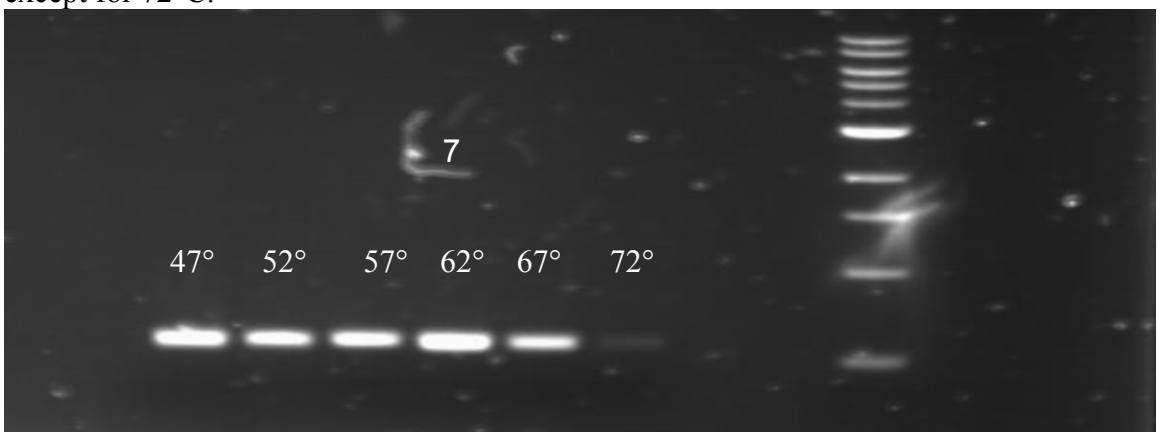
The 1287 nucleotide 5' amplicon introducing SNV COSM217333 was formed with the primers combination, COSM217333 R1 + SLC23A1AttB1 F1 in Phusion HF containing reaction (Figure 61).

Figure 61: Gel image of the gradient PCR for the 1287 nucleotide amplicon for the Primer combination COSM217333R1 + SLC23A1AttB1 F1 using Phusion HF enzyme, successful amplified across all temperatures, with the strongest signal at 52°C was achieved.



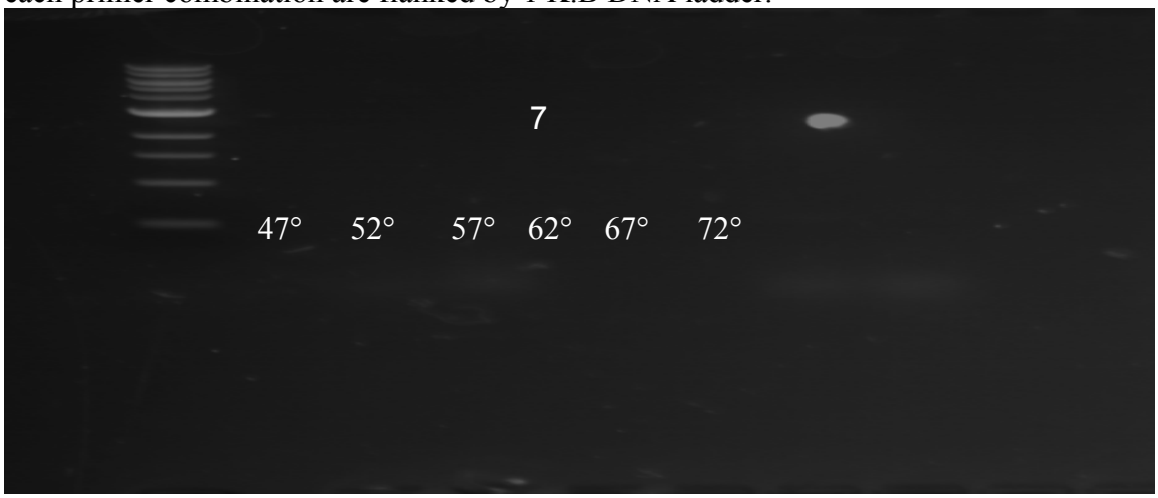
The 574 nucleotide 3' amplicon introducing SNV rs141167635 was formed with the primers combination, RS 139127139 F1 + SLC23A1AttB1 R1 in Phusion HF containing reaction (Figure 62).

Figure 62: Gel image of the gradient PCR for the 574 nucleotide amplicon for the Primer 7 combination COSM217333F1 + SLC23A1AttB1 R1using Phusion HF enzyme, successfully amplified across all temperatures, with the strong signals at all temperatures except for 72°C.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was not successfully amplified at any temperatures (Figure 63).

Figure 63: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (COSM217333 R1 +SLC23A1AttB1 F1) and (COSM217333 F1 +SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon at the correct size. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

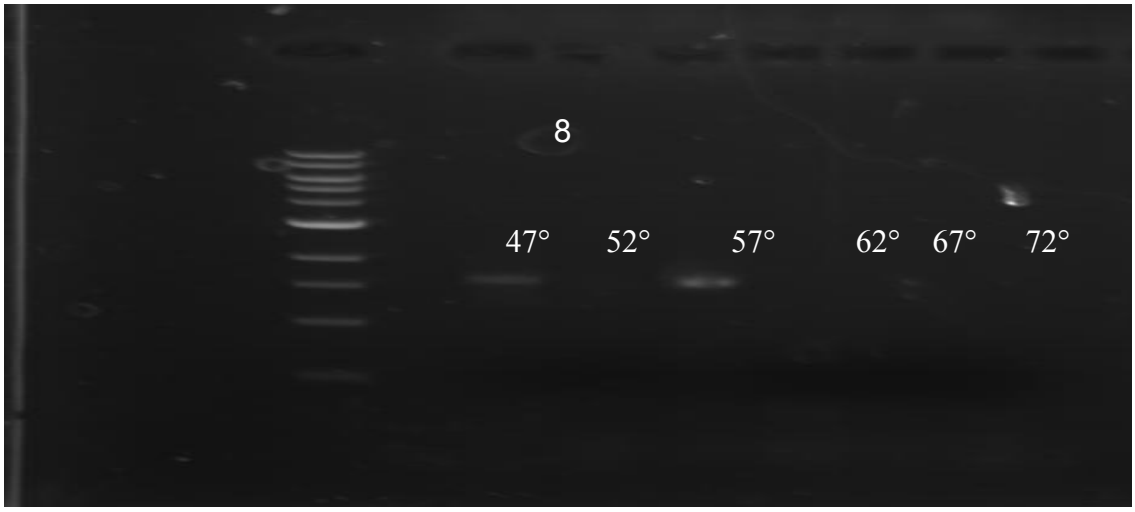


TMP_ESP_5_138707912

This is nonsense SNV, causing a gain stop codon, which signal for a stop codon at the amino acid tryptophan (W). It is located on the exon 14 on SLC23A1 gene, at amino acid number 527.

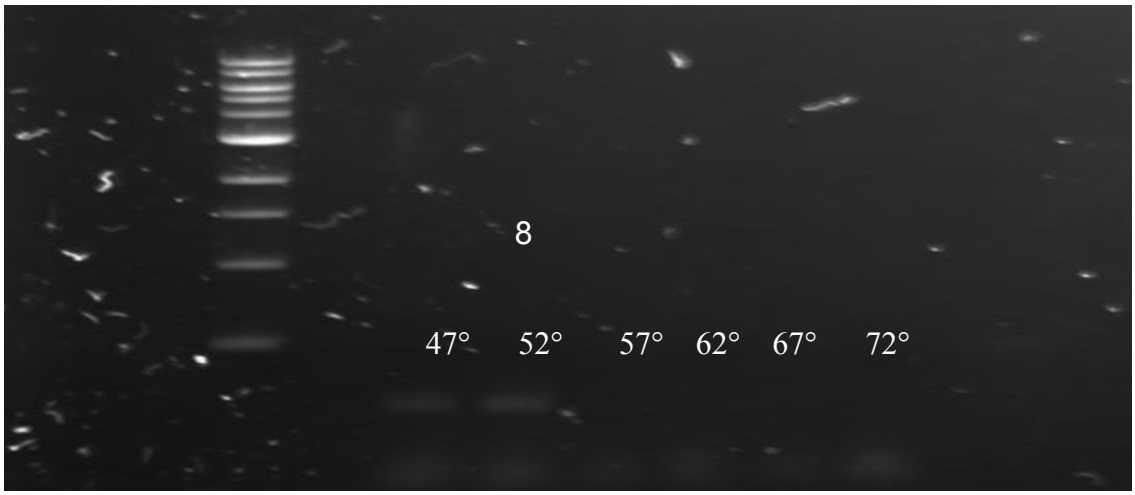
The 1613 nucleotide 5' amplicon introducing SNV TMP_ESP_5_138707912 was formed with the primers combination, TMP_ESP_5_138707912 R1 + SLC23A1AttB1 F1, in Phusion HF containing reaction. However, the amplicon was successfully amplified only at two temperatures 47°C and 57°C (Figure 64).

Figure 64: Gel image of the gradient PCR for first the 1613 nucleotide amplicon for the primer combination TMP_ESP_5_138707912 R1 + SLC23A1AttB1 F1 using Phusion HF enzyme, successful amplification only at two temperatures 47°C and 57°C was achieved.



The 248 nucleotide 3' amplicon introducing SNV TMP_ESP_5_138707912 was formed with the primers combination, TMP_ESP_5_138707912 F1 + SLC23A1AttB1 R1 in Phusion HF containing reaction, and the amplicon was successfully amplified only at two temperatures 47°C and 52°C (Figure 65).

Figure 65: Gel image of the gradient PCR for first the 248 nucleotide amplicon for the primer combination TMP_ESP_5_138707912 F1 + SLC23A1AttB1 R1 using Phusion HF enzyme, successful amplification only at two temperatures 47°C and 52°C was achieved.



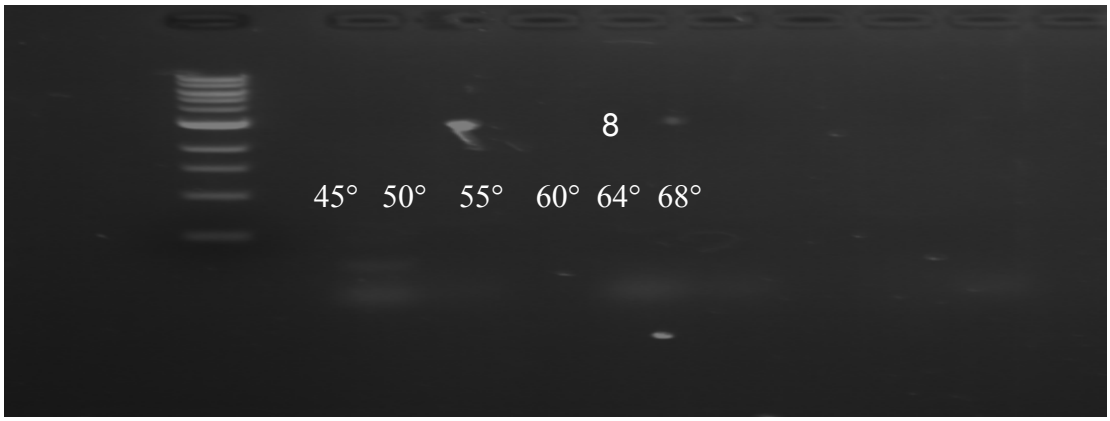
The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was not successfully amplified at any temperatures (

Figure 66 and

Figure 67).

Figure 66: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (TMP_ESP_5_138707912 R1 +SLC23A1AttB1 F1) and (TMP_ESP_5_138707912 F1 +SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon at the correct size. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

Figure 67: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (TMP_ESP_5_138707912 R1 +SLC23A1AttB1 F1) and (TMP_ESP_5_138707912 F1 +SLC23A1AttB1 R1), in a PCR reaction containing regular Taq polymerase. None of the reactions yielded an amplicon at the correct size. Lanes of each primer combination are flanked by 1K.B DNA ladder.

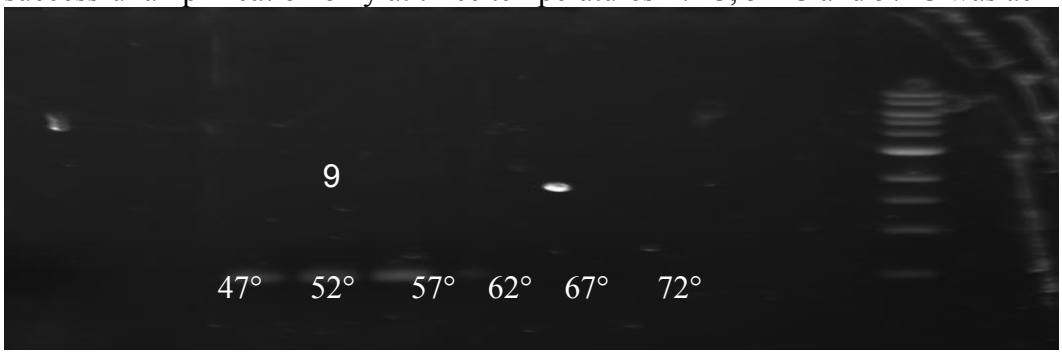


COSM204876

This is a missense SNV that changes the amino acid from glycine (G) to serine (S). It is located on exon 6 on the SLC23A1 gene, at amino acid number 158.

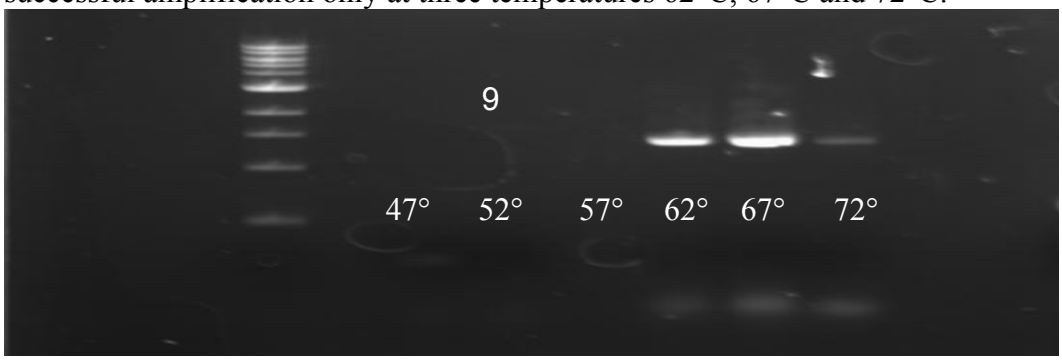
The 505 nucleotide 5' amplicon introducing SNV COSM204876 was formed with the primers combination, COSM204876 R1 + SLC23A1AttB1 F1, in phusion HF containing reaction. However, the amplicon was successfully amplified only at three temperatures 47°C, 52°C and 57°C (Figure 68).

Figure 68: Gel image of the gradient PCR for the 505 nucleotide amplicon for primer combination COSM2 20486 R1+ SLC23A1AttB1 F1 using Phusin HF enzyme, successful amplification only at three temperatures 47°C, 52°C and 57°C was achieved.



The 1356 nucleotide 3' amplicon introducing SNV COSM204876 was formed with the primers combination, COSM204876 F1 + SLC23A1AttB1 R1 in Phusion HF containing reaction, and the amplicon was successfully amplified only at three temperatures 62°C, 67°C and 72°C (Figure 69).

Figure 69: Gel image of the gradient PCR for the 1356 nucleotide amplicon for primer combination COSM2 20486 F1+ SLC23A1AttB1 R1 using Phusin HF enzyme, successful amplification only at three temperatures 62°C, 67°C and 72°C.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was not successfully amplified at any temperatures (Figure 70 and Figure 71).

Figure 70: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (COSM204876 R1 +SLC23A1AttB1 F1) + (COSMCOSM204876 F1 + SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

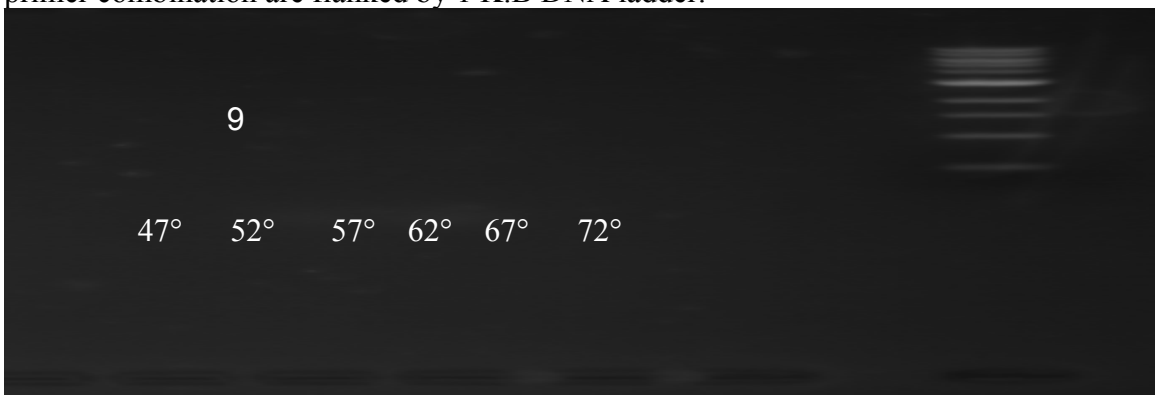


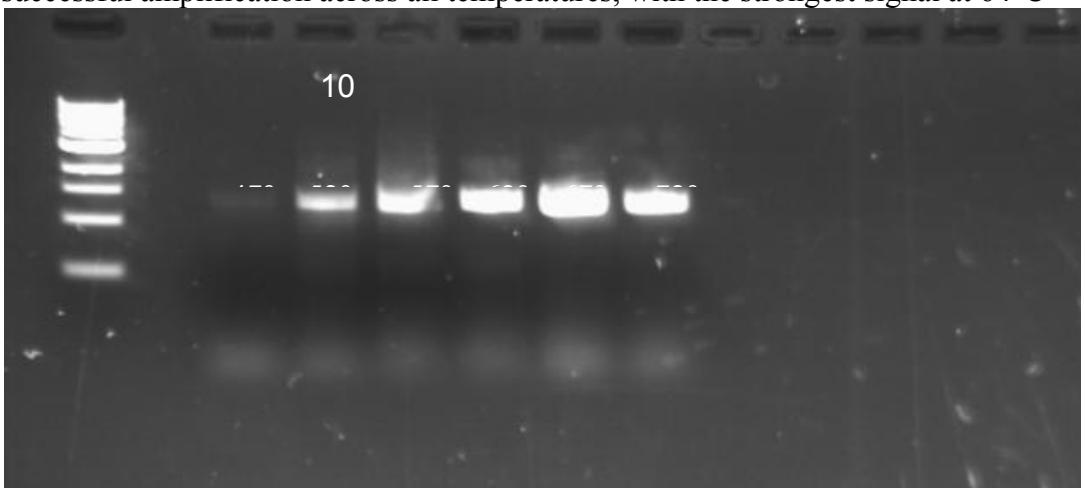
Figure 71: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (COSM204876 R1 +SLC23A1AttB1 F1) + (COSM204876 F1 + SLC23A1AttB1 R1), in a PCR reaction containing regular Taq polymerase. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

Rs6596474

This is a missense SNV that changes the amino acid from serine (S) to alanine (A). It is located on exon 11 on the SLC23A1 gene, at amino acid number 421.

The 1294 nucleotide 5’ amplicon introducing SNV RS 6596474 was formed with the primers combination, RS 6596474 R1 + SLC23A1AttB1 F1, in phusion HF containing reaction (Figure 72).

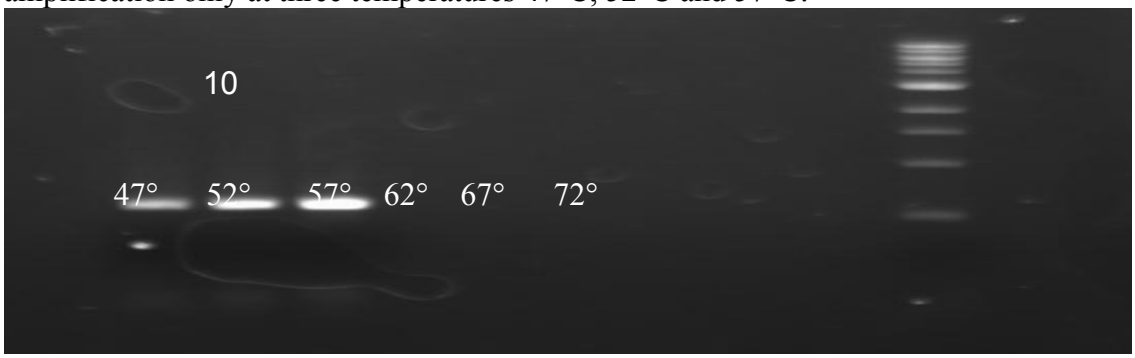
Figure 72: Gel image of the gradient PCR for the 1294 nucleotide amplicon for the primer combination RS 6596474 R1+ SLC23A1AttB1 F1 using Phusin HF enzyme, successful amplification across all temperatures, with the strongest signal at 64°C



The 567 nucleotide 3’ amplicon introducing SNV COSM30850 was formed with the primers combination, RS 6596474 F1 + SLC23A1AttB1 R1 in Phusion HF containing reaction, and the amplicon was successfully amplified only at three temperatures 47°C, 52°C and 57°C (

Figure 73).

Figure 73: Gel image of the gradient PCR for the 567 nucleotide amplicon for primer combination RS 6596474 F1 + SLC23A1AttB1 R1 using Phusin HF enzyme, successful amplification only at three temperatures 47°C, 52°C and 57°C.



The PCR based ligation of the two amplicons was attempted using the respective 5'and 3' amplicons in a dilution of 1:300, with the primer combinations SLC23A1AttB1 F1+ SLC23A1AttB1 R1 in reactions containing Q5 HF and regular Taq polymerases. The expected amplicon size should be 1861 nucleotides, yet all the attempts failed and the amplicon was not successfully amplified at any temperatures (Figure 74 and

Figure 75).

Figure 74: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (RS 6596474 R1 + SLC23A1AttB F1) and (RS 6596474 F1 + SLC23A1AttB1 R1), in a PCR reaction containing Q5 HF enzyme. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.

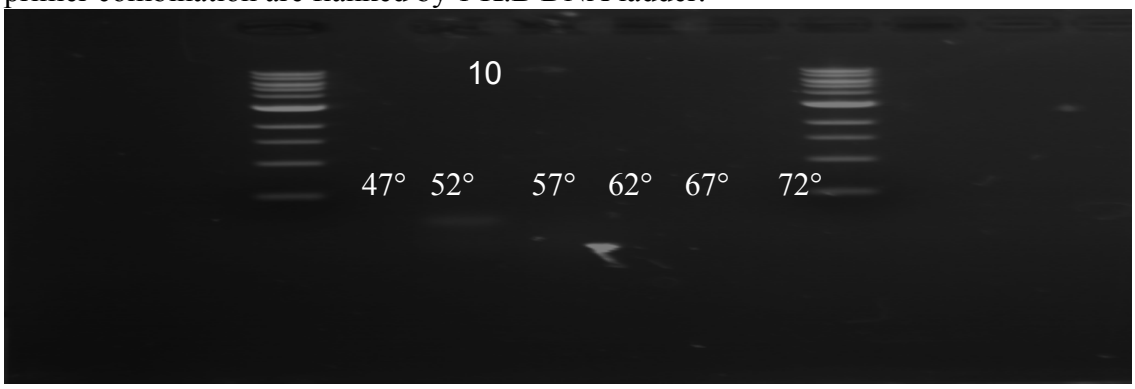
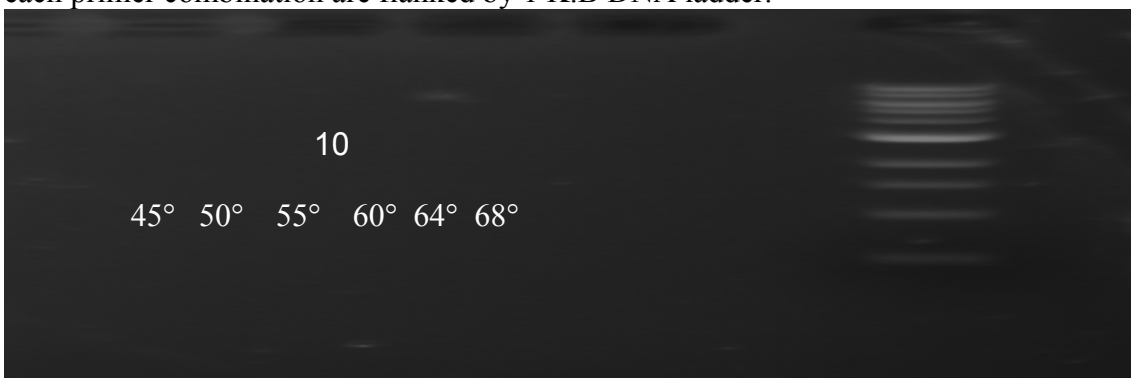


Figure 75: Agarose gel to visualize PCR amplicons of ligating two amplicons, the expected size is 1861 nucleotides for the primer combinations (RS 6596474 R1 + SLC23A1AttB F1) and (RS 6596474 F1 + SLC23A1AttB1 R1), in a PCR reaction containing regular Taq polymerase. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1 K.B DNA ladder.



Appendix E: Introducing SNVs in the 5' and the 3' of the SLC23A1 through amplification of the entire SLC23A1 ORF and tagging with Gateway® flanking sites.

RS143475469

This is missense SNV that changes the amino acid changes on position 18 from aspartic acid to glutamic acid (D 18 E). It is located on exon 2 on the SLC23A1 gene. The nucleotide changes from C to G.

SNV RS143475469 was not successfully introduced using the primers combination, RS143475469 F1 + SLC23A1 ORF R1 untagged in Phusion HF and Q5 HF containing reaction. The expected amplicon size is 1797 nucleotides, yet all the attempts failed and the amplicon was not successfully amplified at any temperatures (figures 69, 70).

Figure 76: Agarose gel to visualize PCR amplicons of the entire SLC23A1 ORF, the expected size is 1797 nucleotides for the primer combinations RS143475469 F1+

SLC23A1 ORF R1 untagged in a PCR reaction containing Phusion HF. None of the reactions yielded an amplicon. Lanes of each primer combination are flanked by 1kb DNA ladder on the sides of the image.

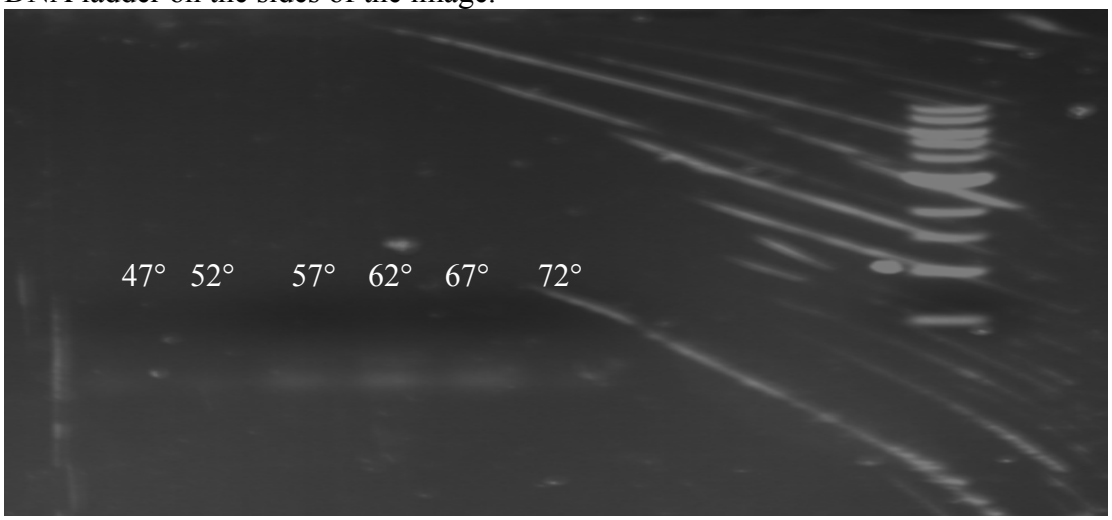
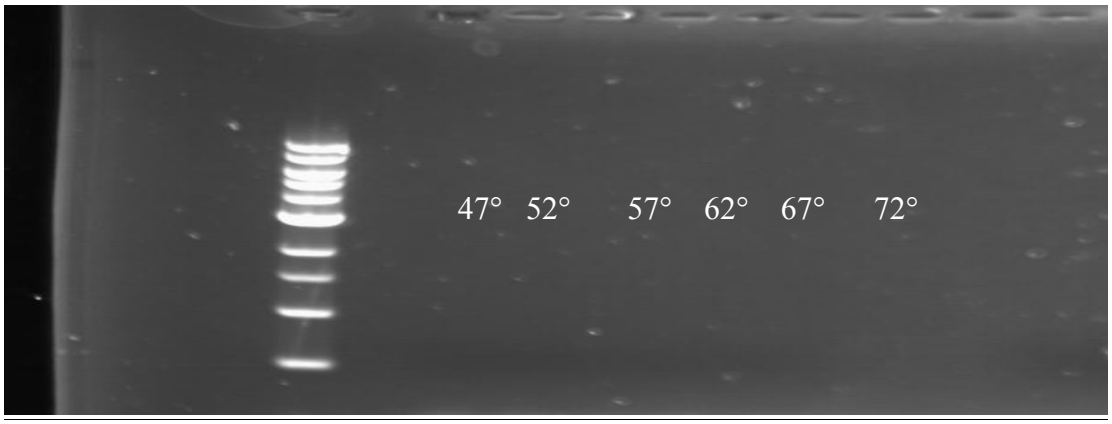


Figure 77: Agarose gel to visualize PCR amplicons of the entire SLC23A1 ORF, the expected size is 1797 nucleotides for the primer combinations RS143475469 F1 + SLC23A1 ORF R1 untagged in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Only the 1 k.b DNA ladder is shown on the side of the image.



Rs138079930

This is a missense SNV that alters the amino acid from glutamic acid on position 8 to lysine (E 8 K). It is located on exon 1 on the SLC23A1 gene. The nucleotide changes from C to G.

SNV RS138079930 was not successfully introduced in the 5' and the 3' of the SLC23A1, using the primers combination, RS138079930 F1 GTW + SLC23A1 ORF R1

untagged in Phusion HF and Q5 HF containing reaction. The expected amplicon size is 1831 nucleotides, yet all the attempts failed and the amplicon was not successfully amplified at any temperatures (figures 71, 72).

Figure 78: Agarose gel to visualize PCR amplicons of the entire SLC23A1 ORF, the expected size is 1797 nucleotides for the primer combinations RS138079930 F1 GTW + SLC23A1 AttB R1 tagged in a PCR reaction containing Phusion HF. None of the reactions yielded an amplicon. Only the1 kb DNA ladder is shown on the side of the image.

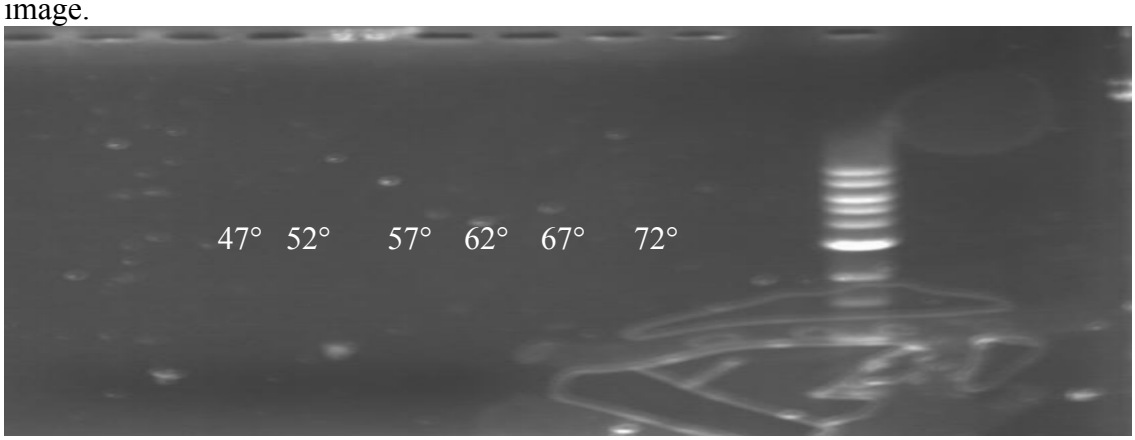


Figure 79: Agarose gel to visualize PCR amplicons of the entire SLC23A1 ORF, the expected size is 1797 nucleotides for the primer combinations RS138079930 F1 GTW + SLC23A1 AttB R1 tagged in a PCR reaction containing Q5 HF. None of the reactions yielded an amplicon. Only the1 kb DNA ladder is shown on the side of the image.