# Genotype dependent gene regulation in visceral adiposity: a study of the 11q23.3 locus

Shanima Shanu Mirza

*This thesis is submitted in partial fulfillment of the requirement for the degree of Master in -Human Nutrition*


Department of Clinical Medicine

&

Hormone Laboratory Research Group

University of Bergen

Norway

May 2022

# Acknowledgements

The project presented here was carried out in the Hormone Laboratory at the Department of Clinical Science from August 2021 until June 2022 as a part of the master program in Human Nutrition at University of Bergen.

Frist and foremost, I would like to express my sincerest gratitude to my supervisor Dr. Jan-Inge Bjune (main supervisor), for believing in me, giving me the opportunity to be part of your group and work on this exciting and innovative project. I am very thankful for the invaluable guidance, knowledge, encouragement, and enthusiasm you have provided me from the very beginning. Also thank you for being patient and supportive all the time. My research would not been possible without the support from you and the group.

I would also like to thank my co-supervisor Dr. Simon Nitter Dankel, I am very grateful for your support, giving me the chance to work on your project, shared knowledge, motivation and always taking the time to answer questions, even when you were very busy.

I am very grateful to all the members of the research group for welcoming me and your willingness to help with even the smallest problems, also making my time here pleasant and educational. Especially, thanks to Margit Hildershavn Solsvik, Regine Åsen Jersin, and Lura Roxana Jonassen for assistance on everyday laboratory work.

Lastly, I would like to thank my family and friends, especially my mom, dad, and brother, for great encouragement and support through this fun, exciting and demanding year. Also thank you for believing in me.

Bergen, May 2022
Shanima Shanu Mirza

# Abstract

**Background:** Overweight and obesity are defined by an excessive accumulation of adipose tissue. Adipose tissue, also known as fat, consist of specialized lipid storing cells called adipocytes. In human, adipose tissue is distributed throughout the body in two major depots: subcutaneous adipose tissue (SAT), and visceral adipose tissue (VAT). While SAT is beneficial for metabolism, VAT is associated with several metabolic disorders and increased risk of premature death. Despite the known pathological effects of VAT, the underlying mechanisms that promote selective fat storage in VAT are poorly understood and largely unexplored. However, genetics are thought to play a major role, and a recent GWAS study has identified 11q23.3 as one of several novel genetic loci associated with VAT mass. Because GWAS associations do not provide functional information, the molecular mechanisms underlying this new association are still unknown.

**Aim:** The overall aim of this study was to investigate enhancer activities in the VAT associated 11q23.3 locus and identify the causal variant(s).

**Methods:** Standard molecular biological methods were used to assess transcriptional activity of the cloned tiles, such as PCR, gel-electrophoresis, In-fusion cloning, and luciferase assays in appropriate adipose cell lines and a control cell line. Site-directed mutagenesis was performed to identify causal SNPs regulating the cis-regulatory (enhancer) activity. Finally, bioinformatic web tools were utilized to search for transcription factor binding sites overlapping the causal to discover upstream regulatory mechanisms.

**Results:** We were able to clone three tiles, from both risk and protective allele types. Our data suggested that tile 8 is involved in genotype-dependent enhancer activity, and that this activity is determined by at least five causal SNPs. We identified multiple transcription binding sites overlapping the SNPs within tile 8 and found a repressive RARB site to be lost in the risk allele of rs1799993.

**Conclusion:** Our findings suggest that tile 8 is involved in genotype dependent enhancer activity of the 11q23.3 locus, and at least five likely causal SNPs, were identified. Among these, the rs1799993 risk variant may disrupt retinol-RARB anti-adipogenic signaling, leading to increased VAT mass.

# Abbreviations

| | |
|---|---|
| ASC | Human adipose derived stem cell |
| BAT | Brown adipose tissue |
| BMI | Body mass index |
| bp | Base pair |
| CVD | Cardiovascular disease |
| eQTL | Expression quantitative trait loci |
| gDNA | Genomic DNA |
| GLP-1 | Glucagon-like peptide-1 |
| GWA | Genome Wide Association |
| GWAS | Genome Wide Association Studies |
| Kbp | Kilo Base pair |
| LD | Linkage disequilibrium |
| PMCA | Phylogenetic module complexity analysis |
| SAT | Subcutaneous adipose tissue |
| SNPs | Single Nucleotide Polymorphism(s) |
| TAG | Triglycerides |
| TBS | Transcription binding site |
| TF | Transcription factors |
| TSS | Transcription start site |
| UPC1 | Uncoupling protein 1 |
| VAT | Visceral adipose tissue |
| WAT | White adipose tissue |
| WHO | World Health Organization |

# Table of contents……VI

# 1. Introduction

## 1.1 Obesity

### 1.1.1 Obesity epidemiology

Overweight and obesity are defined by an excessive accumulation of adipose tissue and are major risk factors for co-morbidities and increased mortality[1,2]. According to the World Health Organization (WHO), the prevalence of obesity and overweight has nearly tripled worldwide since 1975, and obesity is therefore recognized as a global pandemic [3,4]. WHO estimations revealed that in 2016, more that 1.9 billion adults worldwide were categorized as having overweight, of which 650 million adults were obese[3]. Overall, approximately 11% of men and 15% of women globally were estimated to have obesity in 2016[3]. Overweight and obesity is not only restricted to adults but also affects children. According to WHO, an estimated 38.2 million children under the age of 5 were categorized as obese or overweight in 2019[3].

In Norway, data on overweight and obesity of children, adolescents and adults are collected through different studies including Health Studies in Nord-Trøndelag, Tromsø Health Studies, National Service Center, Child Growth Study and the SAMINOR health study[5]. Recent estimation from these different studies have revealed that approximately 15-20% of Norwegian children are overweight or obese, about 25% adolescents are overweight or obese, and almost 25% of adult men and 21% of adult women are obese[5].

### 1.1.2 Nutritional, environmental, and genetic contributions to obesity

Obesity is ultimately the result of chronic excessive positive energy balance[6]. However, the underlying causes driving this disbalance are complex. Diet, physical activity, and genetic predisposition are three important factors that interact with each other to affect energy balance and body weight. In the modern days, a range of unhealthy dietary factors that promotes overnutrition has become a considerable burden in high-income countries and increasing burden for middle-income countries[7]. In these countries most of the population have an unhealthy dietary pattern, which includes low intake of wholefoods, fruits, and vegetables, combined with very high intake of unhealthy ultra-processed food, such as sugar-sweetened beverages and processed meat[8,9]. For example, recent food consumption data from Australian teenagers has demonstrated that one-third of energy intake comes from food that consist of high energy and low nutritional values, high salt, and sugar, high in saturated fat

and low in fiber[7]. These changes in social norms have resulted in food that used be "occasional" are now considered everyday food.

In addition to nutritional aspect, other environmental factors have a huge impact on our energy intake and body weight. Present day "obesogenic environment", which refers to influences such as our surroundings, food availability, socio-economic status, or lifestyle, promote obesity in individuals and in the population, and have a huge impact on human behavior and lifestyle[1010]. For instance, in an obesogenic environment, the community or society has increased marketing, availability, and affordability of energy dense foods, which can partly explain the excess energy intake and weight gain[10]. In such environment individuals must have huge self-control to make responsible decisions about personal diet and physical activity. In addition, other factors that can affect obesity are, sedentary lifestyle, socioeconomic risk factors, education, social acceptance, social media, and gut microbiota[11].

Interestingly, many people living in obesogenic environments remain lean, highlighting the importance of a third factor influencing energy balance, namely genetic predisposition[12–15]. In fact, twin studies suggest that heritability accounts for a striking 40-80% of obesity cases[15] However, obesity genetics is, with the exception of rare monogenic forms of obesity, not completely deterministic, meaning that individuals with risk genes are not destined to be obese, but rather are more susceptible to an obesogenic environment[12,15]. For example, individuals with obesity risk variants in the FTO locus have a 23% higher risk of developing obesity, but this risk has been shown to be preventable with physical activity[16,17]. In contrast, other studies have found limited effect of exercise on obesity[18]. Understanding genetic predisposition is important in determining who are at particular risk of developing obesity, and whether that individual would respond better to nutritional or exercise intervention. Therefore, the focus of this thesis will be on the genetic predisposition of obesity.

### 1.1.3 Characterization and classification of obesity

Body mass index (BMI) is the most widely used classification index for overweight and obesity due to its simplicity and is defined as the person's weight in kilograms divided by height in meters squared ($kg/m^2$). For adults, WHO defines overweight and obesity as BMI $\geq$ 25 and $\geq$ 30, respectively (Table 1.1)[19]. For a person 1.8 m tall, this corresponds to a weight of roughly 80 and 100 kg, respectively.

**Table 1.1: BMI-based classification of obesity.**

| Category | BMI (kg/m$^2$) |
|---|---|
| Underweight | < 18.5 |
| Normal weight | 18.5 – 24.9 |
| Overweight | ≥ 25 |
| Obesity | ≥ 30 |
| Obesity class I | ≥ 30 – 34.9 |
| Obesity class II | ≥ 35 – 39.9 |
| Obesity class III | ≥ 40 |

This table represents values according to WHO guidelines.

Even though BMI is a popular, non-expensive and simple method to estimate overweight and obesity, utilizing BMI as the only marker of obesity can be inaccurate in many situations. For example, BMI does not differentiate between fat mass and fat-free mass, nor accounts for different fat depots in the body, thus giving an inaccurate measurement of the degree or type of obesity[20] In addition, since BMI does not consider bone mass or body composition based on sex, age and race, other methods may be more suitable for accurate measurement of obesity and associated health risks[21].

Non-invasive quantitative methods or anthropometric measurements such as waist circumference (WC), Waist-hip ratio (WHR), Waist-height ratio (WHtR), hip circumference (HC), limb circumference and skinfold thickness, bioelectrical impedance, underwater weighing, can be more accurate than BMI measurements[21–23]. However, these methods still cannot distinguish between all types of adipose tissue types (discussed later). Criterion methods such as dual energy X-ray absorptiometry (DXA), magnetic resonance imaging (MRI) and computed tomography (CT) are the gold standards for accurate measurements of body fat and its distribution[22]. Although these measurement methods provide a better indication of an individual's body fat and its distribution, these methods are not widely available and can be very expensive, time costly and can be difficult to standardize across observers and machines. However, imaging data from about 4,000 people with visceral obesity has recently been combined with other anthropometric measurements to make computational models that can accurately predict the degree of visceral obesity. This model has been used to predict the degree of visceral obesity in nearly 400,000 British people[24].

## 1.2 Adipose tissue

Adipose tissue, also known as fat, is a plastic and heterogenous organ, which consist of specialized lipid storing cells called adipocytes[25]. These cells accumulate as loose connective tissue at specific adipose depots throughout the body and are the master regulators of nutritional hemostasis, energy balance and thermogenesis[25]. Normally these cells maintain the body's energy homeostasis by storing energy in the form of triglycerides (TAG) during feeding and releasing energy in the form of free fatty acids during fasting or in situations with elevated energy demand[26]. In addition, adipocytes secret several endocrine hormones and other factors, collectively known as adipokines. These adipokines plays a key metabolic role and act on other biological systems via regulation of insulin sensitivity (e.g., adiponectin), inflammation (e.g., TNF-$\alpha$, IL-8, IL-18) and insulin resistance (e.g., resistin, lipocalin), reproductive function, blood pressure and angiogenesis[27–29]. An important aspect of adipose tissue endocrinology is that adipose tissue is home to many other cell types besides adipocytes, including monocytes, pluripotent stem cells, pericytes, endothelial cells and macrophages. It is assumed that these non-adipocyte cells might be responsible for some secreted adipokines[29].

### 1.2.1 Types of adipose tissue and their function

In mammals, adipose tissues are categorized in two main types based on anatomy and functions: white adipose tissue (WAT) and brown adipose tissue (BAT)[25]. While both types of adipose tissues take up glucose and fatty acids, WAT stores this energy as lipid droplets, whereas BAT specializes in burning the lipids through generation of heat.

In mammals, including humans, WAT is by far the most abundant form of adipose tissue [30]. The white adipocytes are primarily large adipocytes that possess a single large lipid droplet with relatively few mitochondria[25]. As mentioned above, the principal function of WAT is to control the storage and release of lipids in response to changes in systemic nutritional and metabolic energy levels. Storage of lipids in WAT during periods of positive-energy balance and after feeding is mediated via lipogenesis, also known as lipid synthesis[25]. Lipogenesis is a metabolic process that converts different circulating nutrients such as glucose, long-chain fatty acids and branched-chain amino acids into lipids via esterification fatty acids to acetyl-CoA[31]. In both human and rodents, the major site for lipogenesis is adipose tissue and the liver[32]. In the situation of energy demand due to fasting or physical exercise, adipose tissue

releases free fatty acids and glycerol via lipolysis, which hydrolyzes TAGs to fulfil the energy demand[32]. Both processes are highly regulated by signals from peripheral tissue and the central nerve system[32].

In contrast to WAT, BAT represents only a small proportion of the total adipose tissue mass and is found in the neck/shoulder area in humans Brown adipocytes are highly oxidative and characterized by multilocular lipid droplets with abundant mitochondria which express uncoupling protein 1 (UPC1)[25]. This protein short-circuits the oxidative phosphorylation, leading to oxidation of fatty acids and heat generation without producing ATP[26]. Thus, BAT acts as so-called metabolic sinks that can dispose of excess nutrients[33]. For instance, a recent paper demonstrated that when the thermogenic trait of these cells is fully activated, they can increase whole-body-energy expenditure by over 100% in mice and by 40%-80% in humans [34,35].

The relative contribution of BAT to whole-body energy expenditure in humans is controversial due to the small size of this depot[36]. However, WAT can also participate in the thermogenic process by "browning or beiging", which refers to white adipocytes transforming into brown-like cells in response to cold exposure[33]. This process is highly beneficial due to the superior size of white adipose depots, but unfortunately, achieving effective beiging has proved difficult in obese individuals[36]. Therefore, the remainder of this thesis will focus on the classical white adipose tissues.

### 1.2.2 Distribution of white adipose tissue

WAT is distributed throughout the body in several distinct depots. WAT can be broadly divided into two major depots: subcutaneous adipose tissue (SAT), which is located beneath the skin, and visceral adipose tissue (VAT) that lines the internal organs[30] (Figure 1). Normally, SAT accounts for > 80% of total body fat and VAT accounts for 10-20% in men and 5-10% in women, but these numbers can vary dramatically in obesity[37]. Importantly, studies have shown that SAT possesses multiple beneficial effects on metabolism, while VAT is associated with several metabolic disorders[38,39]. Body fat distribution is dependent on various factors including age, sex, and total body fat content [40]. Thus, it is the distribution of these fat depots, rather than just the total adipose or body mass that determines the risk of developing obesity-associated diseases. Therefore, as mentioned above (section 1.1.3),

accurate measurement or prediction of VAT mass is critical to determine disease risk in obesity.

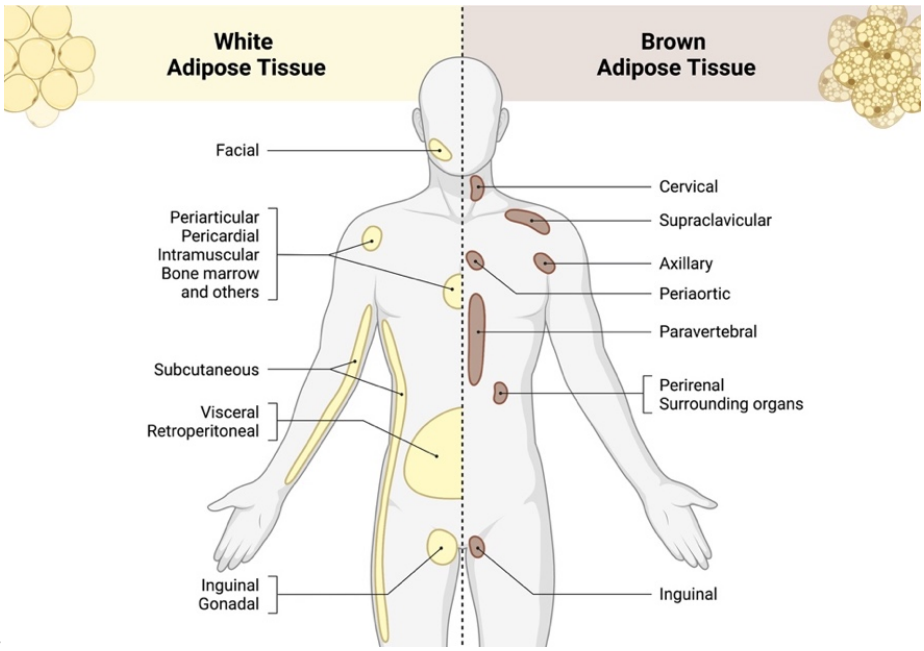



**Figure 1: The major adipose-tissue depots in human**. The main white adipose tissue (WAT) is categorized into two main depots: visceral (VAT) and subcutaneous (SAT). VAT can further be divided into –omental, mesenteric, retroperitoneal, gonadal, and pericardial[37]. SAT is located on upper and lower body region, which can be measured by hip, thigh, and leg circumference. In contrast, brown adipose tissue (BAT) can be found in cervical-supraclavicular, perirenal, and paravertebral regions[37]. This figure is adapted from Biorender.com.

### 1.2.3 Diseases associated with obesity

Throughout most of the human evolution, access to food was occasional and storage of fat was advantageous for survival[25,33]. However, in current time, the constant access to energy-rich foods promoting obesity is highly problematic. Obesity, specifically *visceral* obesity, is an integral part of the metabolic syndrome, a set of unfavorable conditions occurring together, that also includes hypertension, high blood sugar (hyperglycemia), high blood pressure (hypertension), high levels of circulating TAG and low levels of high-density lipoprotein (HDL) cholesterol[25]. These conditions are strongly associated with an increased risk of co-morbidities, such as type 2 diabetes, cardiovascular diseases (CVD), asthma, cancer[41], several other chronic conditions[42], and ultimately increased risk of early death [43]. Nearly 25% of world's populations are affected by metabolic syndrome, according to "The International Diabetes Federation"[43]. In addition to metabolic syndrome, obesity leads to impaired quality of life and decreases life expectancy, as well as socio-economic burden both

in individual and social level[44]. Furthermore, studies have suggested that obesity increases/associated with various types of cancers [41].

## 1.3 Current Treatment options

Obesity is a systematic disease; thus, it requires a multidisciplinary and multimodal treatment. Treatment options surrounding obesity is dependent on the phase of the disease and purpose of the treatment[45]. The following sections describes some of the major treatment options that are available to date.

### 1.3.1 Non-surgerical lifestyle therapy

Non-surgical lifestyle therapy is multifractional interventions, which is designed for individuals or patient groups according to their risk factor status and necessity of the individuals. Non-surgical lifestyle, also known as "basic therapy," includes dietary counseling, physical exercise plan, healthy lifestyle habits and positive behavioral traits. These different aspects of the therapy vary significantly in the weight reduction and weight maintenance phase, also in national and international level[45]. The main goal of the basic therapy is to promote long term body weight reduction in combination with positive change in behavior that contributes to improve in obesity associated risk factors, reduced metabolic syndromes, and improved life quality[45].

### 1.3.2 Drugs supporting treatments

In recent years, FDA has approved several drugs targeting obesity and overeating disorders. According to most international guidelines, patients with a BMI above 25 $kg/m^2$, 30 $kg/m^2$ and at least one weight-related sequelae are considered as eligible for using these drugs [46,47]. In order to gain FDA approval, obesity drugs must achieve >5% weight loss, as well as cardiovascular safety and positive impact on numerous obesity related risk factors [48]. The FDA approved drugs functions in various ways including reduced appetite, increase in energy expenditure or decreased intestinal absorption of food intake (Table 1.2) [49].

**Table 1.2: Current FDA approved pharmacological interventions for obesity.**

| Drugs names | FDA approval | Mechanisms of action |
|---|---|---|
| Orlistat | 1999 | Functions as pancreas lipase inhibitor |
| Lorcaserin | 2012 | Functions as 5-HT2C receptor agonist |

| | | |
|---|---|---|
| Phentramine/topiramate ER | 2012 | Acts as noradrenergic agent/GABA/GLUT modulator |
| Naltrexone SR/bupropion SR | 2014 | Opioid receptor antagonist/dopamine and norepinephrine reuptake inhibitor |
| Liraglutide | 2014 | GLP-1 receptor agonist |
| Lisdexamfetamine dimesylate | 2015 | Central nervous system stimulant |

This table summarizes all the FDA approved drugs to the date and the way these drugs functions[49].

**Orlistat** is the only FDA approved weight loss drug that functions peripherally and is a pancreatic lipase inhibiter that leads to reduced apportion of fat through intestine [50]. **Lorcaserin** is FDA approved in the US and Mexico but didn't get approved in Europe. This drug is a selective 5-HT2C receptor agonist and functions through activating hypothalamus POMC neurons to decreased food intake[51]. The first FDA approved combination drug is **Phentramine /topiramate**, where phentermine causes satiety through norepinephrine and dopaminergic antagonism, and topiramate function through GABA and glutamate signaling. However, the combination effects of these two drugs on weight loss is not known [49]. **Bupropion/naltrexone** is another combinational drug that can be used to treat obesity, where bupropion is a dopamine/norepinephrine reuptake inhibitor and naltrexone is an opioid receptor antagonist[47,49]. **Liraglutide**, is a glucagon-like peptide-1 (GLP-1) receptor agonist and mimics the actions of endogenous GLP-1, that is released from small intestines. Primarily this drug is used to treat type 2 diabetes, however this drug influences weight loss in dose-dependent manner[47,52]. Another GLP-1 receptor agonist, **Semaglutide** recently gained FDA approval, has a much greater effect compared to other drugs, both by itself and in combination with other drugs [53]. Lastly, **Lisdexamfetamine dimesylate** which is originally used to treat deficient hyperactivity disorder (ADHD) and for binge eating disorder. Thus, extended treatment with this drug has a result on obesity[54].

### 1.3.3 Bariatric surgery

At present day, utilization of nonsurgical approaches (lifestyle interventions and drug treatments, except for semaglutide) doesn't lead to a satisfactory long-term weight loss and thus, doesn't reduce the incidence of obesity related co-morbidities[55]. In contrast, bariatric surgery has proved to be highly effective to achieve strong and long-term weight loss in morbidly obese patients[55]. To date, according to American Society of Metabolic and Bariatric

surgery (ASMBS) there are four main types of surgical options are available for morbidly obese patients (with an BMI $\geq$ 40 kg/m$^2$) including – laparoscopic sleeve gastrectomy (SG), Roux-en-Y gastric bypass (RYGB), laparoscopic adjustable gastric banding (LAGB), and biliopancreatic duodenal switch (BPD/DS) [56]. All of these surgeries reduce the physical size of the stomach to a varying degree or bypasses the stomach altogether. The outcome of these surgical treatments lead to huge weight loss, reduction in CVD and improves life quality, but RYGB appears to be more effective[55]. However, there are also multiple issues associated with long-term outcome including recurrence rates, nutritional deficiency, depression, suicide risk, sarcopenia, and other safety issues[45]. Moreover, surgery is only given to people with BMI above 40, or above 35 with at least one comorbidity (such as type 2 diabetes), leaving many overweight or obese patients without treatment options. Thus, the "perfect" cure against obesity is still lacking.

## 1.4 Biological mechanisms contributing to development of obesity and related diseases

Obesity is a complex disease which, as mentioned above, can be caused by a complex interplay between nutrition and genetic factors, social, physiological, and environmental factors. The following sections will introduce some of the biological, molecular and genetic factors that play an important role in development of obesity.

### 1.4.1 Adipose tissue dysfunction

During the development of obesity and obesity related disease, several interconnected processes contribute to adipose tissue dysfunction. Inflammation and insulin resistance are two of the most known mechanisms associated with adipose tissue dysfunction, as well as changes in secretory function of adipocytes [25,57]. In obese individual macrophages are more prevalent, adipose tissue contain M1 and M2 macrophages. Primarily function of these macrophages is tissue repairment and they can store fat like adipose tissue. Therefore, interplay between macrophages and adipose tissue initiate and maintain adipose tissue dysfunction[25,57]. Adipocytes increases in size via hypertrophy because of overnutrition, and these enlarged cells releases more FFAs, thus leads to increase in basal lipolysis. These FFAs binds to macrophage toll-like receptor 4 (TLR-4) leads to increase production of TNF-$\alpha$ production, which then results in increase lipolysis and expression of various genes. Thus, this local paracrine loop leads to pro-inflammatory state both adipocytes and macrophages,

also reduced production of adiponectin[57]. Excessive secretion of FFAs and TNF-$\alpha$ by enlarged fat cells play a crucial role in development of insulin resistance in obese individual. Furthermore, adipogenesis (is the process via differentiation of multipotent stem cells that leads to formation of mature adipocytes), and angiogenesis (formation of blood vessels) are closely related process during adipose hypertrophy[25,57]. Thus, excessive expansion of adipose tissue and angiogenesis leads to hypoxia. Hypoxia inducible transcription factors inhibits transcription of adiponectin genes and PPAR-$\gamma$ activity (involved in adipocytes differentiation and fatty acid trapping)[57].

Obese individuals have high level of leptin, and their appetite and energy expenditure are not regulated. This phenomenon leads to development of hypothalamic leptin resistance, which can lead to persistent hunger in obese people[57]. Moreover, since insulin has the same effects as leptin in the hypothalamus, thus insulin resistance may contribute to leptin resistance. Another characteristic of adipose tissue dysfunction is fibrosis, which characterizes as excessive accumulation of extracellular matrix and tissue stiffness[25]. Fibrosis in adipose tissue is the contributor of functional decline, various factors of gene regulatory networks, cellular mediators and signaling pathways have been associated with adipose fibrosis. These factors such as transcription growth factors, activin A, plated-derived growth factor, connective-tissue growth factor and inflammatory cytokines[25]. Obesity is also related to reduction in abundance and activity of thermogenic adipose tissue[25]. Figure 2 is a summary version of all the major hallmarks associated with obesity development and obesity related diseases.

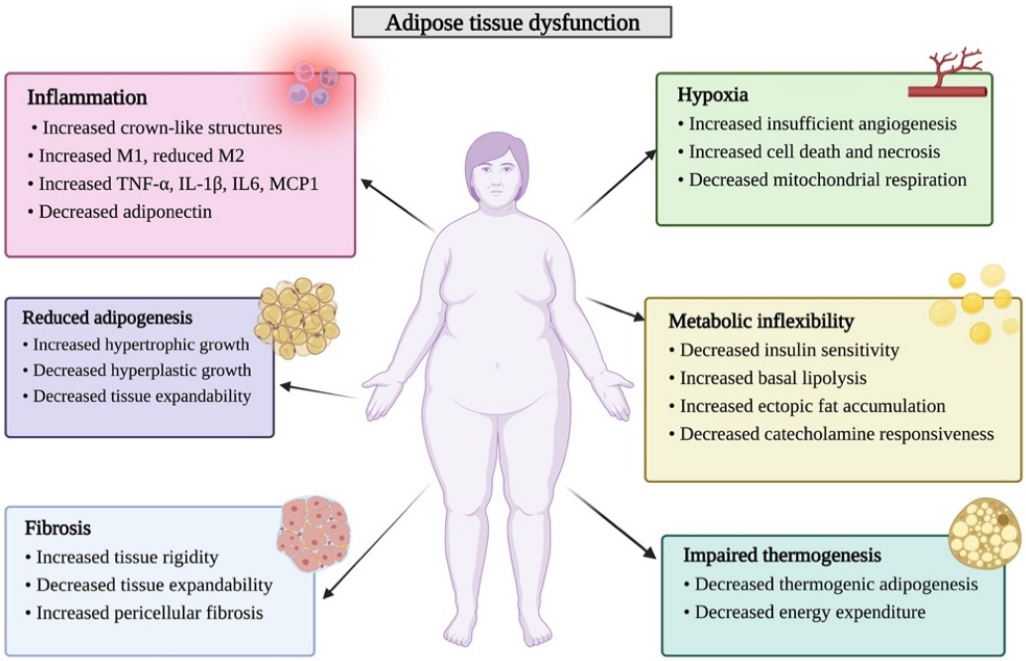



**Figure 2: The hallmarks of adipose tissue dysfunction.** Schematic representation of six main hallmarks, which contributes to adipose tissue dysfunction during the development of obesity related diseases. This figure was adapted from Sakers et al. 2022: "Adipose-tissue plasticity in health and disease" and created in Biorender.com.

### 1.4.2 Genetic mechanisms

Multiple studies have confirmed that within a given population certain variation exist in body size and shape among individuals and many genes have been identified as contributors to severe human obesity. As mentioned above, a meta-analysis of 31 adult twins have shown that BMI variation caused by genetic differences ranged from 47% to 80% [58]. Furthermore, adaptation studies have suggested that BMI of the adapted children strongly correlates with the biological parents and less with the adaptive parents[59].

Non-syndromic monogenic obesity is the consequence of loss-of-function mutations in single genes associated with leptin-melanocortin pathway, and influence approximately 5% of the population[60]. This pathway involves processing signals from the peripheral tissue into the hypothalamus, thus regulates food intake and influence body weight. Mutations in these genes are rare and primarily causes severe over-eating and strong obesity[61]. Genes that have been identified with these kinds of mutations including *LEP*, leptin receptor (*LEPR*), melanocortin 4 receptor (*MC4R*), and pro-opiomelanocortin (*POMC*). Loss-of-function mutation in *MC4R* are the most common form of monogenic obesity with an occurring frequency of 2-6% in obese individuals [60,61]. Therefore, monogenic mutations only explain a tiny fraction of the 47-80% of obese individuals attributable to genetics.

The remaining obesity cases, usually referred to as common obesity, can also be termed polygenic obesity, meaning that multiple genes are influencing the risk of developing obesity[62]. Unlike monogenic mutations, which are rare and usually reside in the coding region of a gene, leading to a disrupted protein production or function with severe impact on cellular function, polygenic mutations, termed "single-nucleotide-polymorphisms (SNPs) are common and most often found in non-coding regions such as promoters, enhancers, and other intergenic regions[63]. Thus, polygenic mutations mostly affect gene expression, each with a very little effect on the phenotype alone, but in combination with other predisposing variants may lead to huge phenotypic influence[62,63].

In recent years, modern technology and completion of Human genome project has led to a better understanding of genes related to obesity. The FTO –fat-mass and obesity associated gene was the first obesity related gene identified via Genome Wide Association (GWA, discussed below) approach in 2007. The most recent meta-analysis of GWAS [64] and several studies have shown that variation in *FTO* gene is associated with increased appetite, energy, dietary fat and protein intake, and reduced satiety, also thermogenesis process of the adipose tissue [65,66]. Other loci have also been implicated in gen-diet interaction, including *IRS, TCF7L2, GIPR, ADIPOQ, FAAH, FTO* or *Visfatin*[65], among many other loci.

### 1.4.3 Genome wide association studies (GWAS)

While the few genes involved in monogenic obesity were identified through linkage studies in family pedigrees during the 90s, this approach was not suitable to identify common variants in the general population because the human genome had not yet been mapped. However, with completion of the Human Genome Project in 2003 the DNA sequences of the entire human genome was revealed, and researchers could then pinpoint various regions of the genome that varies between individuals and in theory search for gene variants occurring more frequently in certain groups of people with specific traits, for example with and without obesity[67]. However, as more genomes were slowly being sequenced, it turned out that there are approximately 10 million SNPs, or bases in the genome that differ from individual to individual. [68]. Due to the time and cost of sequencing at the time, comparing large groups of cases and controls were not feasible[69]. However, the HapMap project, discovered that many SNPs that are inherited together in so-called haploblocks, also known as linkage disequilibrium (LD)-blocks[68]. Therefore, a single tag-SNP can represent hundreds of other SNPs in each block. Thus, a few of these tag-SNPs are sufficient to identify common

haplotypes in a region[68] (Figure 3), reducing the number of SNPs need for sequencing to about 200,000 – 1,000,000 which was manageable.

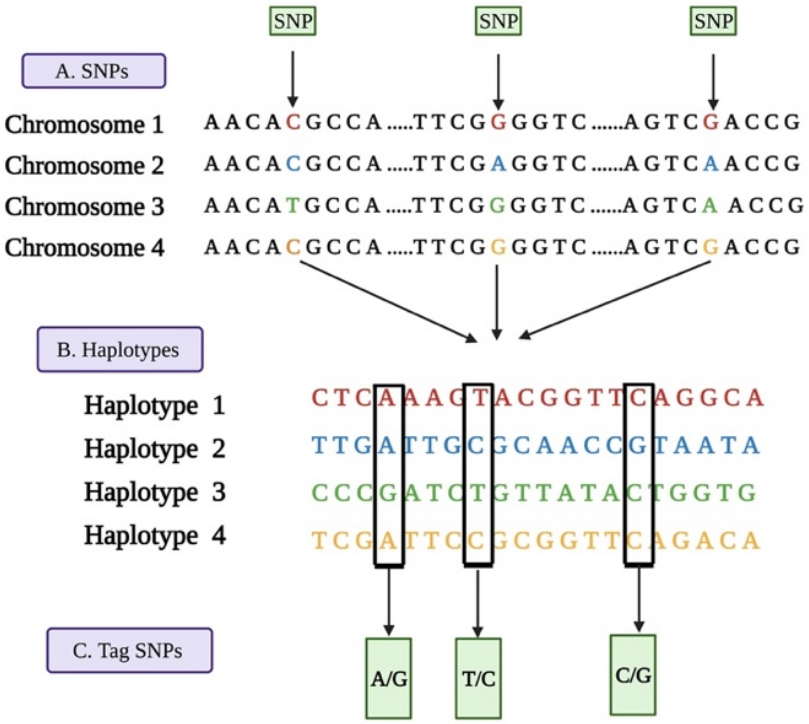



**Figure 3. Schematic representation of single nucleotide polymorphism (SNPs), haplotypes, and tag-SNPs. A)** A short DNA segment of the same chromosome from four different people. Most of the bases in segment are similar, except for the three colored bases in each segment that shows a variation, also known as SNP. **B)** Represents haplotype, that is a series of alleles found at a linked loci on a single chromosome. The figure is showing the extended haplotypes of the same segments of these four chromosomes, containing 20 SNPs and which allele of each SNP the chromosome carries. **C)** tag-SNPs, genotyping only three of the 20 SNPs is enough to identify each of the four haplotypes uniquely. This figure was adapted from "The international HapMap project" and created from Biorender.com.

These projects together with other new research techniques made it possible to deploy GWAS. GWAS is population-based technique, and the aim is to detect associations between trait variants and allele frequencies within a given population[70]. GWAS is conducted by collection of DNA from patients groups with and without the disease being studied. Then genotyping is on the tag-SNPs that represent the total genomic variation. Association tests then are performed to test whether certain genetic variants occur significantly more frequent in the genome of the patients with disease compared to patients without the disease[70]. With improved sequencing technology and larger sample sizes, an exponential number of obesity associated SNPs have been identified in recent years, reaching about 1,100 by 2021[71]. In contrast to the rapid identification of new variants, the biological interpretation of what the

variants mean and how they influence obesity has proved extremely slow due to a number of limitations of GWAS, as explained in the next section.

### A. Challenges with GWAS

Despite having detected thousands of disease associated variants, these variants still do not account for all the genetic contribution of obesity, a situation often termed "missing heritability"[72]. A suggested explanation for this is the genome-wide significant threshold level[73]. With advances in sequencing technology, more SNPs are being sequenced, [73], meaning that more corrections for multiple testing are needed. Thus, only SNPs with the strongest p-values (below $5 \times 10^{-8}$) are considered to be associated with the investigated trait, leading to fewer discoveries[73].. Therefore, it is believed that a very large number of undiscovered SNPs with small effect size together may explain the missing heritability in obesity [70].

Despite the missing heritability, the roughly 1,100 SNPs already identified with the strongest effect should in theory be enough to explain the genetic basis and molecular mechanisms of obesity. However, this has not been the case, because more than 90% of these variants/SNPs are located in non-coding regions, meaning that they cannot directly be mapped to any gene function or biological mechanism[72,74–76]. Many of these variants are likely in promoters and enhancers and therefore affecting regulation of gene expression rather than protein structure. While SNPs in promoters can be assigned to the nearest gene and thereby be assigned a biological function, the target gene(s) of enhancers may be situated hundreds of thousands of bases away due to looping of DNA[77]. Therefore, a SNP in an enhancer may potentially regulate hundreds of genes. Thus, non-coding SNPs must be painstakingly analyzed by comprehensive bioinformatic and functional follow-up studies. Another important limitation of GWAS is that the association does not say anything about in which cells or tissues the SNP may act, making biological interpretation even harder[78].

Finally, even if one were to solve the abovementioned problems, the reported tag-SNP is likely not causal, meaning not being the variant that is driving the association. As mentioned above, a tag-SNP may represent hundreds of SNPs in a block of DNA that are inherited together. There may be only one, or a few of these SNPs that actually have an impact by serving as binding sites for transcription factors. Without knowing the exact causal SNPs, the biological mechanisms cannot be fully understood. Thus, pinpointing casual variants is

essential for not only prediction of the risk associated with disease, but also for understanding the underlying mechanism related to complex trait variation [62].

### B. Solution to the challenges associated with GWAS

Complementary methods are needed to correctly identify the underlying target genes associated with diseases and variants located in the non-coding region. For example, increased sample size can lead to high mapping resolution, where GWAS is performed across multiple populations. Theory of this approach is that causal variants associated with a trait remain consistent across population, while tag SNPs may only be associated with subset of population because of the differences in LD structures[79].

Expression quantitative trait loci (eQTL) analysis is an important approach to understand molecular mechanism underlying complex trait. eQTL analyses are similar to GWAS, but associate SNPs with expression of genes rather than disease status[80]. By intersecting GWAS hits with eQTL, it is possible to investigate whether a SNP associated with a disease is also associated with the expression of a gene in various tissues.[79,81,82]. Thus, eQTL is a powerful method to predict *likely* downstream target genes of disease-associated SNP, although experimental validation is necessary.

Even if a disease-associated SNP from GWAS also has an eQTL hit that suggests a likely downstream gene, these analyses do not show *how* the SNP affects the expression of that gene. As mentioned above, the mechanisms likely involve enhancer activity. Different sets of genes are expressed in different cells and tissues, defining their identity. The expression of these tissue specific genes are regulated by both epigenetic modifications and highly sequence-specific transcription factor regulatory elements. These transcription factor regulatory elements can be divided in *cis* and *trans*-regulatory elements (located in the non-coding region of the gene), and they vary based on key genetic and evolutionary properties. Cis-regulatory elements (CRE) also known as cis-regulatory modulus (CRM), such as enhancers, promoters, and silencers, are clusters of transcription-binding motifs that regulate the gene expression/transcription of nearby genes. Trans-regulatory elements (TRE) encode trans-regulatory factors, more commonly known as transcription factors (TFs), that bind to CREs and mediate the regulation. Normally, gene expression is regulated by complex interaction between multiple CREs and TRE/TFs[81,82], (Figure 4).

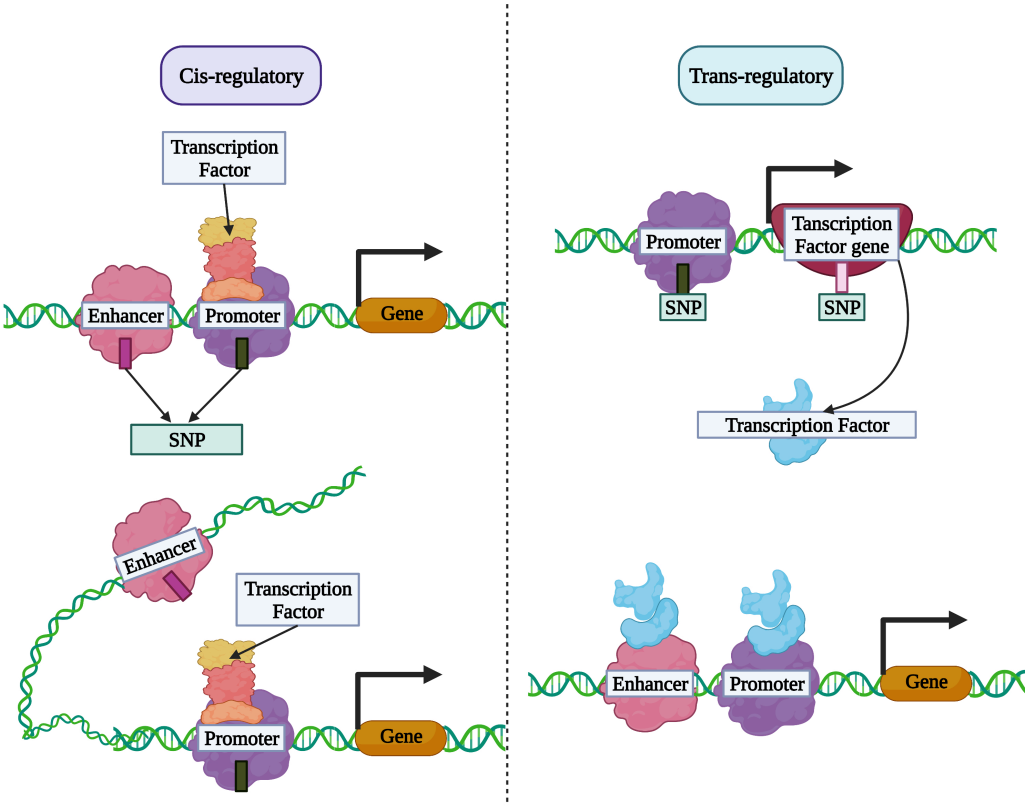



**Figure 4: Schematic representation of variation in *cis* and *trans* regulatory modulus.** As the figure shows cis acting variants/ SNPs can influence both proximal and distal elements, which lead to changes in gene expression. In contrast, variants in trans regulatory elements rather alters the abundance or nature of an intermediate factor and consequently changes expression of the target gene. Moreover, these variants can be both in coding and non-coding regions. This figure is adapted from "Missing heritability in Parkinson's disease: the emerging role of non-coding genetic variation" and created in biorender.com.

SNP variants can occur in either of these regulatory networks, which can lead to changes in coding sequence, change in gene expression level or the whole networks of gene and their product [82]. However, it is variations in CREs that are believed to be most important, meaning that mutations in enhancers of target genes have a bigger effect on phenotype variation than mutations affecting the upstream TF itself[83]. These genetic variations can be studied via utilizing eQTL analysis and eQTL expressions are divided in *cis* and *trans*. A cis-eQTL appear with the location of the underlying genes, in comparison trans-eQTL doesn't necessarily pinpoint to transcription factors, but rather a heterogenous group of transcription regulators [81,82]. Furthermore, trans-eQTL haven been shown to be globally important, however their individual effects are small and therefore their genome-wide identification remains challenging [81].

Thus, mutations that change a TF binding site (TFBS) may be more important than mutations that change the expression level or function of a TF because changes in the TFBS can lead to complete loss of TF binding or introduce a novel binding motif for a completely different TF. It is therefore important to identify whether a SNP sits in a TFBS depending, and whether a risk or protective variant may change the binding site.

One of the major limitations of both GWAS and eQTL analyses is that they only inform about the association between a block of multiple SNPs/variants in LD, but do not reveal which one(s) of them are the actual casual variant(s) responsible for changes in the target genes. In other words, they do not tell which TFBS are affected[84]. Therefore, in 2014, Claussnitzer and colleagues developed a computational method, known as phylogenetic module complexity analysis (PMCA) to narrow down potential casual variants in a GWAS locus[85]. The PMCA approach utilizes algorithm that analyses the phylogenetic conservation between species to detect and identify similarities in complex patterns of co-occurring transcription factor binding sites for combinatorial binding of transcription factors within CRMs. In other words, researchers used the PMCA method as a starting point to identify genetic variation in the non-coding regions that binds to transcription factors and involved in gene regulation, and how they are associated with disease mechanisms [85]. Based on the analysis, "complex regions", which refers to genomic region with significant enrichment of phylogenetically conserved transcription factor binding sites modules, could be identified. In contrast, "noncomplex region" refers to regions with no particular significance in terms of gene regulation [85]. This computational method made it possible to narrow down the number of candidate causal SNPs in an LD block from hundreds to a handfull. This reduction is important, because to pinpoint the exact causal variant(s), each SNP must be experimentally tested *in vitro* and *in vivo* by luciferase reporter assays and genome editing, respectively, which requires a huge effort for even a single SNP.

In summary, Claussnitzer et al. was able to demonstrate a clustering of specific homeobox transcription factor binding sites at 47 type 2 diabetes associated risk GWAS loci. Furthermore, they were able to identify cis-regulatory variant rs4684847 in the non-coding region of the *PPARG* gene, which inhibits the activity of PPARG2 protein. PPARG2 is a master regulator of adipocyte development and fatty acid storage, disrupting glucose metabolisms and initiation of type 2 diabetes[33]. Using the same approach, Claussnitzer et al. was able to demonstrate that the GWAS signal most strongly associated with overall obesity

(in the *FTO locus)* was situated in an enhancer that regulated two genes 0.5 and 1.5 million base pairs away. Moreover, they were able to pinpoint the exact causal SNP and showed that the risk variant disrupted the binding of a repressive TF, leading to increased elevation of the target genes (IRX3 and IRX5), which promoted fat storage in adipocytes.

### C. Identification of new loci associated with visceral adiposity

As explained earlier, fat distribution plays a crucial role in the risk of developing the metabolic syndrome and other obesity related diseases. Among all the different types of adipose tissue, VAT accumulation in the body is more harmful compared to other adipose tissue type[24]. Despite these facts, the underlying genetic mechanisms promoting selective VAT storage are poorly investigated. Until recently, only a few GWAS studies have investigated genetic associations with VAT, and these only provided 4 genes (Karlsson et al. 2019, supplementary: functional analysis). A likely explanation t is the technical and logistical challenges, as VAT associated studies require advanced imaging techniques to precisely measure the fat content surrounding the internal organs[24].

However, a recent study by Karlsson et al. performed a large-scale GWAS for VAT accumulation revealed 200 new associated loci[24]. To achieve this, they developed a sex-stratified, nonlinear prediction model via X-ray absorptiometry (DXA) measurements of VAT mass on 4,198 individuals of white British ancestry from UK biobank. Then they utilized these prediction models to predict the VAT mass of nearly 400,000 white British individuals from UK biobank, of which 325,153 were used in downstream analysis. A large-scale GWAS was performed on these participants with predicted VAT mass. From this study they were able to identify 209 independent VAT mass associations that was distributed over 200 non-overlapping loci. Among these, 71 VAT associated lead SNPs/ tag-SNPs overlapped with eQTL. . One of these VAT-associated tag-SNPs was found in the 11q23.3 and showed an eQTL signal for the *HMBS* and *VPS11* genes in visceral fat. The tag-SNP rs rs1784461 is in LD with (38 SNPs) other SNPs, which all potential causal variants (Figure 5).

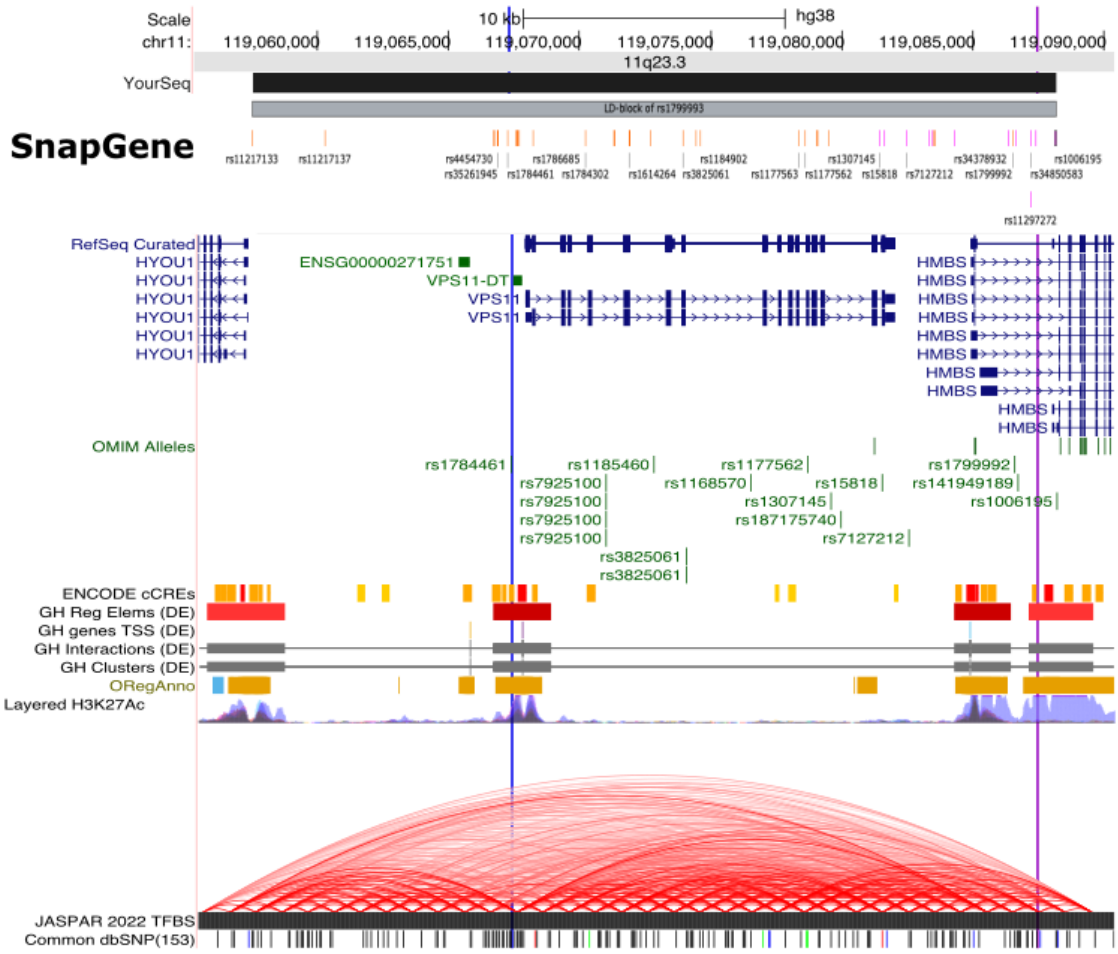



**Figure 5: UCSC genome browser on human locus 11q23.3.** The first bar (gray) represents the chromosome 11q13.3 locus and the second bar are the actual sequence search of the LD-block based on tag-SNP 1784461 from SnapGene. The third bar (dark gray) presents sequence/map in the SnapGene and all the SNPs in the LD-block. Next, presentation of RefSeq data, which shows human protein coding and non-protein coding genes taken from NCBI RNA reference sequences collection, in this scenario three genes and their isoforms –*HYOU1, VPS11* and *HMBS*. The thick dark blue bars of RefSeq data are exons and thin line with pills are introns. OMIM is a compendium of human genes and genetics phenotypes and OMIM alleles represents variants in the OMIM database that have associated dbSNP identifiers. dbSNP contains mapping information of SNP variations, microsatellites, insertions, and deletions for both common variations and clinical mutations. Next track displays ENCODE registry of candidate cis-regulatory elements (cCREs) in the human genome, red bar represents promoter-like signature, orange bar represents proximal-like signature and yellow bar represents distal enhancer-like signature. The next four tracks are GeneHancer data, and they account for information concerning regulatory elements, gene transcription siter (promoter), association between regulatory elements and genes and clustered interactions. Red bars are prompters and gray bars are enhancers. ORegAnno is another data source that represents data including regulatory regions, transcription binding sites and regulatory polymorphisms, where light blue bar represents regulatory region and orange bars represents transcription binding sites. Layered H3K27Ac track characterizes enhancers by histone modification marker H3K27Ac, from the figure a lot of enhancer activity in the *HMBS* region. The JASPAR track represents genome-wide predicted binding sites for each of the transcription factor profiles in the JASPAR CORE collection. The shaded arcs are predicted binding sites and indicates the p-value of the matched position, the darker the shades the lower the p -value. Lastly, common dbSNPs shows the short genetic variations from dbSNP, most variants in dbSNP are rare and some variants are pathogenic. This figure was created in UCSC genome browser along with multiple different data sources.

Notably *HMBS* gene is one of the two key heme biosynthesis related gene and linked to adipogenesis via activities related to mitochondrial respiratory[86]. Karlsson et al. performed a limited functional analysis where they focused on 3 regions with predicted regulatory activity. A subset of SNPs, including tag-SNP rs1784461 were selected for functional follow up, where they performed luciferase assay to predict the differences in luciferase activity between risk and protective alleles. Interestingly they identified one SNP –rs1799993 in the regulatory region, that showed genotype-dependent promoter activity. Thus, they suggested that this SNP is a likely causal variant underlying the associations between the risk haplotype in the 11q23.3 locus, increased *HMBS* expression and increased VAT mass.

A major limitation of the Karlsson et al study is the use of a luciferase reporter system that only assessed promoter activity, because although the proposed causal SNP rs1799993 resides in the promoter region of *HMBS*, this is limited to the short isoforms that are only expressed in blood cells (Figure 5 and Uniprot database[87]), which are not relevant for fat cells. In contrast, most *HMBS* isoforms possess additional exons further upstream, meaning that rs1799993 resides in the intron of these splice variants. It is therefore more likely that in fat cells, this SNP resides in an enhancer, rather than a promoter. This further opens the possibility that rs1799993 regulates not only HMBS, but potentially a range of other genes too (that are too distant to be picked up by eQTL analyses).

A second limitation of the Karlsson study is the lack of a systematic investigation of the entire LD block for regulatory activity, meaning that there may be other undetected causal SNPs present. Finally, because of these limitations, a TFBS, or change has not been identified. To address these shortcomings, our lab has previously attempted to clone different sections (also termed tiles) of the 11q23.3 locus into an enhancer luciferase reporter to assess enhancer activity. However, due to technical challenges with amplification and a traditional restriction enzyme digestion and ligation approach, only a few small parts of the locus were assessed with luciferase assays[88]. Importantly, one of the successfully cloned tiles was 8a, which harbored rs1799993. This tile did show increased luciferase activity, but since only the protective genotype was cloned, it is not known whether this tile has a genotype-dependent effect on gene expression. Moreover, since the tile includes other SNPs as well, it cannot be excluded that one or more of these SNPs may be causal as well. Finally, the enhancer activity of other tiles in the locus are also lacking.

The present study attempts to shed further light on the genetic mechanisms underlying the association between the 11q23.3 locus and visceral obesity, so that in long term these findings can lead to a better understanding of VAT and VAT associated diseases, and perhaps ultimately provide a treatment target against visceral obesity.

# 2. Aims of the study

VAT is particularly strongly associated with metabolic diseases compared to SAT, and the underlying mechanisms that promote selective fat storage in VAT are poorly understood. However, genetics are thought to play a major role, and a recent GWAS has identified several genetic loci that are specifically associated with increased VAT mass.

The main goal of this study is to investigate genotype-dependent cis-regulatory activities in one of the VAT-associated loci (11q23.3) and identify the causal genetic variant(s).

Specific aims include:

1. Design new tiles for the 11q23.3 locus
2. Optimize PCR conditions for each tile from both risk and protective donors
3. Perform in-fusion cloning of the tiles into a luciferase reporter vector
4. Screen for enhancer activity using luciferase assays
5. Identify tiles with genotype-dependent enhancer activity
6. Perform site-directed mutagenesis of individual SNP(s) in active tiles to pinpoint causal variant(s)
7. Identify TFs binding to the causal SNP(s)

# 3. Materials

**Bioinformatics**

**Table 3.1 Online tools**

| Online tool | Description | URL |
|---|---|---|
| Beacon PrimerSoft | Tool for assessing for primer properties | http://www.premierbiosoft.com/qOligo/Oligo.jsp?PID=1 |
| GTExPortal | Provides information about tissue specific eQTL data | https://www.gtexportal.org/home/ |
| HaloReg v4.1 | Tool for exploring variations in haploblocks. | https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php |
| JASPAR CORE 2022 | Open-access database for transcription factor binding profiles | https://jaspar.genereg.net/ |
| NCBI primer BLAST | Website for designing primers. | https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi |
| NEBcloner | Website that provides information about compatibility between restriction enzymes. | https://nebcloner.neb.com/#!/ |
| QuickChange Primer Design (Agilent) | Tool to design primers for site-directed mutagenesis | https://www.agilent.com/store/primerDesignProgram.jsp |
| TaKaRa | Website for designing in-fusion cloning primers | https://www.takarabio.com/products/cloning/in-fusion-seamless-cloning |
| UCSC Genome Browser | Web interface for visualization of genomic DNA. | http://genome.ucsc.edu |
| WashU Epigenome browser | Tool to visualize, integrate and analyze epigenomic datasets | http://epigenomegateway.wustl.edu/browser/ |

**Table 3.2 Analytic software used in this thesis**

| Product | Supplier | Application |
|---|---|---|
| Biorender | Biorender | Image editor |
| GraphPad Prism 9 | GraphPad Software Inc | Statistical analysis |
| Image Lab™ Software Version 3.0 | BioRad Laboratories Inc | Gel visualization |
| Inkscape | Inkscape | Image editor |
| Microsoft® Excel 2022 | Microsoft® | Statistical Data analysis |
| Microsoft® Word 2022 | Microsoft® | Writing Text editing program |

| | | |
|---|---|---|
| Optima Data Analysis | BGM LabTech | Excel export Retrieval of luciferase raw data |
| SnapGene Viewer | SnapGene | DNA sequence visualization |

**Table 3.3 Overview of in-fusion cloning primes used to amplify tiles.**

| Tile | Version | Sequence 5'→3' | Product length |
|---|---|---|---|
| **Tile 1** | **1.1** | F:TGGCCTAACTGGCCGGGGGGAGGCGAAGAAAGAAA<br>R:GGCTAGCGAGCTCAGGGTTTGGGTTGGCATCAAGG | 3958 bp |
| | **1.2** | F:TGGCCTAACTGGCCGTTTGCTCCCTCTACTGGGGT<br>R:GGCTAGCGAGCTCAGGGTTTGGGTTGGCATCAAGG | 4084 bp |
| | **1.3** | F:TGGCCTAACTGGCCGAAGGAGACCCGAGCCTTTTG<br>R:GGCTAGCGAGCTCAGGCAGAACGAACCCCATGCTA | 4003 bp |
| **Tile 3** | **3.1** | F: TGGCCTAACTGGCCGTGCATTTGGGACAAGAAGTGT<br>R:GGCTAGCGAGCTCAGAGCAGGGAAGATTCGAGTGC | 4761 bp |
| | **3.2** | F:TGGCCTAACTGGCCGTGCATTTGGGACAAGAAGTGTT<br>R:GGCTAGCGAGCTCAGACTCACCAAGGGGTTGATGC | 4604 bp |
| | **3.3** | F:TGGCCTAACTGGCCGAGGTTTGACTTTGCTGTAACCC<br>R:GGCTAGCGAGCTCAGTGGATTGCCACCATCTCTCT | 6025 bp |
| **Tile 4** | **4.1** | F:TGGCCTAACTGGCCGGGTTCTTGCCACGTTCCCTA<br>R:TGGCCTAACTGGCCGGGTTCTTGCCACGTTCCCTA | 4762 bp |
| | **4.2** | F:TGGCCTAACTGGCCGGCATCAACCCCTTGGTGAGT<br>R:GGCTAGCGAGCTCAGTTCTAGCCCCTACCTGGAGC | 4714 bp |
| | **4.3** | F: TGGCCTAACTGGCCGGGTTCTTGCCACGTTCCCTA<br>R: GGCTAGCGAGCTCAGTATTGCTGGACAGCCCCATC | 4732 bp |
| **Tile 5** | **5.1** | F:TGGCCTAACTGGCCGCTCTACAGCAAGGGCAACCA<br>R:GGCTAGCGAGCTCAGGGTTCTCCGCCAGATACAGG | 3696 bp |
| | **5.2** | F:TGGCCTAACTGGCCGGTCTGGAGGCACTTTGGGAT<br>R:GGCTAGCGAGCTCAGCGCTTCATGTTGCTCTCTGC | 3990 bp |
| **Tile 8** | **8.1** | F:TGGCCTAACTGGCCGCCCCTTTGGGAAGACACGTT<br>R:GGCTAGCGAGCTCAGGTCACAGGTTAGGCAGATCC | 3096 bp |
| | **8.2** | F:TGGCCTAACTGGCCGAGAGCAAAGGAAGCGCCATA<br>R:GGCTAGCGAGCTCAGTGCCATAGAATGTAATCTGCCAT | 3729 bp |
| | **8.3** | F:TGGCCTAACTGGCCGAGGTCCACTGTCGCAATGTT<br>R:GGCTAGCGAGCTCAGCATGAGAAGGGGCTTGGG | 3716 bp |

Black letters represent gene-specific sequence. The green letters in front of each forward primer (F) and pink letters in front of each reverse primer (R) represent sequences homologous to the ends of the linearized target plasmid required for in-fusion cloning.

**Table 3.4 Primers for site-directed mutagenesis**

| Tiles | Sequence 5'→3' | Length | Mutation |
|---|---|---|---|
| rs1799991 | F: ATATAAAAATTAGGCCGAGCACAGTGGCTCATGCC<br>R: GGCATGAGCCACTGTGCTCGGCCTAATTTTTATAT | 35 bp | G → A |
| rs1799992 | F: GTGGCATGCGCCTGCAGTCCCAGCTACTT<br>R: AAGTAGCTGGGACTGCAGGCGCATGCCAC | 29 bp | T → C |
| rs2846282 | F: GGGTGACAGAGCGCAACTCCCCCTCAAAA<br>R: TTTTGAGGGGGAGTTGCGCTCTGTCACCC | 29 bp | G → A |

| | | | |
|---|---|---|---|
| rs11297272 | F:GTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAAGA GAAATGGGACCTCC<br>R:GGAGGTCCCATTTCTCTTTTTTTTTTTTTTTTTTTAGA CAGAGTCTCAC | 50 bp | G → - |
| rs1799993 | F: GGCAGCACTCTAGGTGAACGAACTTTAGGCAGC<br>R: GCTGCCTAAAGTTCGTTCACCTAGAGTGCTGCC | 33bp | C → A |

F stands for forward and R stands for reverse. All the red annotated letters represent mutation (from protective type to risk type) and SNP rs11297272 had deletion of G in the underlined position.

**Table 3.5 Primers used for sanger sequencing and colony PCR**

| Target | Sequence 5'→ 3' | Length |
|---|---|---|
| pGL4.23-*Luc2*_minP_empty | F: CTAGCAAAATAGGCTGTCCC | 20 bp |
| | R: CTTAATGTTTTTGGCATCTTCCA | 23 bp |

In this study we have used the vector specific primers to perform all the colony PCR and sanger sequencing. To perform sanger sequencing for PCR amplified product of tiles, we have used primers used to amplify the tiles. F stands for forward and R stands for reverse.

**Table 3.6 Overview of biological samples used for PCR optimization**.

| Patient code | Material | Genotype* | Ethical approval |
|---|---|---|---|
| Patient 752 | gDNA | C/A, heterozygous | 2018/2020 |
| Patient V564 | gDNA | C/A, heterozygous | 2018/2020 |
| Patient V584 | gDNA | C/A, heterozygous | 2018/2020 |
| Patient V588 | gDNA | C/A, heterozygous | 2018/2020 |
| Patient V620 | gDNA | C/A, heterozygous | 2018/2020 |
| Patient V621 | gDNA | C/A, heterozygous | 2018/2020 |
| A41WAT (human cell line) | gDNA | C/A, heterozygous | BIDMC, 2009-P-000101 |

Genotype* represents genotypes for SNP rs1799993.

**Table 3.7 Overview of final patient samples used for cloning of tiles.**

| Patient code | Material | Genotype | | Ethical approval |
|---|---|---|---|---|
| | | Genotype for SNP 1799993 | Haplotype (homozygous) | |
| V586 | gDNA | A/A | risk | 2018/2020 |
| V609 | gDNA | A/A | risk | 2018/2020 |
| V601 | gDNA | C/C | protective | 2018/2020 |
| V602 | gDNA | C/C | protective | 2018/2020 |

**Table 3.8 List of plasmids used as starting point for cloning and mutagenesis**

| Plasmid | Cat. No | Description |
|---|---|---|

| | | |
|---|---|---|
| pGL4.23-*Luc2*_minP_empty | E8411 | The empty vector was digested with restriction enzyme, before used for cloning by In-fusion cloning. |
| pGL4.23-minP_Tile_8a | Cloned in our laboratory | The plasmid was used to clone site-directed mutagenesis PCR amplified product. |

**Table 3.9 Overview of commercial kits**

| Product | Supplier | Cat. No | Application |
|---|---|---|---|
| BigDye™ Terminator V3.1 Cycle Sequencing | Termofisher | 4458687 | Sanger sequencing |
| HiSpeed® Plasmid Maxi Kit | Qiagen | 12663 | Plasmid/DNA purification |
| In-fusion® Snap Assenmbly master Mix | TaKaRa | 638947 | Molecular cloning |
| Luciferase Assay HITS kit | Bio Thema | 484-103 | Transcriptional activity |
| Platinum SuperFi II PCR Master Mix | Invitrogen | 12359-050 | Polymerase chain reaction |
| QIAquick Gel extraction Kit | Qiagen | 28704 | DNA purification |
| QuickChange II Site Directed Mutagenesis Kit | Agilient Technologies | 200523-5 | Site directed mutagenesis |
| Taq DNA Polymerase | Qiagen | 201205 | Polymerase Chain Reaction |

**Table 3.10 Materials for restriction enzyme digestion**

| Product | Supplier | Cat. No | Application |
|---|---|---|---|
| HindIII-HF | BioLabs | R3104S | Restriction enzyme |
| KpnI-HF | BioLabs | R3142S | Restriction enzyme |
| CutSmart Buffer | BioLabs | B7204S | Buffer for restriction enzyme |
| PCR grade water | VWR Life Science | 733-2573 | Water for various methods |
| Quick CIP | BioLabs | M0525 | Dephosphorylation of vector |

**Table 3.11 Agarose gel electrophoresis**

| Product | Supplier | Cat. No | Application |
|---|---|---|---|
| Gel loading Dye Purple 6x | BioLabs | B7024S | To dye samples |
| GelRed® Nucleic Acid Stain | Sigma-Aldrich | SCT123 | Detecting nucleic acids |
| GeneRuler 1 kb DNA ladder | Thermo Scientific | SM1333 | Molecular weight standard |
| SeaKem ® LE Agarose | Lonza | 5004 | Gel electrophoresis |

**Table 3.12 Transformation and competent bacterial cells**

| Product | Supplier | Cat. No | Application |
|---|---|---|---|
| One Shot Top 10 competent cells | Thermofisher | C404010 | Transformation |
| XL1-Blue Supercompetent cells | Agilient Technologies | 200236 | Transformation |
| XL 10-Gold Ultracompetent cells | Agilient Technologies | 200315 | Transformation |
| S.O.C medium | Invitrogen | C4040 | Microbial growth medium |
| Bacto™ Agar | BD Diagnostics | 214010 | To make agar plates |
| Difco™ LB Broth (Miller) | Sigma Aldrich | 244610 | Growth medium powder |
| Ampicillin, Sodium Salt | CalBioChem | 1712254 | Antibiotic, used as selection marker |

**Table 3.13 Components used to culture mammalian cells**

| Product | Supplier | Cat. No | Application |
|---|---|---|---|
| AmnioMAX™-C100 (1X) | Lonza | 12-604F | Medium for ME3 |
| DMEM 4.5 g/L Glucose & L-Glutamine | Gibco | 2027416 | Medium for COS-1 and 3T3-L1 cells |
| AmnioMAX™-C100 | Gibco | 2337521 | Supplement |
| Fetal Bovine Serum (FBS), 10% | Gibco | 2243865 | Growth medium |
| Penicillin-Streptomycin (PEST) | Sigma - Aldrich | P0781 | Antibiotics to prevent contamination |
| Trypsin-EDTA | Lonza | 25200056 | Detach adherent cells from the vessel. |
| Newborn Calf Serum (CS), 10% | Gibco | 16010-159 | Growth medium |

**Table 3.14 Reagents used for transfection**

| Product | Supplier | Cat. No | Application |
|---|---|---|---|
| Opti-MEM® 1 (1X) | Gibco | 2185845 | Optimize transfection |
| TransIT-LT1® | Mirus | 02083632 | Transfection reagent |
| TransIT-X2® System | Mirus | 02123410 | Transfection reagent |

**Table 3.15 Materials for luciferase assay**

| Product | Supplier | Cat. No |
|---|---|---|

| | | |
|---|---|---|
| ATP substrate | 200324 | BioThema |
| Luciferin substrate | 200317 | BioThema |
| Tris-EDTA Buffer (ATP reagent buffer) | 250316 | BioThema |

**Table 3.16 Overview of buffers, solutions and medium**

| Title | Components | Application |
|---|---|---|
| 50 X TAE | 50 mM EDTA disodium salt<br>2M Tris<br>1 M Acetic acid | - |
| 1X TAE | 20 mL 50 X TAE<br>980 mL Milli-Q water | Used for separating DNA fragments on agarose gel |
| 1,2% Agarose | 1,2 g Agarose<br>100 mL Milli-Q water | Gel for separating gel |
| Agar plates | 5 g Bacto-agar<br>200 mL LB medium | Solidified growth serf ace for microorganisms |
| LB medium (1 liter), Autoclaved | 25 g Difco LB broth<br>900 mL Milli-Q water<br>pH adjusted to 7.5 with NaOH | Bacterial growth medium |
| Luciferase lysis buffer | 1M (1M TrisAc, 20nM EDTA, pH 7.8)<br>1 M DTT<br>0.5 M EDTA<br>85% Glycerol<br>100% Triton-X<br>a$H_2O$ (autoclaved Milli-Q water) | Lysis cells, prior to measurement of luciferase activity. |
| PBS (Phosphate Buffer Saline) 1 liter | 0.2 g KCI<br>0.24 g KH2PO4<br>1.44 g Na2HPO4<br>6 g NaCl<br>800 mL Milli-Q water<br>pH adjusted to 7.4 with HCL<br>Milli-Q water to a total of 1 L | Cells are washed with PBS |

**Table 3.17 Technical equipment used to perform various experiments.**

| Product | Supplier | Application |
|---|---|---|
| Avanti J-26S XP centrifuge | Beckman Coulter | Harvest lysate |
| Countess II Invitrogen | Invitrogen | Counts cell |
| Dual-Action Shaker KL2 | Edmund Bühler GmbH | Shakes samples in cold room |

| | | |
|---|---|---|
| Electrophoresis Power supply | Amersham Bioscience | Gel electrophoresis |
| Eppendorf Centrifuge 5804 R | Sigma-Aldrich | Centrifuge |
| Eppendorf MiniSpin Microcentrifuge | Sigma-Aldrich | Centrifuge |
| Eppendorf Thermomixer 5436 | Sigma-Aldrich | Heat centrifuge |
| FLUOstar OPTIMA | BGM Labtech | Measures luciferase assay |
| Gel Doc™ EZ Imager | Bio-Rad | Visualize gel electrophoresis |
| GeneAmp® PCR System 9700 | Applied biosystems | PCR |
| Grant QBD1 block heater | Grant Instrument | Heating up samples |
| Heraeus Biofuge Primo centrifuge | Thermo Scientific | Spin down cell suspension |
| Incubator Hood TH30 | Edmund Bühler GmbH | Shaker sample in heat |
| Köttermann Fumehood | Köttermann systemlabor | Ventilation device to limit exposure to hazardous compounds |
| NanoDrop ND-1000 Spectrophotometer | Leica | Measure DNA concentration |
| Nikon Eclipse TS100 | Nikon | Microscopy |
| Ocaptaor® Bio | Erlab | PCR hood |
| Optima Data Analysis | BGM LabTech | Excel export Retrieval of luciferase raw data |
| Panasonic MCO-19 AIC UB-PE CO2 incubator | Panasonic Healthcare | Cell culture |
| Precision Balance MS3002TS/00 | Mettler Toledo | Weighing |
| Spectrafuge™ mini centrifuge | Sigma-Aldrich | Centrifuge |
| Termaks B8054 Incubator | Termaks AS | Incubator for warming up medium |
| Vacusafe Comfort | IBS Integra Bioscience | Aspiration |
| Vortex Mixer | Labnet international Inc | Vortex sample |

**Table 3.18 Other materials**

| Product | Supplier | Cat. No | Application |
|---|---|---|---|
| 1.5 mL Eppendorf tubes | VWRÒ International | 211-2164 | For cell suspension |
| 50 mL falcon tubes | Röhre | 1044623 | For various liquid samples |
| 15 mL tube | Röhre | 1043721 | For various liquid samples |

| | | | |
|---|---|---|---|
| Analytik Jena Ultraviolet Lamp Stand for Hand-Held UV Lamps | Analytik Jena | 18-0063-01 | Visualize bands on agarose gel |
| Beckman Coulter Flask | Beckman Coulter | 356011 | Flask for growth of microorganism |
| Countess™ Cell Counting Chamber Slide | Invitrogen | NANC01001 | Glass slide for cell counting |
| Disposable safety scalpels, Aesculap® | Vavantor, VWR | 10131090 | Sharp blade to cut gels. |
| DpnI | BioLabs | R0176S | Cleaves methylated DNA sequence |
| Disposable safety scalpels, Aesculap® | Vavantor, VWR | 10131090 | Sharp blade to cut gels. |
| Opaque 96 well plate | Porvair | 204003 | 96 well for cell lysate, luciferase |
| Reagent Reservoirs | VWR® International | 613-1174 | Wells for samples |
| Serologische Pipette 5 mL | Serologische | 861253001 | Pipette |
| Serologische Pipette 10 mL | Serologische | 06182021 | Pipette |
| Serologische Pipette 25 mL | Serologische | 861685001 | Pipette |
| Tissue Culture Plates | VWR® International | 10062-900 | 96-well plates |
| Thermo Scientific Nnclon™ delta surface | Thermo Scientific | 150350 | Cell dish |
| VWR® low temperature freezer vails | VWR® International | 10018-734 | Snap freeze cells |

# 4. Methods

Ethical approval of this project was granted by the Norwegian Regional Ethics Committees (REC). Patient genomic DNA was retrieved from the Western Norway Obesity Biobank (WNOB), where all donors have given a written informed consent. A previous master thesis in this lab attempted to characterize the 11q23.3 locus by designing and cloning 8 tiles[88]. However, due to challenges with polymerase chain reaction (PCR) and particularly the traditional restriction enzyme digestion/ligation approach, very few tiles were successfully cloned.

The current study attempts to address these challenges by designing a completely new set of primers and thereby dividing the 11q23.3 locus/ haploblock into 8 more equally sized, partially overlapping tiles. Moreover, to increase the chances of success, three alternative primer pairs for each tile were designed (section 4.1). Finally, instead of relying on restriction enzyme digestion of the tiles followed by ligation with the target plasmid, a ligation-independent cloning method (in-fusion cloning, section 4.3.5) was applied. An overview of the workflow used in this thesis is presented in figure 4.1.

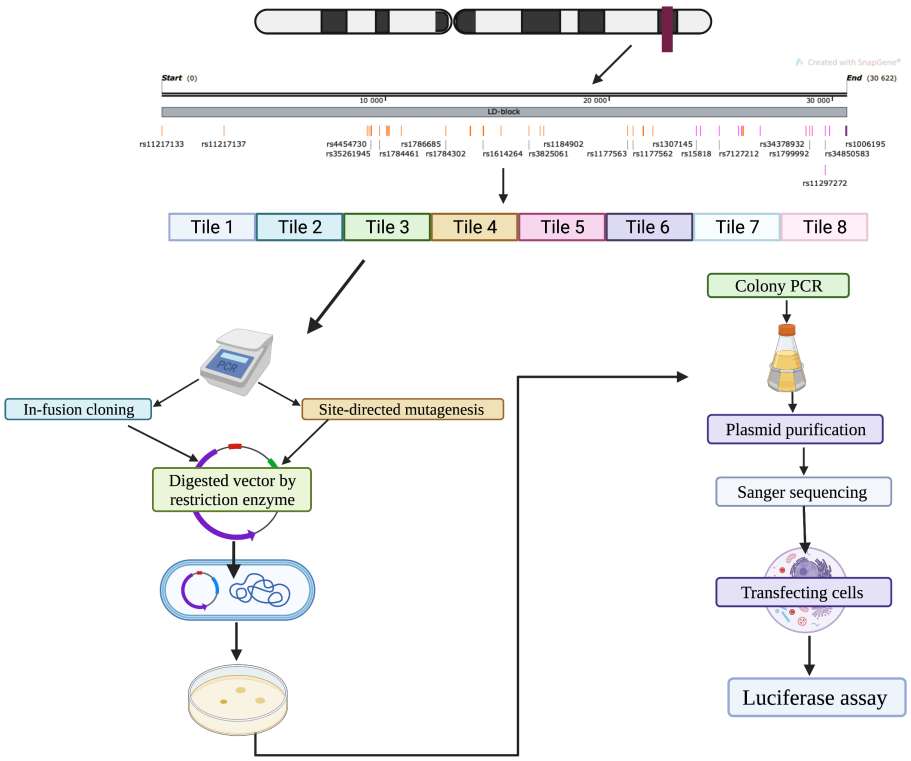



**Figure 4.1 The workflow of the methods used in this study**. The VAT associated 11q23.3 locus/haploblock was divided into 8 equal sized regions (tiles) containing all the SNPs in the

haploblock. Then web tools were used to create primer for each tile, before performing PCR on genomic template DNA. The PCR fragments were analyzed by agarose gel electrophoresis, and purified PCR fragments were used for in-fusion cloning into the restriction enzyme digested pGL4.23-*luc2*_minP_empty vector. Additionally, a pre-cloned plasmid was used for site-directed mutagenesis of individual SNPS to generate multiple plasmids with one mutation each. *E. coli* bacterial cells were transformed with the cloned or mutated vector containing desired sequence. Colony PCR was performed for further verification before further growth of transformed cells and plasmid purification. Following plasmid purification, Sanger sequence was performed to further verify the desired cloned tiles. Mammalian cells were transfected with the plasmid containing desired tiles/sequence, prior to luciferase assay to assess enhancer activity of each tile. This figure was adapted from Biorender.com.

To assess the enhancer activity of each tile, all tiles were cloned into the pGL4.23_*luc2*_minP empty reporter vector (figure 4.2). This vector contains the firefly luciferase reporter gene *luc2* under the control of a minimal promoter (minP). Thus, the empty vector expresses the luciferase gene, but at a very low, basal level. To test the enhancer activity of the different parts of the 11q23.3 locus, PCR-amplified regions were cloned into the multiple cloning site situated just upstream of the minimal promoter. Any enhancer activity would boost the minimal promoter and increase the luciferase gene expression. The vector also contains ampicillin resistance gene, which allows selection in *Escherichia Coli* (*E. coli)*[89].

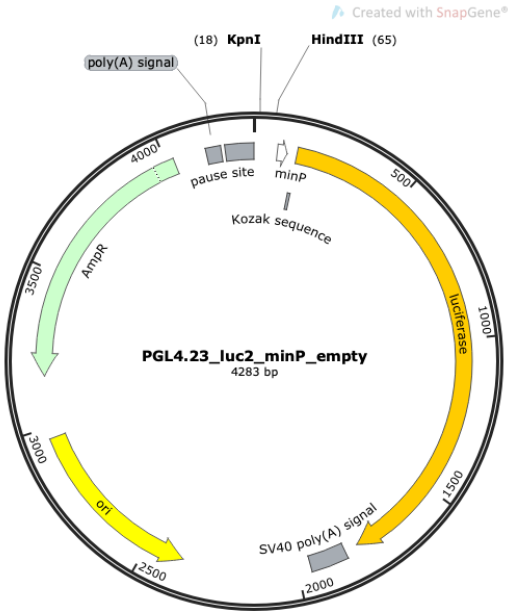



**Figure 4.2 pGL4.23_*luc2*_minP_empty vector map**. Multiple cloning region is situated between 1-70 base pairs, but only the 2 compatible restriction enzyme sites (KpnI and HindIII) are shown. MinP stands for minimal TATA-box promoter with low basal activity. Kozak sequence is the vertebrate consensus sequence for strong initiation for translation. The vector also contains sequence coding *luc2* gene, poly(A)signal, ori, ampicillin gene, and pause site. This figure was retrieved from SnapGene.

## 4.1 Choosing restriction enzymes

To clone PCR amplified tiles into the pGL4.23_*luc2*_minP empty vector, the circular plasmid needed to be linearized. The SnapGene software was used to view all the restriction enzyme (RE) sites present in the vector. Only RE sites unique to the multiple cloning site were considered. Moreover, only REs with a single site within the cloning site were considered. The NebCloner online tool was used to check the compatibility between restriction enzymes and various buffers (table 3.1). When further considering enzymes with high fidelity, 100% activity in the CutSmart buffer, of Time-Saver quality and not sensitive to methylation, and good stock availability in the lab, KpnI and HindIII considered all criteria and were selected for functional testing.

## 4.2 Primer design

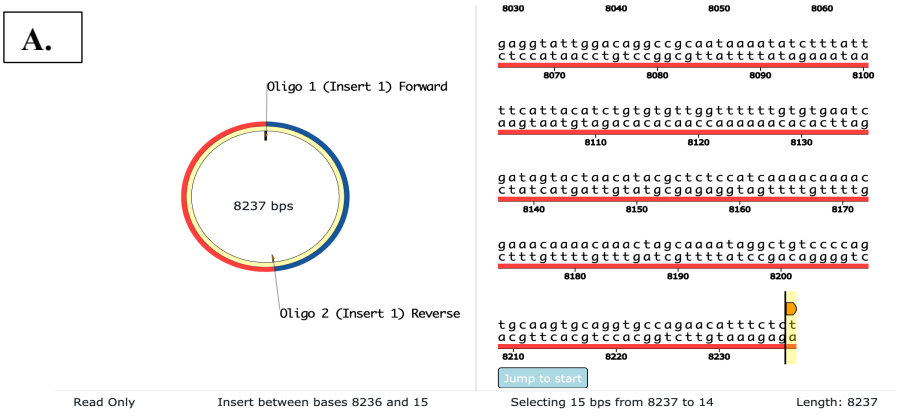
The visceral obesity-associated 11q23.3 locus is a 30 kb haploblock; a human genomic sequence (chr11:119,057,538-119,088,159, GRCh38/hg38) covering 38 SNPs in high LD with rs1799993. While previous studies had cloned minor parts of this haploblock for functional analyses[24,88], most of the sequence had not been cloned. In this thesis we aimed to clone the missing parts of the LD block. The 30 kb genomic sequence was retrieved from the UCSC genome browser and analyzed in the SnapGene software. Here, the total sequence was didivided into 8 regions/tiles of roughly equal size.

Primers for all the tiles used in this study were designed in NCBI primer blast (Table 3.1) by copying the sequence of the desired tile from SnapGene. Three alternative primers were created for each tile and additionally each primer set was checked for cross-dimers, self-dimers, hair-pin values, primer melting temperature (Tm) and GC range in Beacon PrimerSoft website (table 3.1). Primer sets were chosen with Tm between 59-65°C, GC range between 40-60%, cross-dimer, self-dimer, hairpin values between -2.0 and 0, with 0 being optimal (the larger the negative value for these categories the higher the chance to get either cross-dimer, self-dimer, or hairpin).

In addition, once all the primers were chosen based on these characteristics, 15 additional base pairs for in-fusion cloning were added to the 5′ end of each forward primer (5`-**TGGCCTAACTGGCCG-**3′) and 5′ end of each reverse primer (5′-**GGCTAGCGAGCTCAG-**3′) (Table 3.3). These sequences were homologues to the linearized ends of the pGL4.23-*luc2*_minP_empty vector, and were designed in TaKaRa

website (table 3.1), by providing the original sequence of the pGL4.23-*luc2*_minP_empty vector from SnapGene, the desired tile sequence and the selected restriction enzyme. Figure 4.3A shows an example of a typical output from this software.

Primers for site-directed mutagenesis were designed in QuickChange Primer Design (Agilent) website (table 3.1). Five mutations were selected for site-directed mutagenesis (Table 3.4), and for each mutation, 70 base pairs in total (meaning 35 base pairs from each side of the SNP) were copied from SnapGene and used to search for primers on the website. Then based on the SNP position and type of mutation, primers were created with an optimal Tm between 78-80 °C (Figure 4.3B).




2 PCR Fragment Overlaps

```
Destination Vector(+): ...ATTTCTCTGGCCTAACTGGCCGGTAC -3'     5'- CTGAGCTCGCTAGCCTCGAGGA...
Destination Vector(-): ...TAAAGAGACCGGATTGACCGGC -5'     3'- CATGGACTCGAGCGATCGGAGCTCCT...
Insert 1(+)          :       5'- TGGCCTAACTGGCCGGGGGGAG...CCAAACCCTGAGCTCGCTAGCC -3'
Insert 1(-)          :       3'- ACCGGATTGACCGGCCCCCCTC...GGTTTGGGACTCGAGCGATCGG -5'
```

3 Output protocol

**Oligos to Order**

| Oligo Name | Sequence | Length (bp) | GC% | Tm (°C) | Gene Specific Tm (°C) |
|---|---|---|---|---|---|
| Oligo 1 (Insert 1) Forward | TGGCCTAACTGGCCGGGGGGAGGCGAAGAAAGAAAAC | 37 | 59 | 77.7 | 63.8 |
| Oligo 2 (Insert 1) Reverse | GGCTAGCGAGCTCAGGGTTTGGGTTGGCATCAAGG | 35 | 60 | 76.8 | 62.0 |

**PCR Reactions to Run**

| Reaction Number | Forward Oligo | Reverse Oligo | Template Name |
|---|---|---|---|
| 1 | Oligo 1 (Insert 1) Forward | Oligo 2 (Insert 1) Reverse | Insert 1 |

**B.**

**Primer pair → rs1799992: T→ C**

**S29F_rs92_T-C**

5'-gtggcatgcgcctgcagtcccagctactt-3'

**S30R_rs92_T-C**

5´-aagtagctgggactgcaggcgcatgccac-3´

Sequence:

ACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGTAGTCCCAGCTACTTGGGAGGCTAAGAC
AGGAGAAT

**Primer sequences:**

| Primer Name | Primer Sequence (5' to 3') |
|---|---|
| | 5'-aagtagctgggactgcaggcgcatgccac-3' |
| t35c_ | 5'-gtggcatgcgcctgcagtcccagctactt-3' |

**Oligonucleotide information:**

| Primer Name | Length (nt.) | Tm | Duplex Energy at 68 °C | Energy Cost of Mismatches |
|---|---|---|---|---|
| | 29 | 78.81°C | -46.15 kcal/mole | 2.88% |
| t35c_ | 29 | 78.81°C | -44.53 kcal/mole | 3.03% |

**Primer-template duplexes:**

| Primer Name | Primer-Template Duplex |
|---|---|
| | gtggtggcatgcgcctgtagtcccagctacttggg<br>\|\|\|\|\|\|\|\|\|\|\|\|\|\|\| \|\|\|\|\|\|\|\|\|\|\|\|\|\|\|<br>3'-caccgtacgcggacgtcagggtcgatgaa-5' |
| t35c_ | 5'-gtggcatgcgcctgcagtcccagctactt-3'<br>\|\|\|\|\|\|\|\|\|\|\|\|\|\|\| \|\|\|\|\|\|\|\|\|\|\|\|\|\|\|<br>caccaccgtacgcggacatcagggtcgatgaaccc |

**Figure 4.3: Examples of original primer design and inserts from two different websites. A.** Designed 15 homologous base pairs from TaKaRa website. Plasmid sequence –pGL4.23-*luc2*_minP_empty was copied from SnapGene and KpnI as restriction enzyme, tile 1 as insert. From the search two common 15 homologues base pair inserts were created for all the primers. As the 3. Output protocol parameter shows oligo 1 (insert 1) forward and oligo 2 (insert 1) reverse, which is marked with orange underline. **B.** Example of primer designing for site-directed mutagenesis, figure shows primer for SNP rs1799992, where base T replaced with C. Sequence track represents the original sequence from SnapGene and Primer-template duplex track shows the changes in the primers itself.

## 4.3 DNA techniques

### 4.3.1 Restriction enzyme digestion

Restriction enzyme digestion is a method that involves enzymatical fragmentation of DNA. Restriction enzymes, also known as endonuclease, are naturally occurring enzymes that recognize short nucleotide sequence and cleave double stranded DNA at specific sites[90]. These enzymes can either cut DNA in a staggered manner (single-strand overhang) or cut in a manner that produces blunt end. Here, restriction enzyme digestion was only needed on the vector. As mentioned above, KpnI and HindIII were found to be suitable enzymes for this vector.

**Procedure:**

In order to test which of the two RE performed best, restriction enzyme digestion reactions were set up according to Table 4.1. Both KpnI or HindIII reactions were incubated at both 15 minutes and 1 hour at 37 °C. After digestion, the samples were dephosphorylated by addition

of 2.5 $\mu L$ quick CIP and incubated for 10 min at 37 °C to prevent self-annealing of digested vector, prior to heat incubation for 2 minutes at 80°C. Finally, the digested products were run on an agarose gel by gel electrophoresis (section 4.3.2) for either verification analysis or DNA fragment purification (section 4.3.3).

**Table 4.1: General restriction digestion set up**.

| Components | Amount | Samples | First incubation |
|---|---|---|---|
| Plasmid DNA | 2 µg | Plasmid DNA + KpnI-HF | 15 minutes |
| rCutSmart Buffer (10X) | 5 $\mu L$ | Plasmid DNA + KpnI-HF | 1 hour |
| Restriction enzyme (20 units) | 1 $\mu L$ | Plasmid DNA + HindIII-HF | 15 minutes |
| Nuclease free water | To a final volume of 50 $\mu L$ | Plasmid DNA + HindIII-HF | 1 hour |
| **Total volume** | **50 $\mu L$** | Positive control | 1 hour |
| | | Negative control | |

For this set up uncut pGL4.23_*luc2*_minP_empty was used as plasmid DNA. Controls used in the restriction enzyme experiments were: plasmid + no enzyme (positive control) and enzyme + no plasmid (negative control).

### 4.3.2 Agarose gel electrophoresis

Agarose gel electrophoresis is a method used to separate various sizes of DNA fragments, ranging from 100 base pairs to 25 kilobases [91]. This method is used to verify fragment isolation, plasmid, and estimation of DNA concentration. In gel electrophoresis DNA is separated based on the phosphate backbone of the DNA, which causes the DNA to migrate through the agarose polymer matrix when electric current is applied[91]. In other word DNA fragments migrate towards positively charged anode, and the size of the DNA fragment determines how fast the fragments migrate through gel. Thus, longer DNA fragments migrate slower that shorter fragments. In this study 1,2% agarose gels were used based on the size of the DNA fragments. Furthermore, GelRed® was used for post-electrophoresis gel staining. GelRed® is a very sensitive, stable, and environmentally safe fluorescent nucleic dye that was designed to replace the very toxic ethidium bromide[92]. Figure 4.4 represents the general workflow of gel electrophoresis.

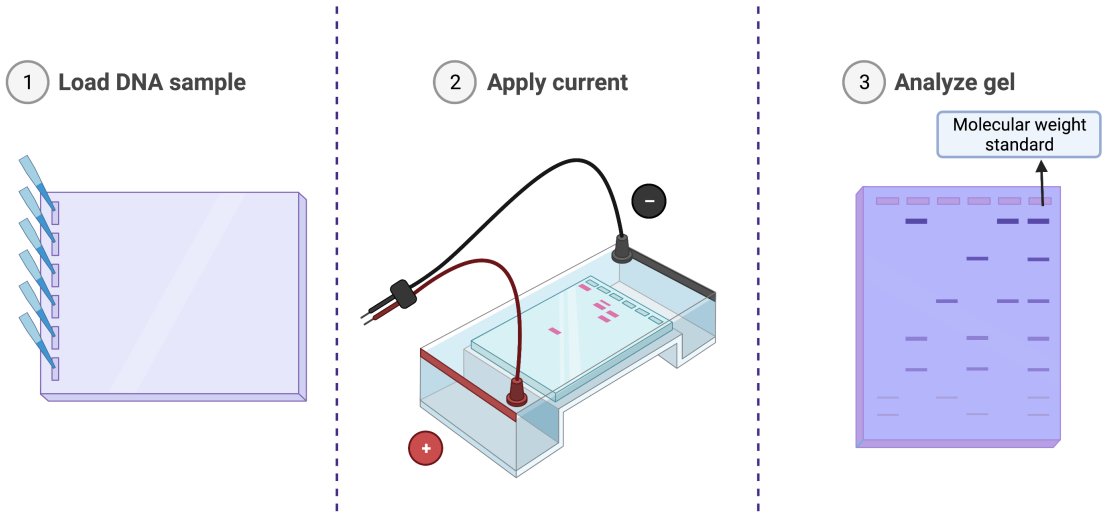



**Figure 4.4: The workflow of gel electrophoresis**. As the figure shows, first one must make the gel, prior to loading the samples and molecular weight. Once the samples are loaded, then the electrical current is supplied. Lastly, analyzing the gel after the desired DNA fragment have migrated through the gel. This figure was adapted and created in the Biorender.com.

**Procedure:**

1,2% agarose gel was prepared by heating up the 1xTAE buffer and agarose powder mixture in the microwave. The mixture was microwaved until all agarose powder was dissolved; the agarose gel solution was cooled down to 60°C before pouring into a gel cast. For 30 ml agarose gel solution 3 $\mu L$ GelRed® was added to stain the gel. Once the gel solidified, it was placed in the gel tank and filled with 1xTAE buffer, until the gel was submerged completely in the buffer. The samples were mixed with 6X loading dye before loading in the gel wells and a molecular weight standard. The gel was run at 80-100 V for 45-90 minutes depending on the DNA fragment sizes. Lastly, the gel was visualized in the Gel Doc™ EZ system and Image lab software (table 3.2) was used to edit and analyze the gel image.

### 4.3.3 DNA purification

DNA purification is a process where desired DNA fragments are isolated from agarose gel by using standard DNA purification kits. The principle of DNA purification is binding of the DNA fragment to a silica membrane in the presence of chaotropic salts. These salts allow removal and washing of any contaminated associated with the extracted DNA, before eluting purified DNA[93]. In this study, QIAquick Gel Extraction Kit was used to purify extracted DNA fragments.

**Procedure:**

DNA bands were visualized by Ultraviolet Lamp Stand at 218 nm, and the desired band in the gel was sliced with a clean scalpel. Sliced gel containing the DNA fragment was placed in a 1.5 ml tube. After weighing the gel, DNA was purified according to the manufacturers protocol. Figure 4.5 shows the basic steps of this DNA purification method. Briefly, 3 volumes Buffer QG to 1 volume gel (100 mg gel ~ 100$\mu L$) was added to the extracted gel and incubated at 50°C for 10 minutes. Then 1 gel volume isopropanol was added to the mixture and mixed. The mixture was applied to the QIAquick column collection tube and centrifuge for 1 minute at 13000 rpm. 500 $\mu L$ Buffer QG was applied to the QIAquick column collection tube and centrifuged for 1 minute at 13000 rpm. The QIAquick column collection tube was washed with 750 $\mu L$ Buffer PE for 1 minute at 13000 rpm. QIAquick collection tube was discarded and QIAquick column was placed into a 1.5 mL microcentrifuge tube. 30 nuclease free water was added in the center of the QIAquick column and centrifuged for 1 minute at 13000 rpm.

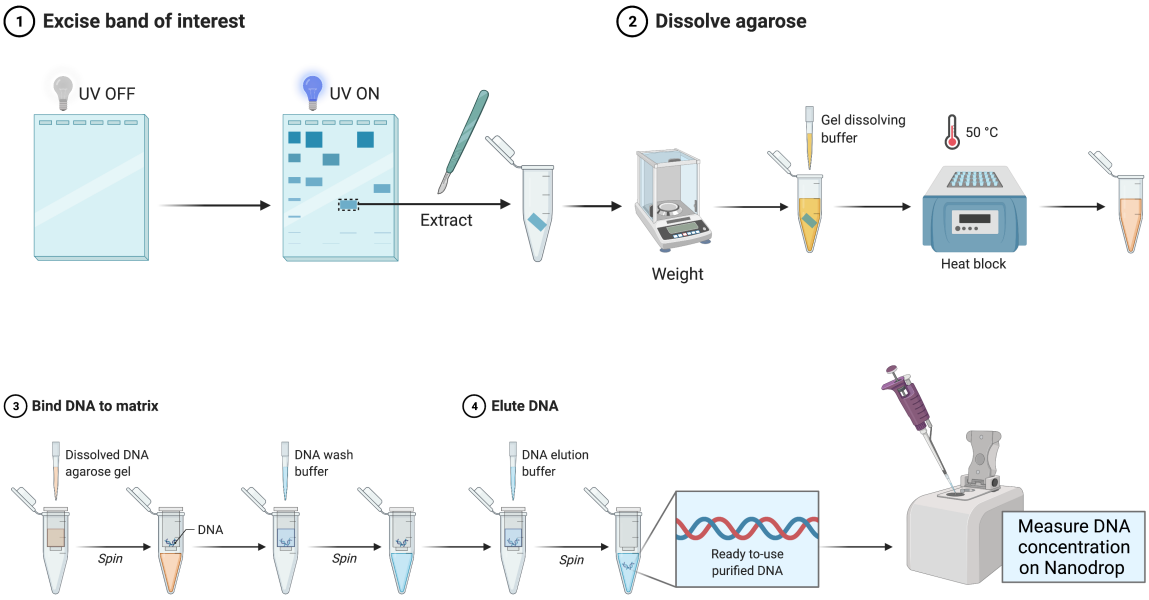



**Figure 4.5 The workflow of DNA extraction and purification**. After visualizing the band extracting DNA and weighing, gel dissolving buffer was added and incubated on 50°C for 10 minutes in a heating block. Isopropanol was added and mixed thoroughly. The sample then placed on spin column and spined down, before adding washing buffer into the column and spinned down again. Lastly, the column was placed in a 1.5 ml tube and to elute nuclease free water was added, before spinning down for the last time. Concentration of the purified DNA was measured on the NanoDrop ND1000 program. This figure was adapted from Biorender.com.

### 4.3.4 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) is a well-known highly sensitive method for DNA amplification. PCR method involves three main cycling steps: DNA denaturation, primer

annealing and DNA extension via heat-stable DNA polymerase. Denaturation of the double stranded DNA refers to disrupting the hydrogen bonds that hold the DNA strands together when heat is applied (approximately 95°C). Secondly, the heat is reduced in a manner so that the primers can anneal to their specific single-stranded DNA template. Lastly, the heat-stable polymerase extends the annealed primers to synthesize new DNA strand from the 3´to 5´ directions by assembling free-nucleotides in the PCR solution. This cycle is repeated around 20-35 cycles and results in exponential replication of the desired DNA template, as visualized in figure 4.6. In this study, different PCR kit was used depending on the type of experiments, these includes Platinum™ SuperFi II Green PCR Master Mix (PCR reaction for all the tiles), QuikChange II Site-Directed Mutagenesis Kit (site-directed mutagenesis, section 4.3.6), Taq DNA polymerase (colony PCR, section 4.3.7) and BigDye™ Terminator v3.1 Cycle Sequencing Kit (sanger sequencing, section 4.3.8). In addition, in this study multiple sets of PCR programs were used for different experiments. In this section only PCR program for amplification of different tiles for cloning are shown (Table 4.2) and the other PCR programs are shown later.

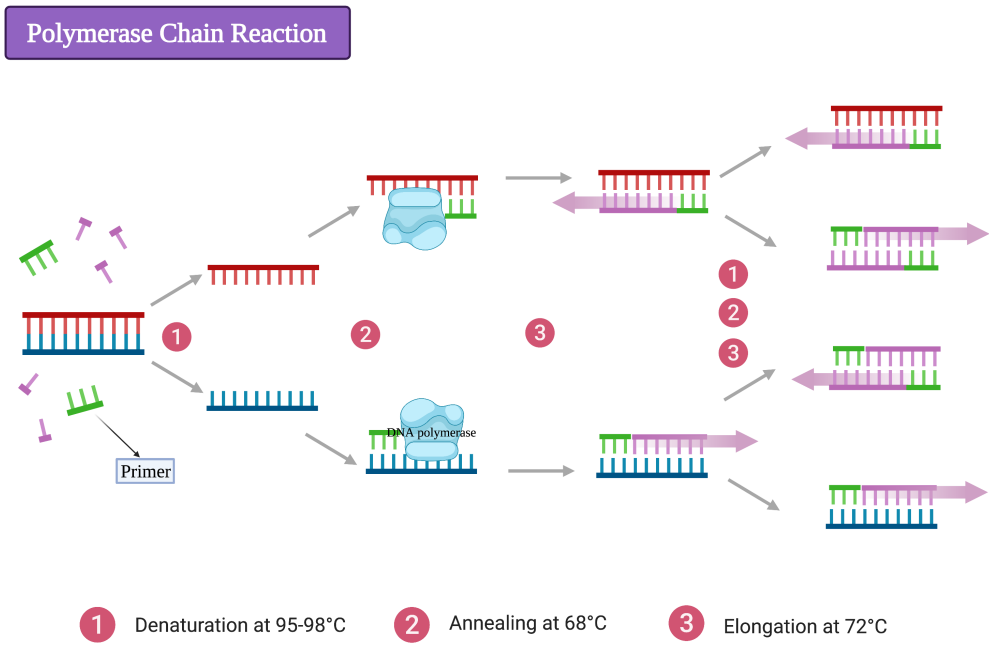



**Figure 4.6: Schematic representation of polymerase chain reaction (PCR).** The figure represents the three main step in PCR: denaturation, annealing and extension. This figure was adapted from Biorender.com.

**Table 4.2: The thermal cycling program used in the PCR amplification of tiles from genomic DNA.**

| Set up* | | |
|---|---|---|
| | **Temperature** | **Time** |

| | | | |
|---|---|---|---|
| Initial denaturation | | 98° C | 30 seconds |
| 35 cycles | Denaturation | 98° C | 15 seconds |
| | Anneal | 65° C** | 15 seconds |
| | Extend | 72° C | 2,5 minutes |
| Final extension | | 72° C | 5 minutes |
| | | 4° C | Hold |
| **Set up** | | | |
| | | **Temperature** | **Time** |
| Initial denaturation | | 98° C | 30 seconds |
| 10 cycles | Denaturation | 98° C | 30 seconds |
| | Anneal | 60° C ** | 15 seconds |
| | Extend | 72° C | 2 minutes |
| 30 cycles** | Denaturation | 98° C | 15 seconds |
| | Anneal | 55° C** | 15 seconds |
| | Extend | 72° C | 2 minutes |
| Final extension | | 72° C | 5 minutes |
| | | 4° C | Hold |

* Represents the most utilized PCR program in the amplification of DNA fragments. ** Represents the parameters that were modified multiple times throughout the whole study.

### 4.3.5 In-fusion cloning

In-fusion cloning is a very easy and very fast cloning method that allows ligation-independent, directional cloning of any PCR fragments or multiple fragments into any vector[94]. In this study, In-Fusion® Snap Assembly Master Mix kit was used to perform all the cloning. Principle of this method involves linearized vector (digested vector with 15 base pair homologous ends, section 4.3.1) and desired amplified inserts with homologous ends via PCR (section 4.3.4) is combined with in-fusion snap assembly master mix and incubated for 15 minutes. The incubation leads to enzyme in the solution to remove nucleotides from the 3´primers ends of the linear DNA, allowing complementary base pairs of the vector and insert to join and anneal in the right direction[94], figure 4.7.

**Procedure:**

Prior to in-fusion cloning, the desired tiles were amplified via PCR and the PCR samples were run in agarose gel (gel electrophoresis, section 4.3.2) to detect the right insert. Once the right insert was detected on the gel, DNA purification was performed (section 4.3.3). Then the purified fragments along with linearized vector was combined with In-fusion Snap Assembly Master mix, according to table 4.3. The solution was incubated for 15 minutes at 50° C and then frozen at -20° C for later use. Positive and negative controls were provided by the kit.

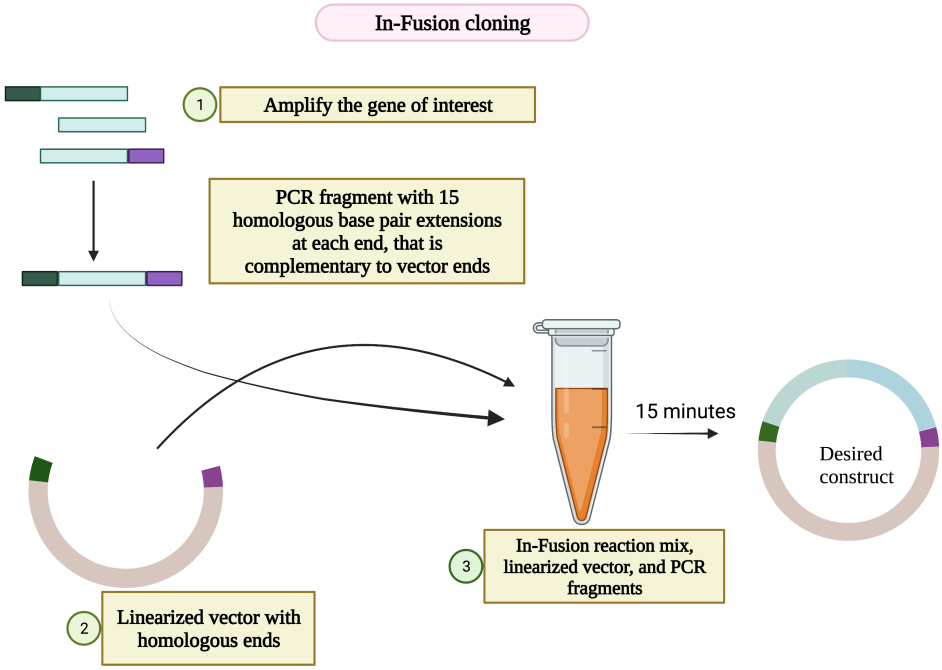



**Figure 4.7: Workflow of in-fusion cloning**. The figure describes the main steps of in-fusion cloning. This figure was adapted from Biorender.com.

**Table 4.3: In-Fusion® Snap Assembly Master Mix kit, reaction set up**.

| Components | Volume |
|---|---|
| 5X In-fusion Snap Assembly | 2 µL |
| Linearized vector (pGL4.23_digested) | 1:2 molar ratio |
| Purified PCR fragments (Tiles) | 1:2 molar ratio |
| Nuclease free water | To a final volume of 10 µL |
| **Total volume** | **10 µL** |

Optimal amount of vector and insert used in cloning reaction was retrieved from in-fusion molar ration calculator: https://www.takarabio.com/learning-centers/cloning/primer-design-and-other-tools/in-fusion-molar-ratio-calculator

### 4.3.6 Site-directed mutagenesis

Site-directed mutagenesis is a technique that involves custom made primers and PCR to create specific point mutation, substitution, or deletion in double stranded plasmid DNA[95]. In this study, QuikChange II Site-Directed Mutagenesis Kit was used to introduce 5 different mutations in pGL4.23_Tile 8a protective type (discussed later). The kit contains PfuUltra high-fidelity DNA polymerase suited for mutagenic primer-directed replication[95]. Furthermore, the kit contains DpnI endonuclease that only cleaves its recognition site when the DNA is methylated, thus digests the parental DNA template, meanwhile phosphorylates the PCR product so that mutation containing synthesized DNA can ligate and circularize[95].

**Procedure:**

Site directed mutagenesis reactions were assembled in small PCR tubes according to table 4.4, including positive control, which was provided by the kit. The reagents were mixed thoroughly and placed in a thermocycler. The thermal cycle was performed according to table 4.5. After amplification by PCR, the reaction mixed was digested with DnpI for 1 hour at 37°C. All the samples including control was run in agarose gel (gel electrophoresis section 4.3.2) to detect the right band size. Once the right band was detected the samples were frozen at -20°C for later use.

**Table 4.4 Quick Change II Site-Directed Mutagenesis Kit reaction set up**

| Components | Volume |
|---|---|
| Nuclease free water | 17,7 µL |
| 10X Pfu buffer | 2,5 µL |
| Primer 1 (100 ng/ µL) ca. 7 µM | 0,6 µL |
| Primer 2 (100 ng/ µL) ca. 7 µM | 0,6 µL |
| dNTP mix | 1 µL |
| PfuUltra HF DNA polymerase (2.5 U/ µL) | 0,5 µL |
| Double stranded DNA template 10 ng/µL | 2,0 µL |
| **Total volume** | **25 µL** |

For this experiment double stranded DNA was used –pGL4.23_Tile 8a_WT_protective, WT stands for wild type.

**Table 4.5 The thermal cycling program used in the site-directed mutagenesis.**

| Cycles | Temperature (°C) | Time |
|---|---|---|
| 1 (hold) | 95 °C | 30 sec |
| 14 cycles | 95 °C | 30 sec |
| | 60 °C | 1 min |
| | 68 °C | 10 min* |
| Hold | 4°C | ∞ |

* Represents time –1 minutes/kilo base pairs of plasmid size.

### 4.3.7 Transformation of *E. coli*

Bacterial transformation is a method where competent bacterial cells can take up foreign plasmid DNA from the environment under the right conditions, such as heat shock. The introduced plasmid DNA will be replicated with the bacterial genome when the bacterial cells multiply[96]. This technique is mostly used for amplification of plasmid DNA. In this study, initially three commercially available competent *Escherichia Coli* cells including TOP10 *E*

*Coli.*, XL 10-gold *E Coli.*, and XL-1 Blue *E Coli*. cells were tested to identify the best performing competent cells.

**Procedure:**

For the transformation test, all three types of competent cells mentioned above were thawed on ice. Then different amounts of plasmid DNA (positive control from in-fusion cloning (section 4.3.5) was used as plasmid DNA) were added to 25 µL of competent cells (figure 4.8). These samples were incubated on ice for 30 minutes, before heat shocking the samples for 30 seconds at 42°C and placed them on ice for 2 minutes. 125 µL S.O.C medium was added to each sample tube and placed on shaking incubator at 37°C for 1 hour at 225 rpm. After incubation, 80 µL of samples were plated on agar containing 100 µg/ml ampicillin and the agar plates were incubated upside-down at 37°C overnight. The next day, transformation efficiency was evaluated visually based on the number of colonies formed. To screen for positive clones, 6 individual colonies were selected for further analyses by colony PCR (section 4.3.7, subsection A) and gel electrophoresis (section 4.3.2). Furthermore, positive clones were amplified and subjected to plasmid DNA purification (section 4.3.7, subsection C) and sanger sequencing (section 4.3.8).

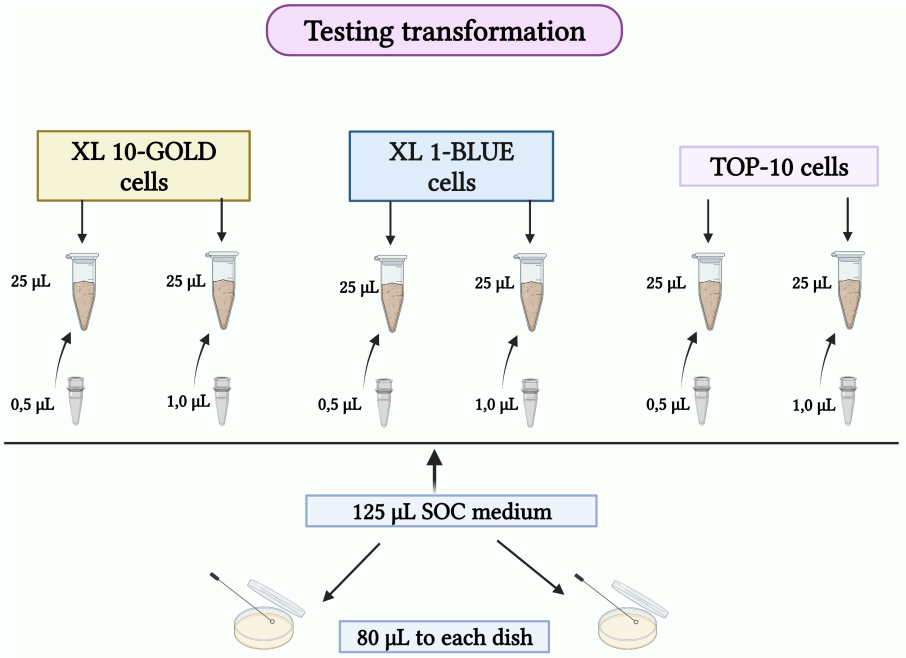



**Figure 4.8 Optimalization of transformation**. As the figure shows tow tubes of 25 µL XL-10 Gold, XL1-Blue and TOP-10 competent cells. For each type of cells 0.5 µL and 1.0 µL of plasmid DNA was added. 125 µL of S.O.C medium was added to each tube before plating on agar plates contained ampicillin. This figure was adapted from Biorender.com.

### A. Colony PCR

Colony PCR is a screening method that distinguishes the colonies carrying the desired construct in contrast to colonies carrying only the selective marker[97]. This method involves custom designed primers that will only yield a specific product with known size in the presence of desired construct. Following PCR amplification, the DNA fragments are visualized on agarose gel. Thus, this method is quite powerful as it rapidly screens through many colonies to distinguish false positives from true positives[97]. Figure 4.9 shows the workflow of general colony PCR.

**Procedure:**

Six single colonies of each construct were added to six PCR tubes containing 10 µL of sterile Phosphate Buffer Saline (PBS). The mixture was mixed carefully and thoroughly by pipetting up and down with filter tips. PCR reaction mix was set up according to table 4.6 and 2.5 µL of PBS colony mixture was added to PCR reaction mix. The PCR solution was then placed on a thermocycler and followed the thermal cycle program according to table 4.7. Following PCR amplification, the PCR fragments were visualized on agarose gel, and based on the result, a single colony was further expanded (section 4.3.7, subsection B).

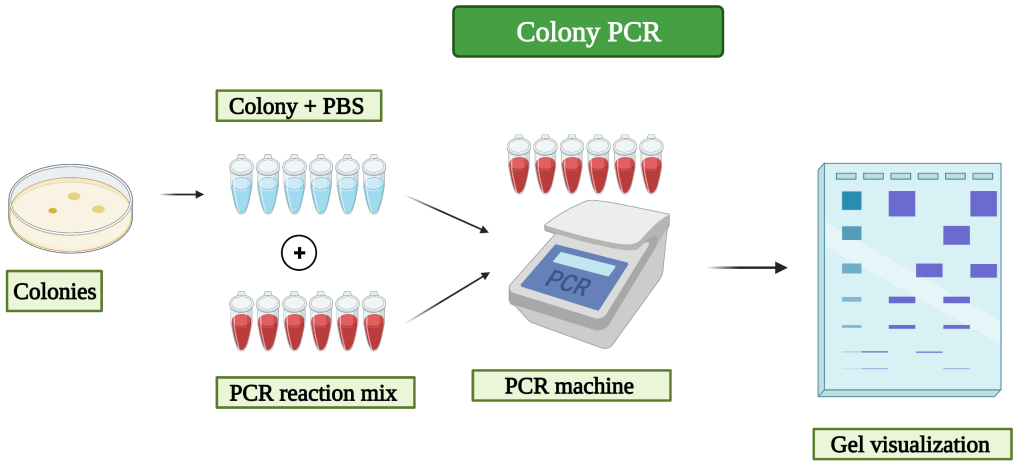



**Figure 4.9 Schematic representation of colony PCR.** Individual colonies are picked up and placed in tubes containing Phosphate Buffer Saline (PBS). After mixing gently PBS colony reaction mix is transferred to PCR reaction mix. After PCR amplification the samples are visualized on gel. This figure was adapted from Biorender.com.

**Table 4.6 QIAGEN Multiplex PCR Plus Kit reaction set up.**

| Components | Volume |
|---|---|
| 10X coral red buffer | 2,5 µL |
| dNTP mix 10 mM each | 0,5 µL |

| | |
|---|---|
| Primer 1 (20 µM) | 0,5 µL |
| Primer 2 (20 µM) | 0,5 µL |
| Q-solution | 5 µL |
| Taq-polymerase | 0,1 µL |
| Nuclease free water | 13,4 µL |
| **Total volume** | **22,5 µL** |

Primers used for colony PCR were forward primer: 01F_RVprimer3_pGL3 and reverse primer: 579R_RVprimer_pGL4 (Table 3.5).

**Table 4.7 The thermal cycling program used for colony PCR.**

| Set up | | | |
|---|---|---|---|
| | | **Temperature** | **Time** |
| Initial denaturation | | 96° C | 5 minutes |
| 10 cycles | Denaturation | 96° C | 30 seconds |
| | Anneal | 60° C | 15 seconds |
| | Extend | 72° C | 5 minutes |
| 25 cycles | Denaturation | 96° C | 30 seconds |
| | Anneal | 55° C | 15 seconds |
| | Extend | 72° C | 5 minutes |
| Final extension | | 72° C | 5 minutes |
| | | 4° C | Hold |

### B. Growth of *E coli.*

After verification via colony PCR (section 4.3.7, subsection A) positive clones were grown in 5 ml lysogeny broth (LB) medium with appropriate amount of ampicillin in shaker incubator at 37°C overnight at 225 rpm. The next day 2.5 ml of bacterial growth culture was transferred into 150 ml LB medium with appropriate amount of ampicillin, before placing the flasks into shaker incubator at 37°C overnight at 225 rpm. Plasmid purification (section4.3.7, subsection C) and glycerol stock (section4.3.7, subsection D) was performed on the bacterial cultures from the big flasks.

### C. Plasmid DNA purification

Plasmid DNA purification is a method that involves extraction and purification of nucleic acids from competent cells. To this date, various plasmid DNA isolation/purification kit exist and the general steps include lysis of cell membranes, removal of histone proteins, lipids, and RNA and purification[98]. In this study QIAGEN Plasmid Plus Kit was used to purify plasmid DNA.

**Procedure:**

Bacterial culture from the big flasks (section 4.3.7, subsection B), was centrifuged for 20 minutes at 6000 rpm at 4°C, then the supernatant was discarded. Following centrifuge, the plasmid purification was performed by QIAGEN Plasmid Plus Kit and mostly followed manufacturers protocol with few modifications. 10 ml P1 buffer with 0,03 mg/ml RNase A and 300X LyseBlue was used to resuspend bacterial pellet. Then 10 ml P2 buffer was added and incubated at room temperature for exactly 5 minutes. After incubation, 10 ml cold P3 buffer was added, and the contents then transferred to QIAfilter Maxi Cartridge plugged with a Qiafilter™ cartridge cap and incubated for 10 minutes in room temperature. In the meantime, Hispeed® Maxi Tip was equilibrated with 10 ml QBT buffer before filtering the contents from QIAfilter Maxi Cartridge with Qiafilter™ Plunger through Hispeed® Maxi Tip. The Hispeed® Maxi Tip was washed with 60 ml QC buffer and eluted with 15 ml QF buffer. Following elute 10.5 ml isopropanol was added to the elute and incubated at room temperature for 5 minutes. After incubation the elute was filtered through Qia precipitator and 2 ml 70% ethanol was filtered through Qia precipitator respectively. Lastly, the Qia precipitator was placed on 5 ml syringe and 1 ml TE buffer was filtered through the precipitator twice to collect the purified plasmid DNA. Concentration of the purified plasmid DNA was measured via NanoDrop with ND1000 program.

#### D. Glycerol Stock

Bacterial glycerol stock is a method to preserve plasmids in bacteria for long term at -80°C. Thus, with this storage method one can quickly expand positive clones without having to transform again.

**Procedure:**

Prior to plasmid DNA purification (section 4.3.7, subsection C), 2.5 ml bacterial cultures were set aside. Then the bacterial culture was added to 835 µL of glycerol stock. The solution was mixed thoroughly by vertexing for 60 seconds, before transferring 1 ml of the mixture into cryotubes. The solution was snap-freeze with liquid nitrogen before storing at -80°C.

### 4.3.8 Sanger sequencing

Sanger sequencing, also known as chain termination method, is designed to determine the nucleotide sequences of a given DNA[99]. This method involves denaturation of the DNA template, so that sequencing primer can bind to the single DNA strand. Following primer binding, DNA polymerase starts synthesis of the DNA template in the 5′ → 3′direction in the

presence of four deoxynucleotide triphosphates (dNTPs: A G, C and T). Termination of synthesis reaction is used to determine the nucleotides of the DNA template in the presence of four dideoxynucleotide triphosphates (ddNTPs: ddATP, ddGTP, ddCTP, and ddTTP) labeled with fluorescent dye. The termination process occurs through removal of oxygen atom from the ddNTPs which inhibit the ddNTPs linking to the next nucleotide. Following synthesis, the chain terminated oligonucleotides are then separated via gel electrophoresis by size. Lastly, gel analysis and determination of DNA sequence is performed via laser excites the fluorescent tags in each band and a computer detects the light emitted by the laser[99], figure 4.10. In this study Big dye® Terminator v3.1 Cycle Sequencing kit was used, and PCR amplification was performed in our laboratory, while rest of the sanger sequencing steps were performed by the core facility at the Haukeland University hospital. The final data was analyzed by website tool application –SnapGene.

**Procedure:**

To verify the correct sequence of the cloned products or PCR fragments, Big dye® Terminator v3.1 Cycle Sequencing kit reaction was set up according to table 4.8. The thermal cycle program used for PCR amplification was according to table 4.9, before sending the amplified samples for further analysis to the core facility at the Haukeland University hospital.

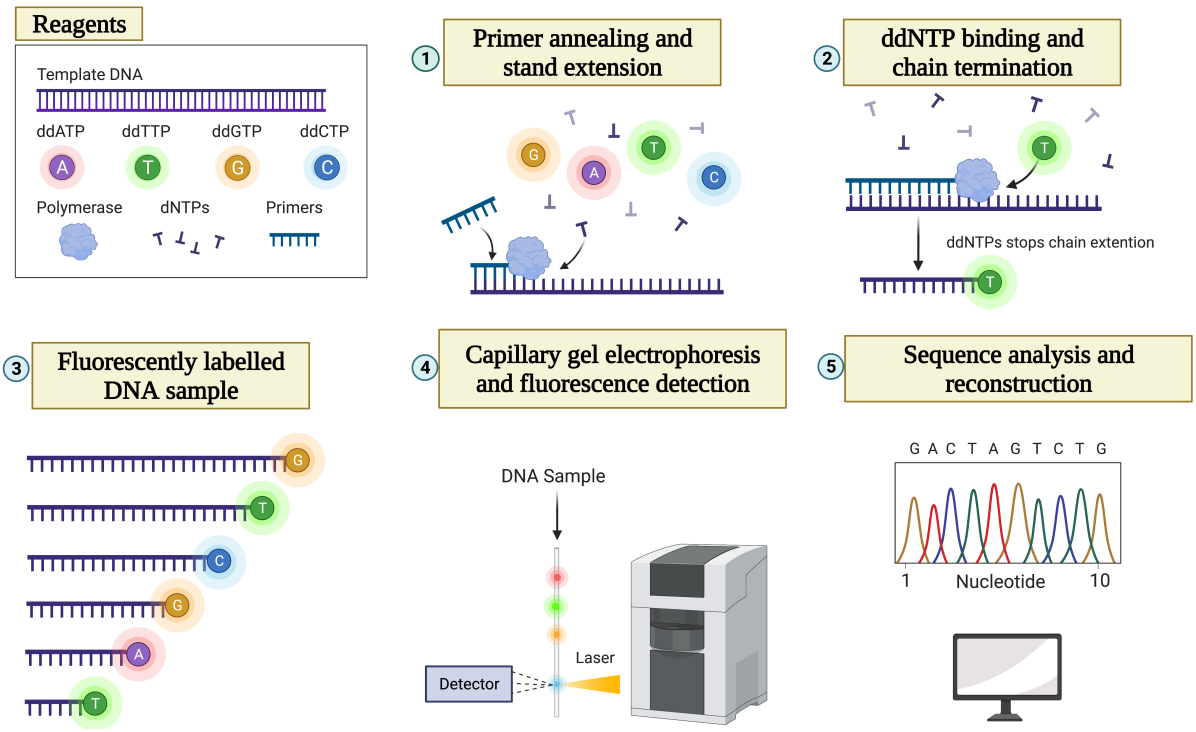


**Figure 4.10: Schematic representation of Sanger sequencing.** This figure was adapted from Biorender.com.

**Table 4.8: Big dye® Terminator v3.1 Cycle Sequencing kit reaction set up**.

| Components | Volume |
|---|---|
| Nuclease free water | To a final volume of 10 µL |
| Primer (2 µM) | 1,6 µL |
| Seq buffer 5X | 2 µL |
| Big Dye | 1 µL |
| Template DNA | X ng |
| **Total volume** | **10 µL** |

Amount of template DNA varied depending on the DNA type. For plasmid DNA 200 ng and for PCR fragments 15 ng.

**Table 4.9 The thermal cycling program used for Big dye® Terminator v3.1 Cycle Sequencing.**

| Set up | | | |
|---|---|---|---|
| | | **Temperature** | **Time** |
| Initial denaturation | | 98° C | 30 seconds |
| 35 cycles | Denaturation | 96° C | 10 seconds |
| | Anneal | 55° C | 5 seconds |
| | Extend | 60° C | 4 minutes |
| Final extension | | 4° C | Hold |

## 4.4 Cell culture methods

### 4.4.1 Cell line models

In this study the following three cells were used: COS-1, ME3 and 3T3-L1. The COS-1 and ME3 cell lines were used most extensively. An overview of the cell lines is provided in table 4.10 and further described below:

**Table 4.10 Overview of the cell lines used in this study.**

| Cell line | Cell type | Generated from | Animal | Dt (hours) |
|---|---|---|---|---|
| COS-1 | Kidney | Epithelial cells | Ape | 48 |
| ME3 | Adipose (beige/brown) | Embryonic fibroblast cells | Mouse | 24 |
| 3T3-L1 | Adipose (white) | Embryonic fibroblast cells | Mouse | 48 |

Dt stands for: the approximate doubling time of the cells in the medium.

**COS-1:**

The COS-1 cell line were generated by transformation of monkey epithelial cell line –CV-1 with a defective mutant of SV40 [100]. The transformed COS-1 cells express nuclear large T-antigen, moreover these cells contain all the proteins that are necessary for replication of circular genomes. There are several advantages with COS-1 cell line, including transfection of these cells are very easy and rapid, selection for transfectants is not required. Furthermore, from the cell line evenly expression of both wild-type and mutant forms of proteins can be expected if expressed protein is stable and nontoxic[100].

**ME3:**

ME3 cells are beige/brown pre-adipocytes mouse embryonic fibroblast $Rb^{-/-}$ line 3[101]. ME3 cells were generated by mating wild-type FVB mice with disrupted *RB-1* gene heterozygous FVB mice. Consequently, heterozygous $RB^{+/-}$ mice were generated and mated to create homozygous $RB^{-/-}$ embryos. ME3 cells were initiated at day 13 of gestation and the cells were cultured in DMEM with 10% newborn Calf Serum[102].

**3T3-L1:**

3T3-L1 cell line is a pre-adipose line and were established from Swiss mouse fibroblast line 3T3, based on their ability to accumulate lipid [103]. Preadipocytes 3T3-L1 cells can differentiate into mature adipocytes in the presence of classic hormonal cocktails, including insulin, synthetic glucocorticoids, and phosphodiesterase inhibitor 1-methyl-3-isobutyl xanthine [104]. 3T3-L1 cells have been frequently used to identify molecular markers, transcription factors and various interaction that are necessary for pre-adipose differentiation [105].

### 4.4.2 Aseptic technique in the cell laboratory

The aseptic technique is a mandatory method that is designed to use practices and procedures to prevent microbial contamination in cell cultures. This technique requires strict rules to minimize the risk of infection, simultaneously creating a barrier between microorganisms in the environment and the sterile cell culture[106].

**Procedure:**

All the experiments involving cell culture were performed in a mammalian cell laboratory on a laminar air flow bench. Before working with cell culture, laboratory coats and gloves were always worn. All the equipment used for cell culture and experiments were sterile. The

laminar air flow bench was sprayed and cleaned with 70% ethanol before placing cell cultures in the bench, moreover all the bottles and reagents were sprayed with 70% ethanol before placing inside the bench.

### 4.4.3 Cell culture

Cell culture is a technique where cells are collected from animal or plant and maintained *in vitro* under suitable artificial environment. Cell culture required strict maintenance of the controlled conditions that mimic the cell's natural condition and must therefore through a growth medium be supplied with essential nutrients such as amino acids, carbohydrates, vitamins, minerals, growth factors and hormones[107]. To avoid contamination and infection of the cells both handling and storage of cells are performed under sterile environment (section 4.4.2).

**Procedure:**

In this study, all the mammalian cell cultures were mainly cultivated in either Dulbecco's Modified Eagle Medium (DMEM) or Grand Island Biological Company (Gibco™) medium. The medium was supplemented with additional nutrients including 10% fetal bovine serum (FBS), 10% newborn calf serum (CS) and AmnioMAX™ –C100 supplement. In addition, to protect the medium from bacterial infection a mixture of penicillin and streptomycin (PEST, 50 000 U penicillin and 50 mg streptomycin) was added. Table 4.11 represents components needed to make a complete medium that was used to culture different cell types. Cultivated cells were grown in cell culture petri dishes and they were incubated at 37° C with 5% $CO_2$ in a humidified condition between handlings. To prevent overgrowth of the cells, the cells were looked at 2-3 times a week and sub-cultured 2-3 times a week (section 4.4.4).

**Table 4.11 cell lines and the growth media used to culture them.**

| Cell lines | Basal medium | Additives |
|---|---|---|
| COS-1 | DMEM, with 4.5 g/L glucose and L-Glutamine. | 10% FBS, 1% PEST (50 000 U penicillin and 50 mg streptomycin) |
| ME3 | Gibco, AminoMAX™ -C100 (1X) Basal medium. | 7.5% AminoMAX™ –C100, supplement, 7.5% FBS, 1% PEST (50 000 U penicillin and 50 mg streptomycin), 1% L-glutamine. |
| 3T3-L1 | DMEM, with 4.5 g/L glucose and L-Glutamine | 10% CS, 1% PEST (50 000 U penicillin and 50 mg streptomycin) |

### 4.4.4 Cell subculturing

Subculture or passaging is a method that involves removal of medium and transfer of cultivated cells from old/previous culture into new and fresh growth medium [107]. This method enables the cell lines to divide continuously and prevents the cells from reaching confluence under controlled conditions. In general, subculture is performed when the cultivate cells are in the log phase (70-80% confluency), Figure 4.11. Furthermore, this method involved enzymatical detachment of adherent cells from the cultured dish. Trypsin-EDTA is used for detachment process and widely used due to its digestive strength. Prior to treatment with trypsin-EDTA, the cells mut be washed with Phosphate Buffer Saline (PBS). PBS is a balanced salt and removes chelators from the culture. Lastly, the trypsin-EDTA activity is inhibited by FBS which contains a1-antitrypisn[107].

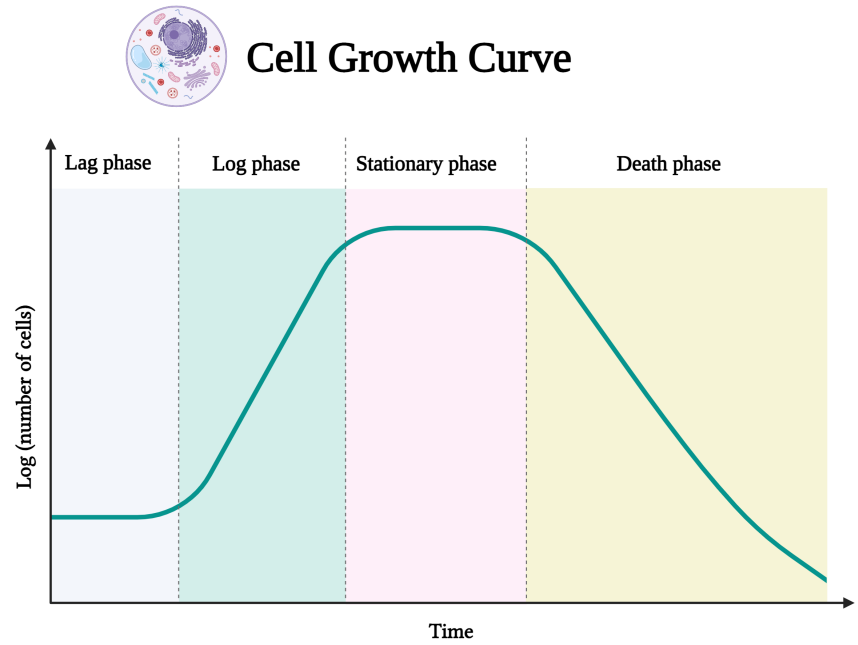



**Figure 4.11 Growth curve of cultured cells.** In general, cultured cells should be passaged within the log phase growth, which is also known as exponential growth and refers to 70-80% confluency. The figure was adapted from Biorender.com.

**Procedure:**

During the whole passaging process, aseptic technique (section 4.4.2) was performed under the laminar air flow bench. Cell culture medium was removed and discarded prior to washing with pre-warmed PBS at 37° C in heat box. PBS solution was removed and discarded before adding pre-warmed Trypsin-EDTA at 37° C heat box. Following addition of Trypsin-EDTA, the cell dishes were incubated for 2 minutes at 37° C in the humidified incubator. After incubation the cell dishes were observed under the microscope for detachment. When all the cells were detached, pre-warmed medium was added to inhibit trypsinization before

centrifuge at 300 rpm for 3 minutes. Supernatant was discarded and pellet was resuspended with 1 mL pre-warmed medium. Then according to the passaging factor, a fraction of cell suspension was transferred to new cell dishes containing fresh medium and the dishes were cultures at 37° C with 5% $CO_2$ in a humidified incubator. Cell lines used in this study were passaged according to table 4.12.

**Table 4.12 Overview of sub-cultivation requirements**.

| Cell types | Passaging factor | PBS | Trypsin-EDTA | Medium* | Cell suspension | Medium (without cell suspension) |
|---|---|---|---|---|---|---|
| COS-1 | 1:15 | 5 ml | 1 ml | 3 ml | 1 ml | 10 ml |
| ME3 | 1:10 | 5 ml | 1 ml | 3 ml | 1 ml | 10 ml |
| 3T3-L1 | 1:10 | 5 ml | 1 ml | 3 ml | 1 ml | 10 ml |

Medium* represents the amount of medium used to inhibit trypsinization. Cell suspension is the amount of medium used after centrifuge.

### 4.4.5 Seeding of cells

Different experiments require a specific cell concentration and to achieve this cell in cell suspension are counted and diluted appropriately to attain the desired cell concentration for seeding. Cell counting is done by trypan blue, which is a dye that stains the dead cells. As a result, trypan blue makes sure that live cells and dead cells are counted separately in the Countess™ II Automated Cell Counter [108]. In this study, to obtain the desired cell numbers to seed for different experiments equation 4.1 was used.

**Equation 4.1:** formula used for cell seeding:

$$\frac{Desired\ number\ of\ \frac{cells}{dish}}{Counted\ cell_{mL}^{cells}} = mL\ of\ cell\ solution\ to\ use$$

**Procedure:**

Cells were collected from cell suspension (cell subculturing, section 4.4.4), and the sample were prepared by adding 20 µL of cell suspension to 20 µL of 0.4% trypan blue stain. The solution was mixed by pipetting up and down. 10 µL of solution was loaded into the sample loading area of the Countess™ II Automated Cell counting chamber slide and inserted the slide into the slide port of the Countess™ II Automated Cell Counter. Then the concentration and percentage of viable cells in the solution was noted down and equation 4.1 was used to appropriately dilute the cell suspension to get the desired concentration.

### 4.4.6 Thawing of cell cultures

Cryopreservation is a process where cultured cells can be stored and persevered for long-term storage in liquid nitrogen tank or freezer. Thawing of cryopreserved cells is a stressful procedure to frozen cells, thus quick performance is necessary to ensure high viability, recovery, and functionality of the frozen cells for downstream applications [109].

**Procedure:**

Appropriate 10 mL medium was pre-warmed at 37° C in heat box and the cryopreserved cells were taken out of the nitrogen tank and placed on ice immediately. The cryotube was placed in heat box at 37° C for 2-3 minutes. After incubation some warm medium was added to the tube before placing back the tube at heat box for 1-2 minutes. Then thawed cell suspension was gently transferred to pre-warmed 10 mL medium and warmed medium was pipetted into the tube for further thawing. Once thawed completely, all the cell suspension was transferred to warm 10 mL medium. The cell suspension then centrifuges at 300 rpm for 3 minutes to remove the DMSO. The supernatant was discarded and resuspended with 1 mL medium. The whole cell suspension was transferred to petri dish containing fresh medium and incubated at 37° C with 5% $CO_2$ in a humidified incubator.

## 4.5 Luciferase assay

Luciferase is an oxidative enzyme, which exists in several species, and the enzyme is used for bioluminescence by the organisms. In other words, these enzymes enable the species to emit light and the most famous of these enzymes is the firefly luciferase [110]. These enzymes are extremely sensitive thus making it a great candidature for reporter assay. Luciferase assay is a reporter assay method that is used to determine if a protein can activate or repress the expression of a target gene with the help of luciferase that is used as a reporter protein. Thus, this method establishes a functional connection between presence of a protein and the amount of gene product. In order to perform luciferase assay, a construct is needed with regulatory region or promoter region fused with coding sequence for luciferase. Then the DNA sequence of interest is cloned upstream of the luciferase gene fused in the expression vector. The recombinant construct is transfected into a cellular system. Once the promoter of the recombinant construct is activated that would lead to transcription of luciferase reporter gene and production of luciferase protein. With appropriate cell lysis buffer, the transfected cells are lysed and releases all the protein including luciferase. The cell lysate is transferred to

multi-well plate and substrate containing luciferin and other necessary co-factors are added to the cell lysate. Addition of substrate catalyzes the reaction, and the light is emitted. The emitted light is measured by a luminometer, and the luciferase activity is determined by the amount of bioluminescent signal detected by the laminator [110]. In this study, pGL4.23-*Luc-2*/minP vector was used as expression vector with desire DNA sequence, and cell lines used including COS-1, ME3 and 3T3-L1.

**Procedure:**

Cells were sub-cultured (section 4.4.4) and seeded (section 4.4.5) 7500 cell/well in 96-well tissue cultured plates. These cells were then incubated at 37° C with 5% $CO_2$ in a humidified incubator overnight. On the second day either0.04 µg/well or 0.08 µg/well of each reporter plasmid pGL4.23 containing the desired DNA sequence (in-fusion cloning, section 4.3.5 and site-directed mutagenesis, section 4.3.6) or overexpressed plasmid was added to PCR tubes. The plasmid mixes were diluted with OptiMEM followed by addition of either TransIT-LT1 (COS-1 and ME3 cells) or TransIT-LX2 (3T3L-1 cells) transfection reagents. The reaction mix was then incubated for 30 minutes at room temperature. In the meantime, the old medium of the seeded cells was removed and 90 µL of pre-warmed fresh medium was added to each well, before adding 10 µL of transfection mix to each well. Transfected cells were incubated at 37° C with 5% $CO_2$ in a humidified incubator. On the fourth day the lysis buffer was made according to table 4.13 and placed on ice. The seeded cells were washed with 100 µL cold PBS and observed under microscope for detection of detachment. After washing, 30 µL of cold lysis buffer was added to each well and was placed on horizontal shaker machine for 20 minutes at 4°C at 375 rpm. 20 µL of cell lysate was transferred to a white opaque 96 well plate. Luciferase activity was assessed according to the BioThema Luciferase assay Kit A protocol on the FLUOstar OPTIMA instrument. This instrument used automated pipetting of 50 µL luciferin and 50 µL ATP per well. Then the data was then exported for further analysis by Microsoft excel and GraphPad (table3.2).

**Table 4.13: Lysis buffer set up.**

| Components | Set up | Volume (1 ml) |
|---|---|---|
| | Nuclease free water | 819 µL |
| 85% Glycerol | 10% Glycerol | 117 µL |
| 0,5 M Tris Ac, 20 mM EDTA | 25 mM TAE, pH 7,8 | 50 µL |
| Triton $X_{100}$ | 1% Triton | 10 µL |
| 0,5 M EDTA | 1 mM EDTA | 2 µL |

| | | |
|---|---|---|
| 1M DTT | 2 mM DTT | 2 µL |

This table only represents the amount of components needed for 1 ml, however depending on the experiments and number of wells this volumes were multiplied by the desired number or ml needed for the experiments.

## 4.6 Statistical analysis

All the statistical analyses and graphing of luciferase results in this study was performed in the GraphPad Prism version 9.2.0 software. Raw data were normalized to the average of a control treatment (most often the empty reporter plasmid) in excel prior to analysis in Graphpad. To verify whether our data was normal distributed, Shapiro-Wilk (W) test was performed for normality ($P< 0.05$). Moreover, Bartlett´s test was performed to check whether the standard deviations were significantly different ($P < 0.05$). One Way ANOVA with Holm- Šídák post hoc test was performed for multiple comparison of cloned tiles and relative to pGL4.23-*luc2*/minP_empty vector. As for data sets with two factors, two-way ANOVA with Holm- Šídák post hoc test were used, and all results are presented as mean ± standard deviation (SD). Differences between means were considered significant at $p<0.033$ (*), $p<0.002$ (**), and $p<0.001$ (***).

**Equation 4.2:** Normalize value

$$\frac{Measured\ \ value\ to\ be\ normalized}{Value\ of\ the\ reference}$$

# 5. Results

## 5.1 Design and characterization of tiles in the 11q23.3 locus

The visceral obesity-associated 11q23.3 locus was divided into multiple regions/tiles.[24,88] Overall, 14 new tiles were successfully designed, consisting of 2-3 primer pairs each for 5 different areas of the locus (Figure 5.1, table 3.3 and supplementary data figure 10.4) to get a deeper understanding of the 11q23.3 locus, publicly available data from the UCSC genome browser and the WashU Epigenomic Browser was retrieved and visualized alongside the tiles (Figure 5.1). The locus contains three genes including *HYOU1*, *VPS11* and *HMBS* gene. As the figure 5.1 shows collected information from various data sources indicating different regions of the LD block having both transcription and enhancer activity. GH Reg Elems (DE) data indicates promoter activity, where dark red represents high promoter activity and light dark represents less promoter activity. According to the data *VPS11* and *HMBS* gene start sites have more promoter activity compared to start site of the *HYOU1* gene. However, we were more interested in enhancer activity, thus we retrieved epigenomic data specifically for adipose tissue. As the data from WashU Epigenomic Browser shows different level of enhancer activity depending on types of adipocytes, for example all three adipose tracks showed enhancer activity in tile 2/3 junction, but in tile 8, only the adipose derived mesenchymal stem cell cultured cells (E025) showed a predicted enhancer, while the mesenchymal stem cell derived adipocyte cultured cells (E023) and adipose nuclei (E063) did not. Tile 1 covered the most upstream region of the LD block with very low predicted activity and harbors very few SNPs. In contrast, tile 3 covers a region with both promoter and enhancer activity with many SNPs and VAT associated *VPS11* gene start site is located in tile 3. Even though, tile 4 and 5 harbors more SNPs compared to tile 1, both WashU Epigenomic Browser and UCSC Genome Browser data shows very little promoter or enhancer activity in these regions. These claims are further verified by ENCODE cCRE and ORegAnno data, that shows level of enhancer activity (orange/yellow bars indicates high enhancer activity). Layered H3K27Ac track also represents enhancer activity by blue bar peaks and as the figure shows within tile 4 and 5 there are almost no enhancer activity. Moreover, tile 8 covers various transcription start site of *HMBS* gene (isoforms) and adipose derived mesenchymal stem cell cultured cells (E025) data shows two of the short isoforms have very high enhancer activity. The ENCODE cCRE and ORegAnno data also showing high enhancer activity in the *HMBS* region and GH Reg Elems (DE) data indicating less promoter activity within the

*HMBS* region covered by tile 8, thus further indicating more enhancer activity. And lastly, Layered H3K27Ac track also shows very high enhancer activity in the *HMBS* locus, which is annotated by the blue bar peaks.

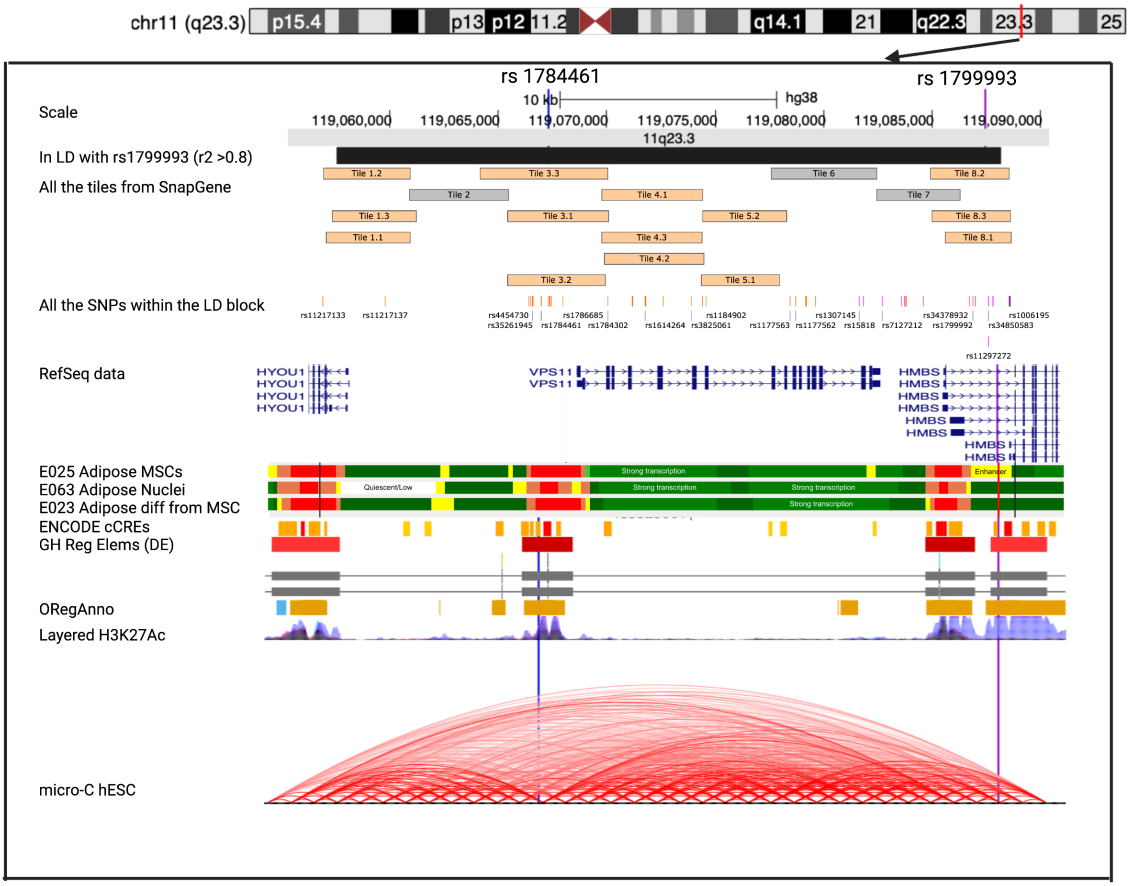



**Figure 5.1 Design and characterizion of tiles in the 11q23.3 locus.** The upper navigation bar represents the position within chromosome 11 for the expanded visualization of *HYOU1, VPS11* and *HMBS* gene locus below. The box below shows an enlarged version of the locus, and the yellow bar represents the LD block with SNPS (r2≥ 0.8). Numbers above the bar represent genomic position. The "Tiles from SnapGene" parameter shows the position of previously cloned tiles (gray), as well as the newly designed tiles. The next parameter shows the position of all the SNPs in the LD block, including tag-SNP rs1784661 (pink bar) and rs1799993 (purple bar). The next track shows RefSeq genes and their isoforms. The thick blue bars are exons (protein coding sequence), and the thin lines are introns (non-protein coding region). The next 3 tracks are data from WashU Epigenome Browser, which is an integration and analysis tools for epigenomic datasets and includes the Roadmap Epigenomics and the ENCODE projects. E023, E025 and E063 are chromatin annotations from three different adipose specific tracks including mesenchymal stem cell derived adipocyte cultured cells (E023), adipose derived mesenchymal stem cells (E025) and adipose nuclei (E063). Within each track –orange/red color represents active transcription start sites, green regions indicate strong transcription, dark green shows weak transcription and lastly yellow regions represents enhancer activity. The next three parameters dsiplay merged regulatory activity across several cell types: ENCODE registry of candidate cis-regulatory elements (cCREs) in the human genome, GeneHancer data (regulatory elements and promoter activity) and OregAnno that also represents regulatory regions and transcription binding sites. Finally, layered H3K27Ac track characterizes enhancers activity. The last panel shows micro-C hESC data that represents chromatin-chromatin interactions.

### 5.1.1 PCR Amplification of tiles from genomic DNA

After having designed the tiles *in silico*, we next aimed to amplify the corresponding DNA fragments by PCR for in-fusion cloning. Importantly, each tile needed to be cloned from gDNA of at least 2 individual donors; one with homozygous risk and one homozygous protective haplotype. However, due to limited availability of gDNA from homozygous patients, primer pair efficiency and specificity were initially screened on a pool of heterozygous patients as described in the following sections. During screening, the PCR conditions were optimized for each primer pair (table 4.2).

### A. Initial screening of primer pairs and optimization of PCR

To increase the chances of successful cloning, 3 primer pairs were designed for each of the 5 tiles. In order to test the specificity and efficiency of primers for each tile, we first used a pool of gDNA derived from multiple patients (table 3.6) The gene pool was created according to table 3.6 by mixing all the different gDNA together and the concentration of the mixture was retrieved from NanoDrop, prior to performing PCR, as the figure 5.2 A shows.

In this experiment tile 1.3 (section 5.1.2, B. genotype dependent PCR amplification) was used as control, since initially primers for tile 1 was tested on patient samples presented in table 3.7 (homozygote risk and protective, supplementary data: figure 10.1). Due to limited amount of patient sample, we decided to test efficiency of other primers on the gene pool samples, prior to test the final primers on the desired patient samples (table 3.7). As mentioned earlier, PCR program used for the test experiment (figure 5.2 A) was according to table 4.2 with modification on annealing temperature, 10 cycles annealing at 60º C and 25 cycles at 65º C. PCR amplified products were then investigated for the correct DNA fragment size according to table 3.6 on agarose gel.

Figure 5.2 A) shows the tiles with approximately right DNA fragment size on the gel, including tile 3.1 (lane 2, 4761 bp), tile 4.2 (lane 6, 4714 bp), tile 5.1(lane 8, 3696 bp), and 8.2 (lane 11, 3729 bp) for gene pool of heterozygote patients. Positive control tile 1.3 did not show any visible band. Furthermore, all the alternative primers for tile 3 had approximately right sizes but compared to tile 3.1, tile 3.2 (lane 3, 4604 bp) had a band more towards 4000 bp and even though tile 3.3 (lane 4, 6025 bp) had the correct size, amplicon size was too big compared to other amplicons. As for tile 4, there are no right sized visible band for tile 4.1 (lane 5, 4762 bp) and tile 4.3 (lane 7, 4732 bp). DNA fragment of tile 5.2 (lane 9, 3990 bp), and tile 8.3 (lane 12, 3726 bp) had no visible band, and the amplified products migrated

through the gel in a smeared manner. Tile 8.1 (lane 10, 3096 bp) had a visible band round 3000 bp, which is short compared to other amplicons.

Outcome of the PCR amplification from the previous experiment wasn't optimal, thus we decided to perform second gene pool experiment with few modifications. The final gene pool experiment (figure 5.2 B) was carried out similarly to the first experiment, except for modification on PCR program. For the final gene pool experiment, two sets of PCR thermal programs were used in parallel for each tile, with 35 cycles of annealing at either 60º C or 65º C. The PCR amplified products were then analyzed and verified by agarose gel electrophoresis based on the fragment band sizes.

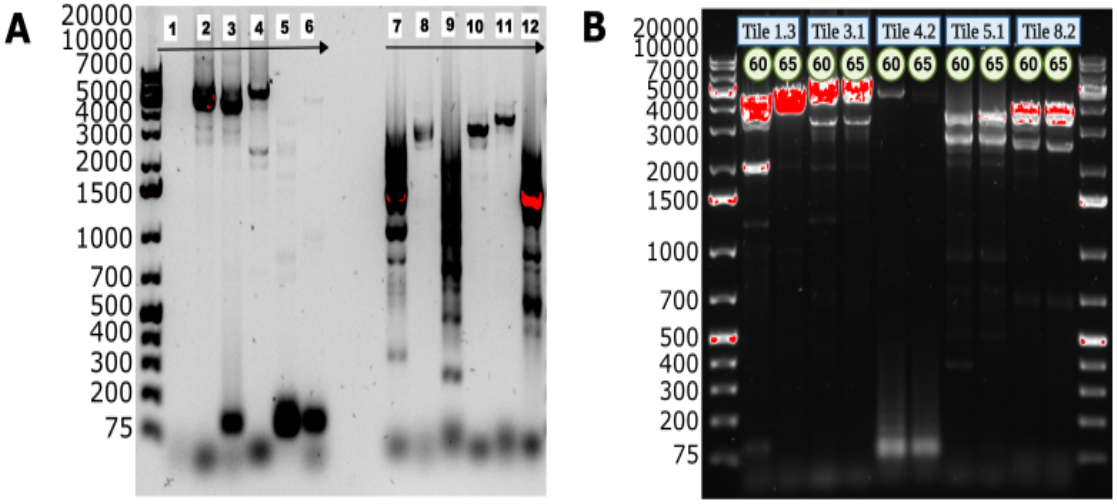



**Figure 5.2 Gel electrophoresis picture of amplified constructs on 1,2% agarose gel**. **A)** represents amplified tiles tested on gene pool and numbered from 1-12. Numbered lanes in the figure from 1-12 shows tile 1.3 (1), tile 3.1 (2), tile 3.2 (3), tile 3.3 (4), tile 4.1 (5), tile 4.2 (6), tile 4.3 (7), tile 5.1 (8), tile 5.2 (9), tile 8.1 (10), tile 8.2 (11), and tile 8.3 (12) respectively. **B)** shows gel electrophoresis result of tile 1.3, tile 3.1, tile 4.2, tile 5.1, and tile 8.2 respectively at annealing temperature at 60º C and 65º C.

In the second experiment as figure 5.2 B) shows, both tile 1.3 (positive control) and tile 3.1 had right size visible bands round 4000 and 5000 bp respectively both at 60º C and 65º C. But at 60º C tile 1.3 had a second band round 2000 bp and tile 3.1 had very weak bands round 3000 bp at both temperature conditions. Thus, for both of these tiles, 65ºC was found to be better than 60ºC. For Tile 4.2, like in the previous amplification, only a very weak band was visible at around 5000 bp at 65ºC. Reducing the temperature to 60ºC increased the band intensity. Both tile 5.1 and tile 8.2 had right size visible bands at both temperatures, but both tiles also had a second band around 3000 bp and 2500 bp respectively. Thus, results from

these experiments suggested that the best results were obtained with primers for tile 1.3, tile 3.1, tile 8.2 and possibly 5.1 using 35 cycles at 65°C. For tile 4, the 4.2 version performed best, but still produced a band too faint for further purification. Notably, despite producing strong bands of correct sizes, the primer pairs for tiles 3, 5 and 8 also resulted in an additional weak unspecific band. However, because these unspecific bands were clearly distinct from the correct band, gel excision of the correct band would be possible. Therefore, further optimization of PCR conditions was not performed.

### B. Genotype-dependent PCR amplification from homozygous patients

Based on the results from the previous section, we decided to perform genotype-dependent PCR amplification of tiles 1.3, 3.1 and 8.2 from homozygous risk and protective patients to obtain amplicons for cloning (Table 3.7). As the figure 5.3 shows each primer was tested on 15 ng of gDNA and PCR program was set according to table 4.2 with 35 cycles of annealing at 65° C. Gel electrophoresis was performed for further verification and analyzation of PCR amplified products. As the figure 5.3 A) shows, tile 1.3 had very strong and correct band size for both the risk and protective haplotype. Similarly, tile 3.1 also had correct band size for all four samples, even though a very weak band was present around 3500 bp.

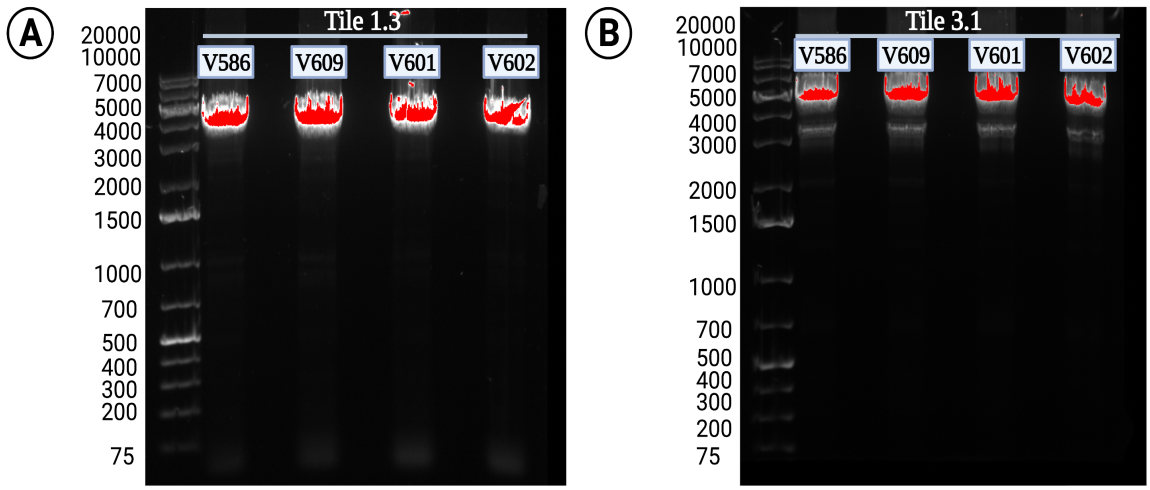



**Figure 5.3 Genotype-dependent amplification of tile 1 and tile 3. A)** shows PCR amplified tile 1.3 products both on homozygote risk and protective patient, furthermore, all the DNA fragments had the correct band size around 4000bp. **B)** represents PCR amplified products of tile 3.1 on homozygote risk and protective patient gDNA. All the fragments had right band size around 5000 bp, even though all the samples had very weak band around 3500 bp.

As for tile 8 same procedure was performed as tile 1 and tile 3, and as the figure 5.4 shows 1,2% agarose gel image presenting amplified DNA fragments on both homozygote risk and

protective type at 65º C. Amplified DNA fragments of protective patient gDNA had right band size around 4000 bp, in contrast the risk patient gDNA had very weak band round 4000 bp.

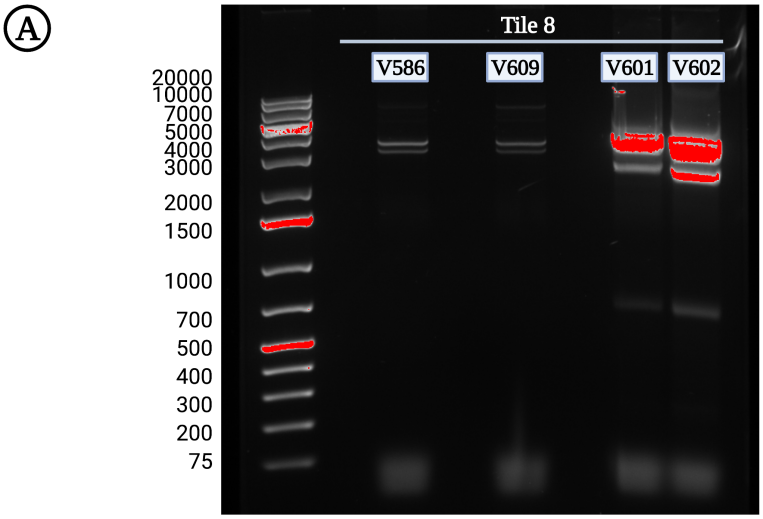



**Figure 5.4 Gel electrophoresis image of amplified DNA fragment of tile 8**. The ladder on the left side represents molecular weight/GeneRuler 1 kb plus DNA ladder. The first two lane represents amplified homozygote risk patient gDNA and the other two lane on the right side represents amplified homozygote protective patient gDNA.

Based on the result, we decided to optimize PCR conditions specifically for tile 8.2 with risk genotype in order to obtain a stronger band, (figure 5.5 A). Each primer was tested on 15 ng of the two-homozygote risk patient gDNA with 35 cycles annealing at either at 55º C or 60º C. At 55º C, amplified gDNA products of both patients had stronger and correct band size around 4000 bp. In contrast, at 60º C amplified DNA fragments had correct size but very weak band as seen before. After verification, we performed one more PCR on risk patient gDNA with 35 cycles annealing at 55º C and verified the result by gel electrophoresis, (figure 5.5 B).

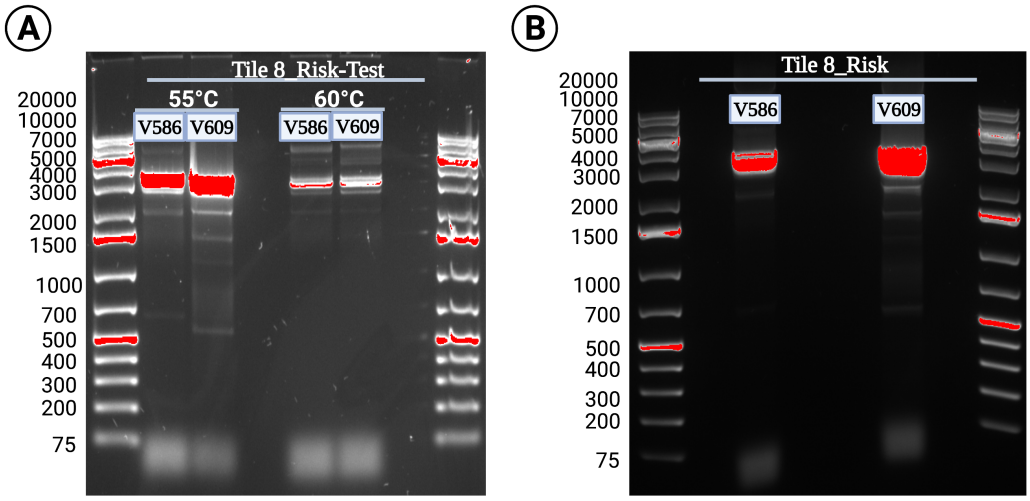


**Figure 5.5 Visualization of PCR amplification of tile 8 on risk gDNA.** A) represents visible band size of V586 and V609 risk gDNA on tile 8 both at 55º C and 60º C. B) 1,2% agarose gel image representing strong band around 4000 bp of the two amplified risk gDNA.

### C. Purification of amplified products

Following PCR and verification by gel electrophoresis, the correct band of each tile on both homozygote risk and protective patient gDNA was excised and purified by a DNA purification kit (section 4.3.3 DNA purification). Most of the genotype-specific PCR amplified products had good concentrations and yields after purification, except for tile 8 from V586 homozygote risk and V601 homozygote protective gDNA which had lower values, table 5.1.

**Table 5.1 Overview of genotype-depended purified amplified products.**

| | | Tile 1 | | Tile 3 | | Tile 8 | |
|---|---|---|---|---|---|---|---|
| **Patient code** | **Genotype** | **ng/$\mu L$** | **Yield** | **ng/$\mu L$** | **Yield** | **ng/$\mu L$** | **Yield** |
| V586 | Risk | 24,2 | 484 | 16,7 | 334 | 7,7 | 231 |
| V609 | Risk | 24 | 480 | 18,2 | 364 | 16,4 | 492 |
| V601 | Protective | 27,3 | 546 | 23,5 | 470 | 7,2 | 216 |
| V602 | Protective | 29,2 | 584 | 24,5 | 490 | 10,4 | 312 |

Yield was calculated by ng/µL multiplied by the elusion volume from the DNA purification kit.

### D. Verification of amplified products via Sanger sequencing

After DNA purification of the PCR amplified products, sanger sequencing was performed (as described in section 4.3.8) to ensure the correct bands had been excised. Figure 5.6 A shows the sequencing result for tile 1 from the two homozygote protective patients. The sequencing result was aligned to the original genomic sequence retrieved from the UCSC genome browser. The right gDNA fragment had been amplified, suggesting that genotype-dependent PCR amplification from protective patients was successful. Figure 5.6 B shows the sequencing result of tile 3 from both homozygote risk and protective patient gDNA. As the alignment with the genomic sequence revealed that the PCR amplified product was the correct sequence. In this study, sanger sequencing on PCR amplified fragment was only performed on tile 1 (protective types) and tile 3 (both protective and risk types), due to troubles with in-fusion cloning (section 4.3.5) which is discussed later (section 6.2.3).

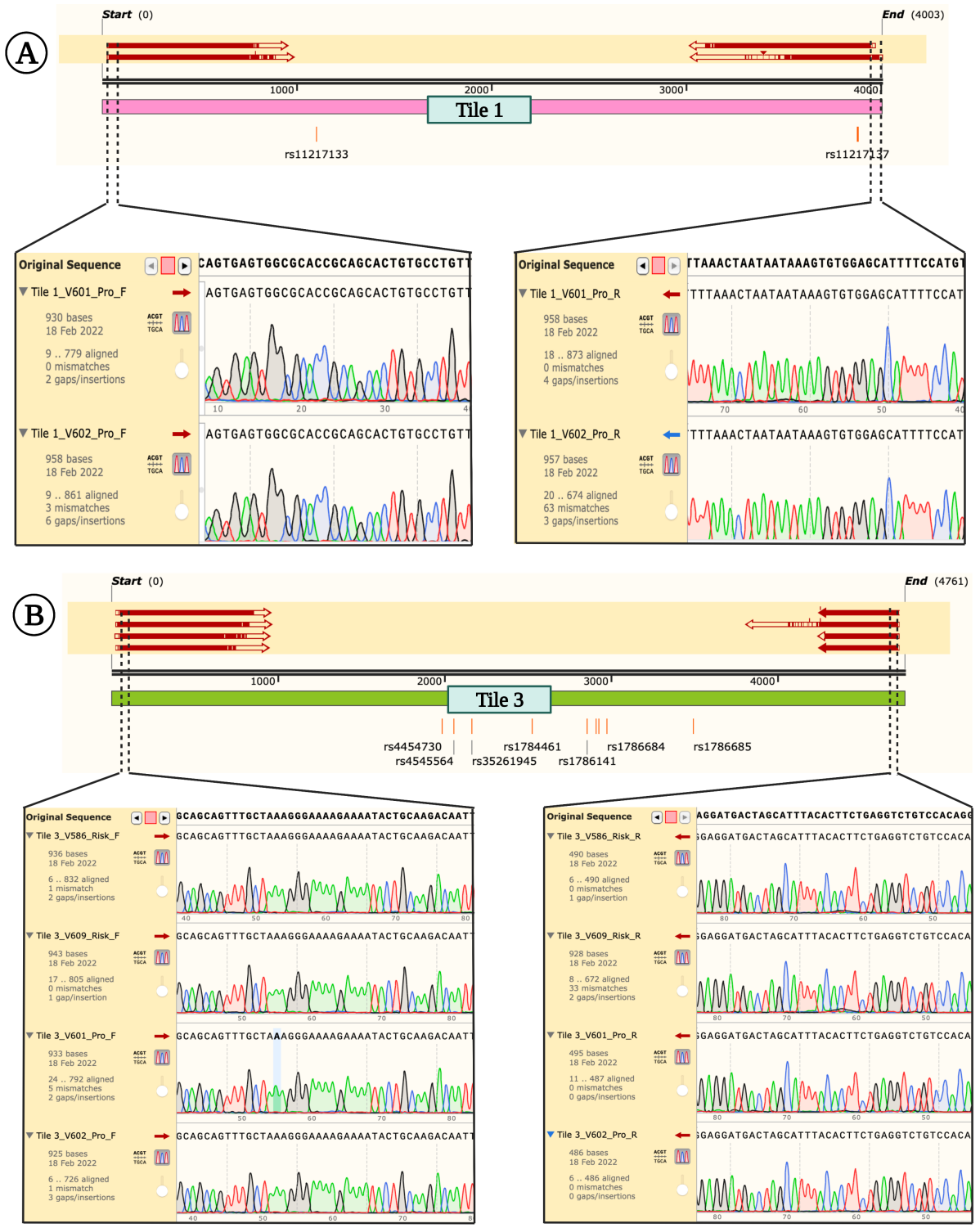


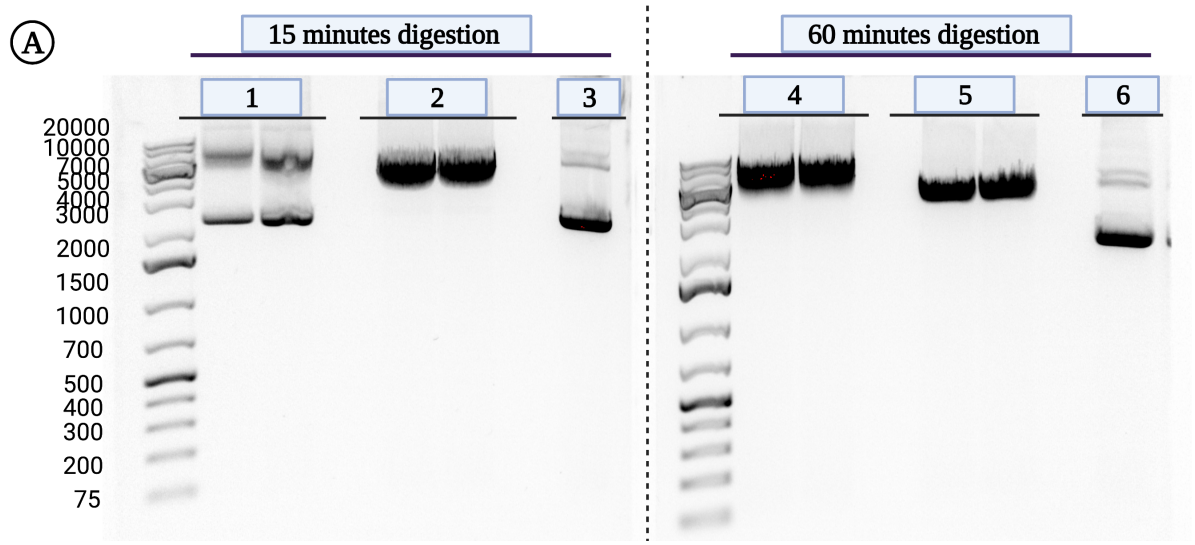

**Figure 5.6 Sanger sequencing result of tile 1 and tile 3. A)** The red bars above the navigation scale represent the coverage of the sequencing data, the electropherogram below represents enlarged version of the original sequenced data. Each peak represents a single nucleotide, and the four bases have different colors. The electropherogram on the left shows sequences from 5'→ 3' (forward primers) direction of tile 1 of protective patients (V601 and V602). The electropherogram on the right side shows sequences from 3'→ 5' direction (reverse primers). **B)** Sequencing data of tile 3 from both homozygote risk (V586 and V609) and protective (V601 and V602) patients.

## 5.2 In-fusion cloning of tiles into the pGL4.23-luc2-minP plasmid

In order to compare the enhancer activity of the different tiles within the 11q23.3 locus by luciferase assay the tiles needed to first be cloned into the pGL4.23-*luc2*-minP-empty luciferase vector. In this section, all the prior steps leading to measurement of luciferase activity are presented and discussed.

### 5.2.1 Restriction enzyme verification

To obtain the optimal digestion of the pGL4.23-*luc2*-minP-empty vector, single-enzyme digestion was first tested with either of the two compatible restriction enzymes HindIII-HF or KpnI-HF. Additionally, digestion was performed for both 15 minutes or 60 minutes at room temperature (figure 5.7A). While both enzymes cut the vector after 60 minutes of digestion, only KpnI-HF fully linearized the vector, and did so after only 15 minutes of digestion (Figure 5.7A). In contrast, HindIII-HF produced two strong bands after 15 minutes of digestion –one at the same position as the uncut, supercoiled position of the negative control (ca 2.5 Kb), and one at either the linearized or relaxed/nicked position (ca 5kb/7kb), which means that the plasmid wasn't digested completely by the enzyme after 15 minutes digestion with restriction enzyme. Moreover, even after 60 minutes, the band from HindIII-HF migrated slower than that of KpnI-HF, further suggesting that HindIII-HF only nicked one strand instead of fully cutting both strands. Thus, based on the result, we decided to utilize KpnI-HF to linearize the vector for cloning. A second scaled-up digestion was then performed using KpnI-HF for 60 minutes (figure 5.7B)



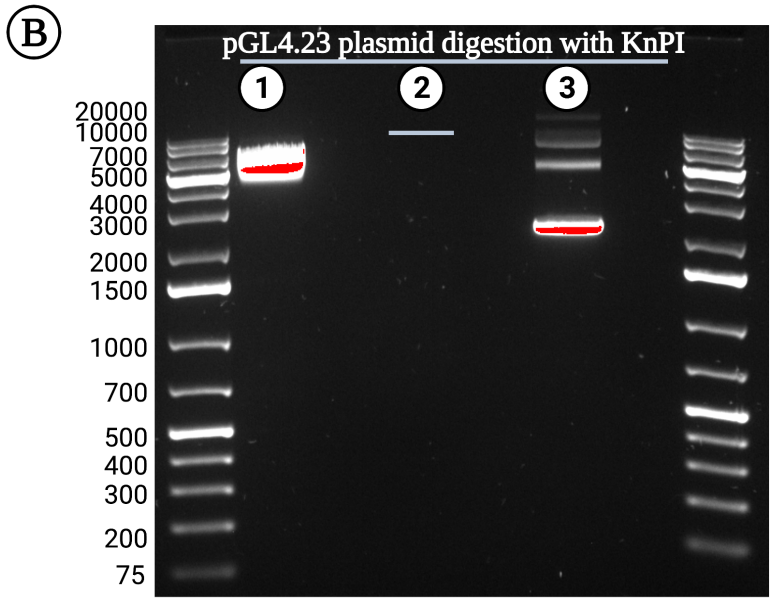



**Figure 5.7 Digestion of pGL4.23-*luc2*-minP-empty vector with HindIII-HF or KpnI-HF restriction enzymes. A)** Pilot digestion. The vector was digested for either 15 minutes (left panel) or 60 minutes (right panel) with either HindIII-HF, KpnI-HF or no enzyme. Our plasmid has a size of 4283 bp, and the last lane on the right side of both gel shows the 3 typical bands for a non-digested plasmid (no restriction enzyme); the top band represents nicked (single strand break), the middle band represents digested plasmid band (linearized), and the last band represents supercoiled, uncut plasmid. **B)** Full-scale digestion. The vector was digested for 60 minutes using KpnI-HF. Sample numbered as 1 is the digested/ linearized plasmid, Sample 2 represents no DNA negative control (only the enzyme) and the third lane shows the no enzyme/non-digested plasmid migration control. The first lane on the left side of each gel represents the molecular weight/DNA ladder.

### 5.2.2 Optimization of competent cells

Next, we optimized transduction of competent cells, by testing different amounts of uncut, positive control in-fusion plasmid DNA in different types of competent bacterial cells. Three different types of component cells, including XL 10-Gold cells, XL 1-Blue cells and TOP-10 cells were tested for this experiment. Either 1.0 µL or 0.5 µL of 50 ng/µL plasmid DNA was added to 25 µL of competent cells (section 4.3.7) ,as the figure 5.8 shows.

Figure 5.8 A) shows that XL 10-Gold competent cells formed very large, unevenly shaped and dense colonies, with both 0.5 µL and 1.0 µL plasmid DNA. In contrast, XL 1-Blue cells had less dense colonies, rather the colonies were distributed in an equally spaced manner with a round morphology (figure 5.8 B). Furthermore, for the XL 1-Blue competent cells, less DNA (0.5 µL) resulted in more colonies than more DNA (1.0 µL). Lastly, the TOP-10 competent cells had very poor colony growth compared to the other competent cells (figure 5.8 C). Overall, the result suggested that XL 1-Blue cells had the best outcome, specifically 0.5 µL plasmid DNA with 25 µL cell suspension.

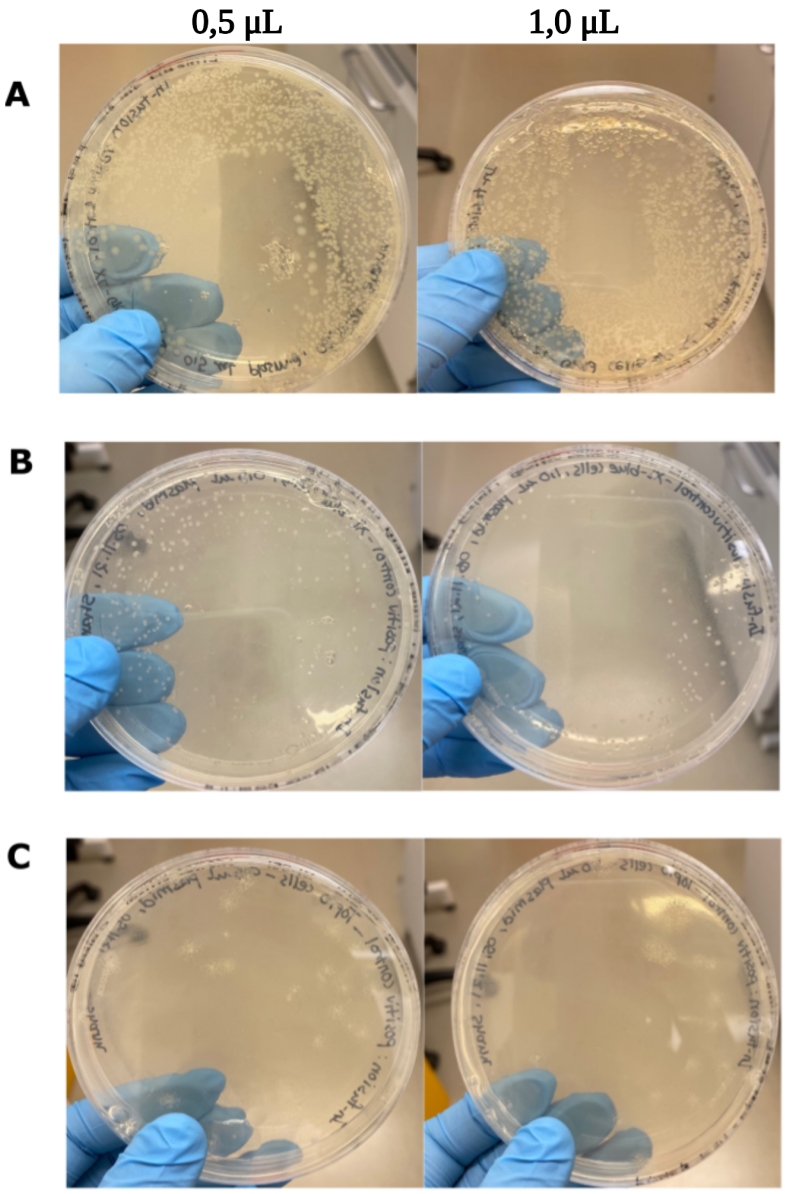



**Figure 5.8. Optimization of competent cells. A)** represent the colony formation image of 0.5 µL plasmid DNA within 25 µL XL 10-Gold competent cells on the left and on the right side 1.0 µL of plasmid DNA and 25 µL of competent cells. **B)** XL 1-Blue cells, on the left side 0.5 µL plasmid DNA and left side 1.0 µL with 25 µL competent cells. **C)** Colony formation of TOP-10 competent cells, on the left side 0.5 µL plasmid DNA and left side 1.0 µL.

### 5.2.3 In-fusion cloning of genotype-specific PCR amplified products

First, we performed a pilot in-fusion cloning of the PCR amplified product of tile 1 from the homozygote risk patients. The in-fusion master mix was mixed according to table 4.3, before incubating at 50ºC for 15 minutes. After incubation, 25 µL of the XL 1-Blue cells were transformed with 0.5 µL of in-fusion master mix and plated on agar plates containing ampicillin, before incubation at 37º C overnight. As shown in figure 5.9, the transformation was successful, resulting in multiple colonies in the positive control plate. Moreover, the

cloning appeared to work as well, since the sample plates transformed with the in-fusion cloning product of pGL4.23-*luc2*-minP-tile-1 derived from patients V586 or V609 gDNA also produced multiple colonies. Importantly, the negative control (transformed with the digested vector only) produced very few colonies, suggesting that most of the colonies in the cloning plates resulted from successful cloning rather than vector self-annealing. Moreover, verification by Sanger sequencing (section 5.2.5) confirmed the right insert within cloned plasmid.

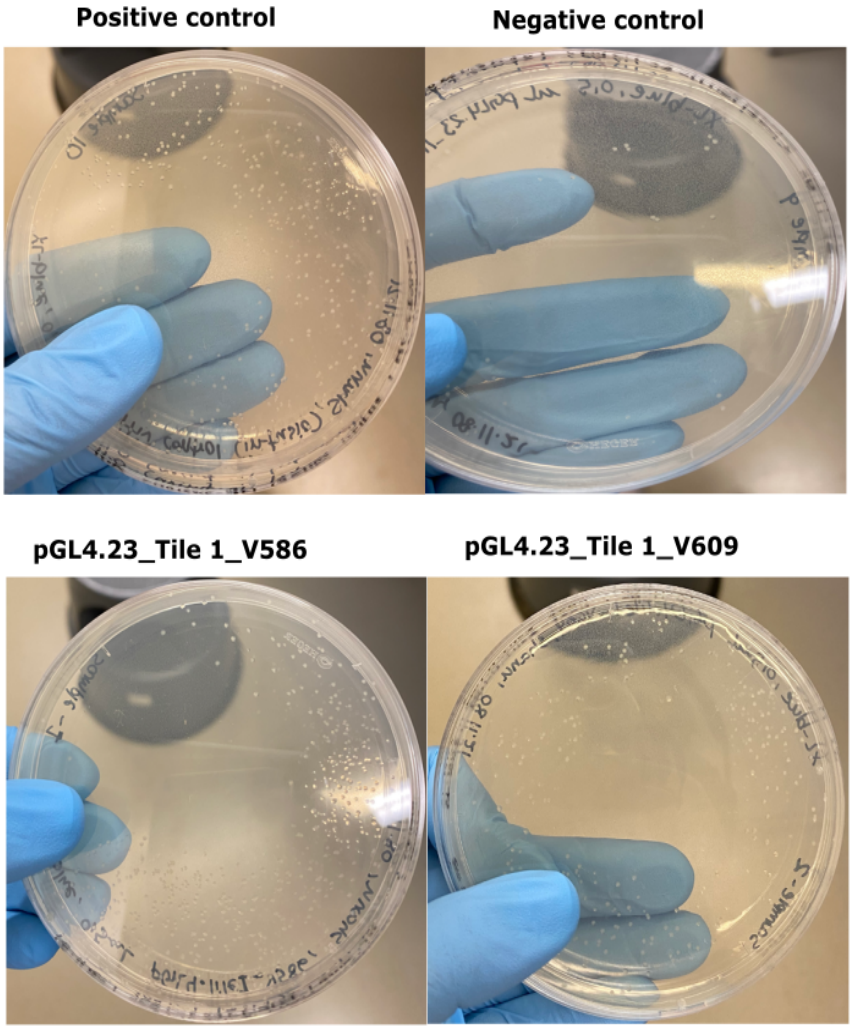
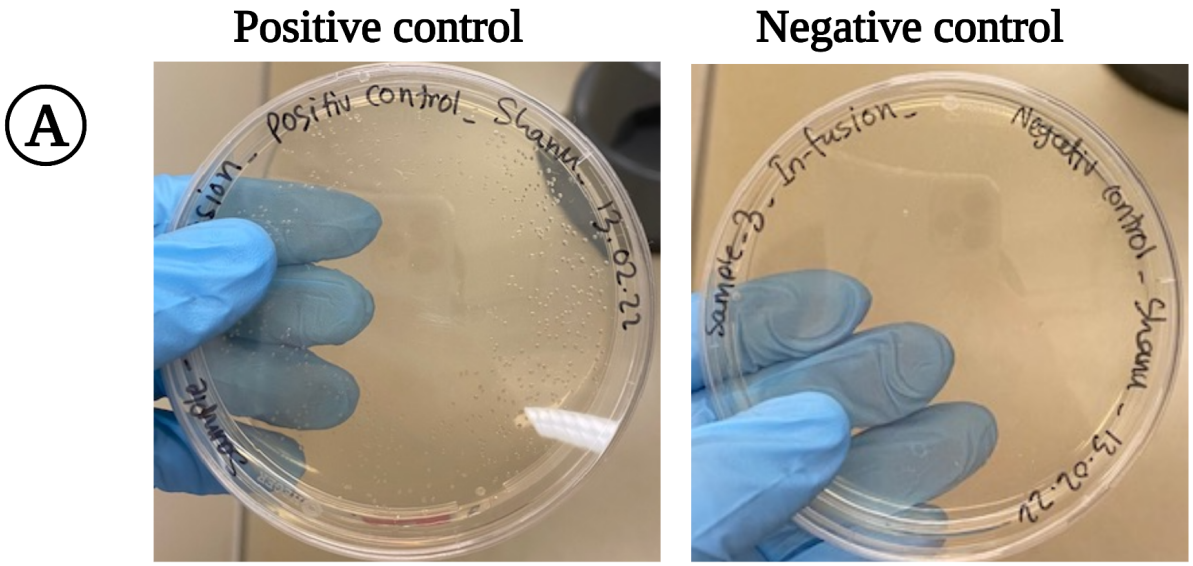
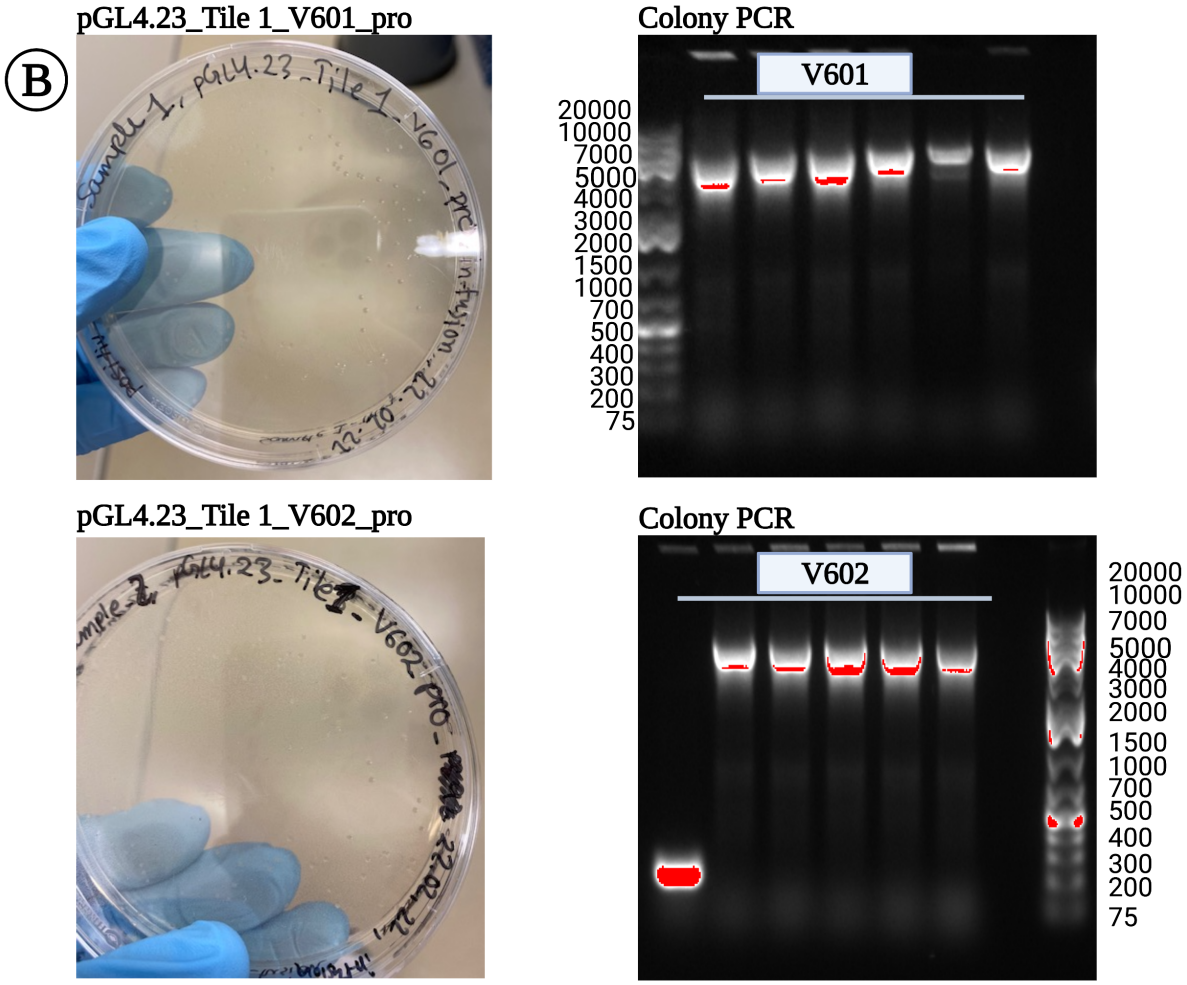



**Figure 5.9 Colonies of pGL4.23-*luc-2*-minP containing genotype-dependent PCR fragment of tile 1**. The first two image on the top representing positive control (from the kit) –colony formation as expected and the negative control (only linearized pGL4.23-*luc-2*-minP vector) very few colonies. The third and fourth image represents pGL4.23-*luc-2*-minP vector containing tile 1 of risk patients.

Following the successful cloning of tile 1 from two risk patients we next performed in-fusion cloning of the remaining PCR amplified tiles (tile 1 protective types, as well as protective and risk types of tile 3 and tile 8) using the same protocol (figure 5.10). Additionally, we

performed colony PCR (section 4.3.7, subsection A) on six individual colonies for each cloned tile to quickly identify positive clones for expansion and plasmid purification.



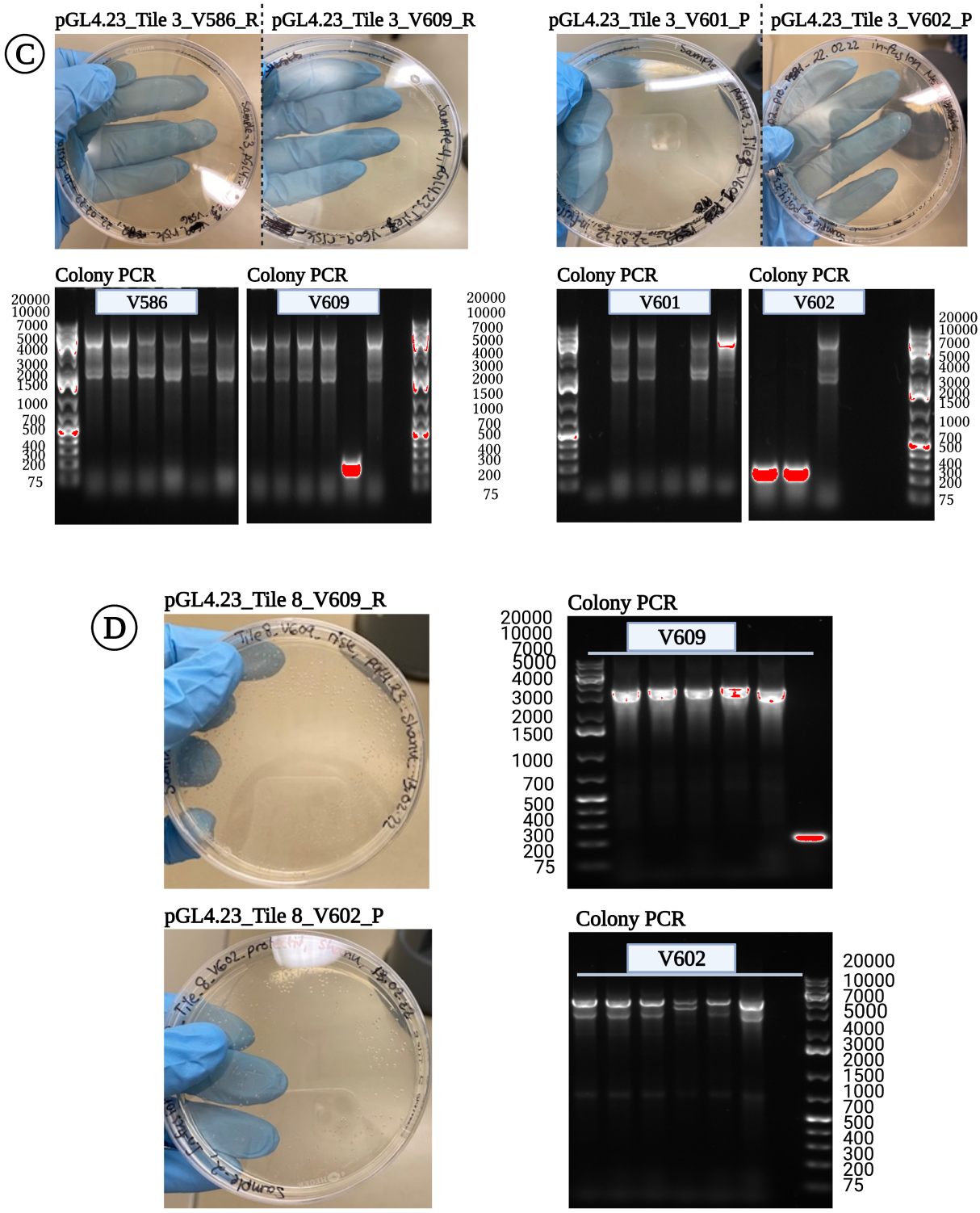



**Figure 5.10 Colony growth of cloned tiles and verification by colony PCR. A)** represents positive control (provided by the kit) and negative control (only linearized pGL4.23-*luc-2*-minP) used for alle the cloned tiles. **B)** image showing colonies with pGL4.23_Tile 1_V601_Protective and pGL4.23_Tile 1_V602_Protective construct and colony PCR respectively. **C)** the first lane represents all the constructs of tile 3 including pGL4.23_Tile 3_V586_Risk, pGL4.23_Tile 3_V609_Risk, pGL4.23_Tile 3_V601_Protective and pGL4.23_Tile 3_V602_Protective. The second lane is the presentation of PCR amplification of individual colonies of each construct. **D)** first two image represents colonies containing pGL4.23_Tile 8_V609_Risk construct and colony PCR of the construct. Last two image showing colonies with pGL4.23_Tile 8_V602_Protective construct and colony PCR of the colonies.

The positive control contained many colonies and the negative control none (figure 5.10 A). All the cloned constructs had colony growth (figure 5.10B-D). Figure 5.10 B) shows colony growth pattern of pGL4.23_Tile-1_V601 and pGL4.23_Tile-1_V602 (protective genotype) and the corresponding1.2% agarose gel image verifies that bacterial clones had the right insert (around 4000 bp). In contrast, tile 3 risk constructs had more colonial growth than protective constructs (figure 5.10 C). As for colony PCR, almost all the individual colonies had the right construct with the right DNA fragment size around 5000 bp. The pGL4.23_Tile 3_V602 (protective) construct had only three colonies and only one of the colonies contained the right construct with right size insert as colony PCR verified. The pGL4.23_Tile 8_V609 risk and pGL4.23_Tile 1_V602 protective constructs had many colonial growths and all the individual colonies chosen for colony PCR had the right insert (4000bp). Tile 8 had only one patient type from each group due to very low yield from DNA purification of amplified PCR fragments, which didn't allow to perform in-fusion assay (table 5.1). Thus, the results suggested that we were able to successfully clone all the tiles according to table 5.2.

### 5.2.4 Purification of cloned tiles

After verification by colony PCR, the selected colony of each construct was further grown in mini culture, before transferring 2.5 ml of the mini culture to 150 ml LB medium flasks . Plasmid DNA was then extracted by a plasmid DNA isolation/ purification kit (section 4.3.7, subsection C). The concentration of the purified plasmids were assessed by NanoDrop. As shown in table 5.2, tile 1 risk types and tile 8 both risk and protective type had very high yields. In contrast, all the tile 3 constructs had lower yield compared to the other constructs.

**Table 5.2. Overview of cloned tiles**.

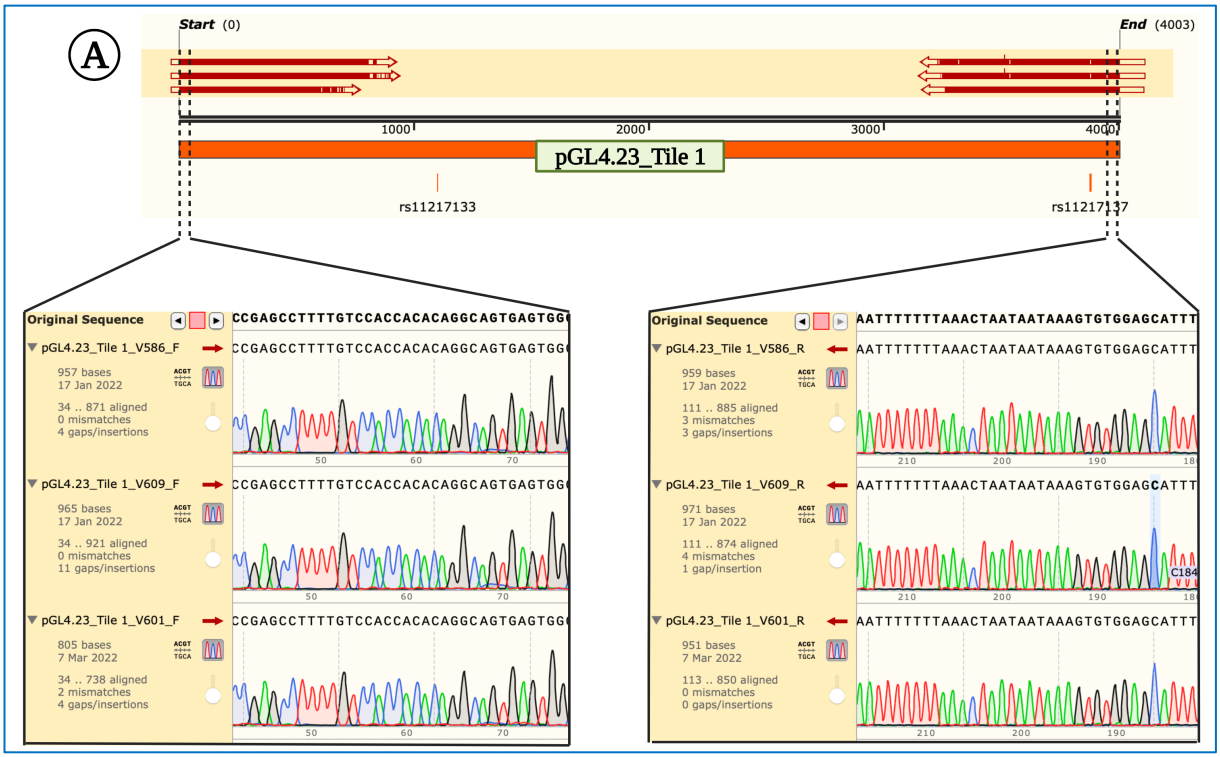
| Tiles | Cloned tiles | $ng/\mu L$ |
|---|---|---|
| **Tile 1** | pGL4.23_Tile 1_V586_R | 1603 |
| | pGL4.23_Tile 1_V609_R | 1430 |
| | pGL4.23_Tile 1_V601_P | 478,4 |
| | pGL4.23_Tile 1_V602_P | 513,7 |
| **Tile 3** | pGL4.23_Tile 3_V586_R | 444,5 |
| | pGL4.23_Tile 3_V609_R | 333,3 |
| | pGL4.23_Tile 3_V601_P | 340 |
| | pGL4.23_Tile 3_V602_P | 379,9 |
| **Tile 8** | pGL4.23_Tile 8_V609_R | 1845,2 |
| | pGL4.23_Tile 8_V602_P | 1696,2 |

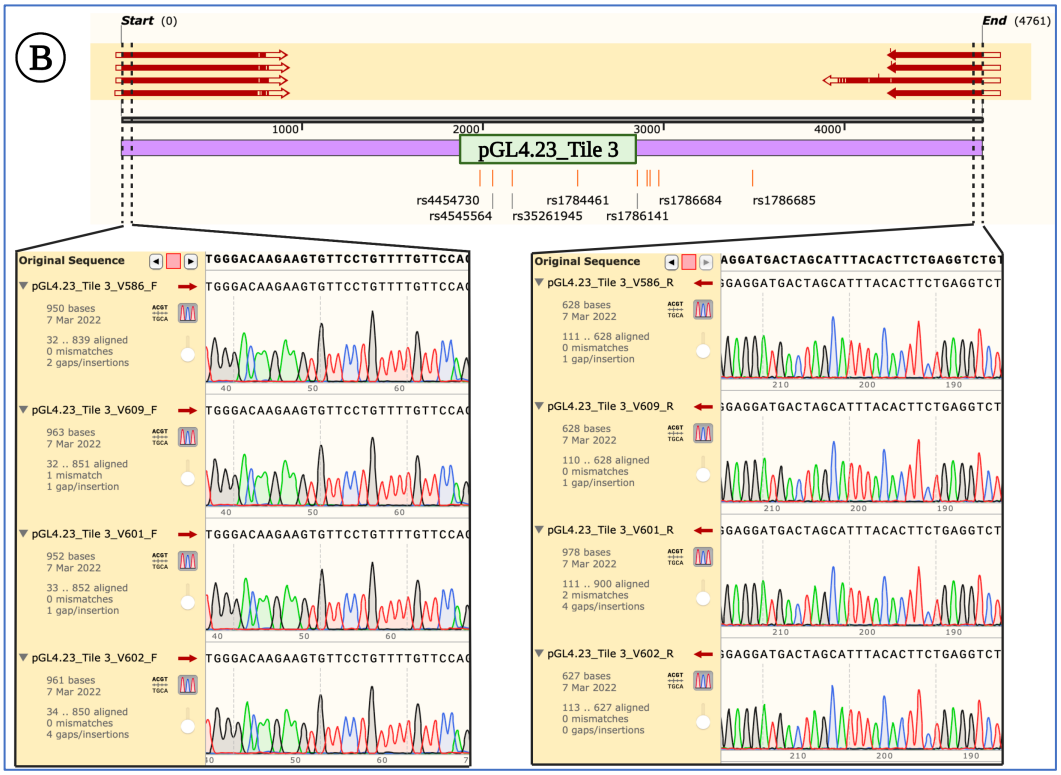
R stands for homozygote risk type and P stands for homozygote protective type.

### 5.2.5 Verification of cloned tiles by sanger sequencing

Sanger sequencing was performed for further verification of the purified cloned tiles. Big dye® Terminator v3.1 Cycle Sequencing kit reaction was set up according to table 4.8 and the thermal cycle program was used according to table 4.9. Moreover, unlike the previous sanger sequence (section 5.1.1, subsection D) experiment gene-specific primers were used, a vector-specific primer pair was used in this experiment for efficient batch PCR of all the constructs. The resulting sequencing files were visualized and analyzed in the SnapGene web tool.

As expected from colony PCR, the sequencing verified that all tiles analyzed contained the correct insert (figure 5.11). Unfortunately, tile 1 from patient 602 was not analyzed due to a methodological error. Visualization and analyzing the raw data from sanger sequence in the SnapGene verified that all the tile 3 and tile 8 constructs had the correct insert. To summarize, both colony PCR (section 5.2.3) and sequence data assessed in this study strongly suggested that we successfully cloned all the tiles presented in table 5.2.





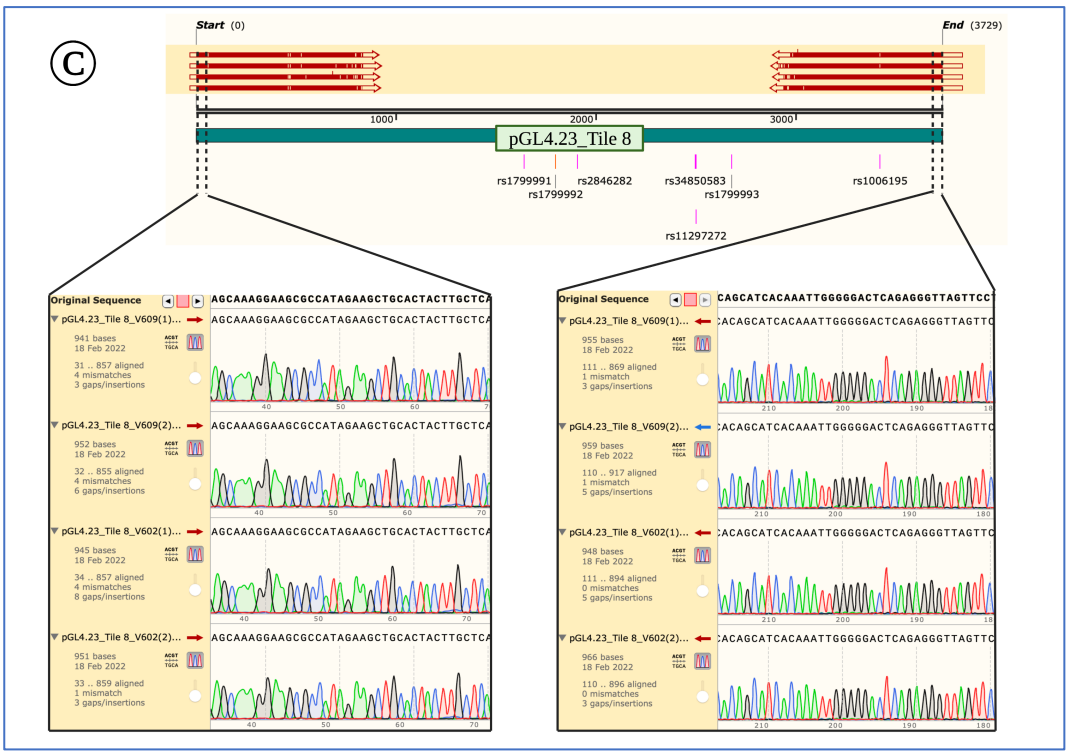




**Figure 5.11 Representation of sequence data of tile 1, 3 and 8. A)** The electropherogram shows cloned constructs of tile 1 of both risk and protective gDNA both in 5'→3' end and 3'→ 5' direction. Also shows the two SNPs in the tile 1. **B)** represents sequencing data of tile 3 constructs, including pGL4.23_Tile 3_V586, pGL4.23_Tile 3_V609, pGL4.23_Tile 3_V601 and pGL4.23_Tile 3_V602. The electropherogram on the left side shows amplified data from 5'→ 3' direction and the right side electropherogram shows sequences from 3'→ 5' direction. **C)** The first two red bars both on the left

and right side are the two parallels of pGL4.23_Tile 8_V609_risk and below the enlarged version of the marked region. The last two bars are parallels of pGL4.23_Tile 8_V602_protective.

## 5.3 Assessment of genotype-dependent enhancer activity

Having successfully cloned three tiles (1, 3, and 8), with both the risk and the protective genotypes into the luciferase reporter vector, we next sought to investigate the enhancer activity of each tile. Figure 5.12 shows an overview of the genomic position of the cloned tiles within the locus, as well as SNP density and predicted enhancer activity.

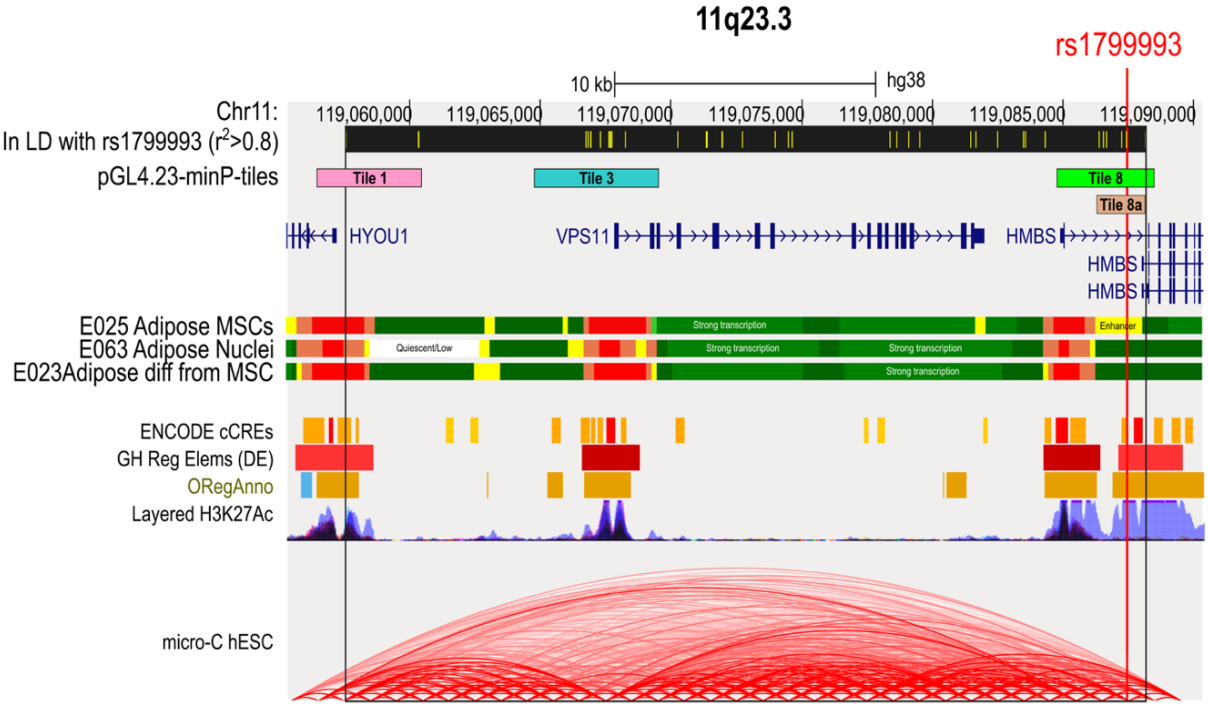
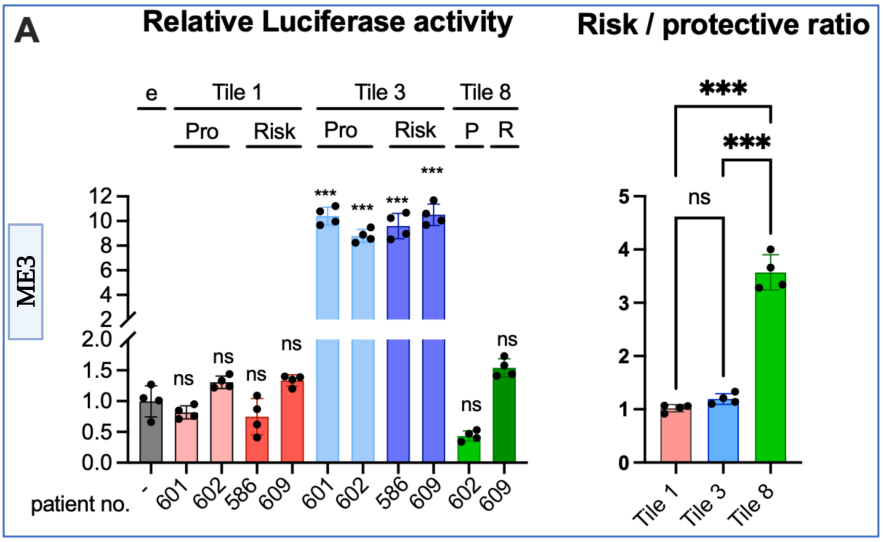
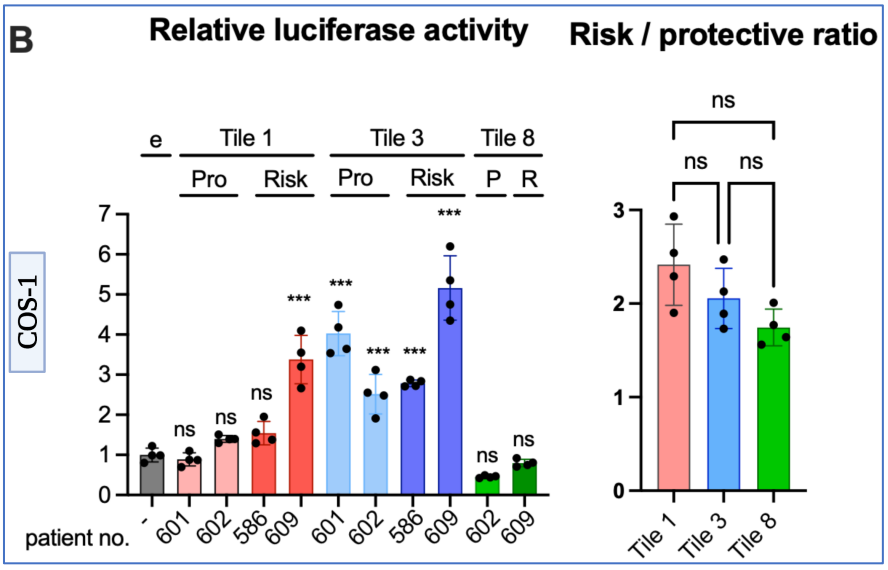



**Figure 5.12. Representation of all the cloned tiles.** The successfully cloned tiles which was verified by both colony PCR and sanger sequencing are highlighted with different color –tile 1 (pink), tile 3 (light blue), and tile 8 (green). Furthermore, the image is also showing tile 8a, which was created in the previous study. Data from WashU Epigenome browser (E025, adipose mesenchymal cells) is indicating enhancer activity (highlighted with yellow color) in the intron of the long isoform of the *HMBS* gene. The four parameters (ENCODE cCREs, GH Reg Elems (DE), OregAnno, Layered H3K27Ac) are also indicating both high promoter and enhancer activity. As for tile 1 and tile 3 all these various data sources are indicating toward strong active transcription start site/ promoter activity.

Each tile, as well as the empty reporter plasmid, was transfected into three cell lines, including COS-1, ME3 and 3T3-L1. After 48h, the cells were lysed and the luciferase activity was measured (figure 5.13).


A
Relative Luciferase activity
Risk / protective ratio
ME3
e
Tile 1
Tile 3
Tile 8
Pro
Risk
Pro
Risk
P
R
patient no.
-
601
602
586
609
601
602
586
609
602
609
Tile 1
Tile 3
Tile 8
ns
***


B
Relative luciferase activity
Risk / protective ratio
COS-1
e
Tile 1
Tile 3
Tile 8
Pro
Risk
Pro
Risk
P
R
patient no.
-
601
602
586
609
601
602
586
609
602
609
Tile 1
Tile 3
Tile 8
ns
***

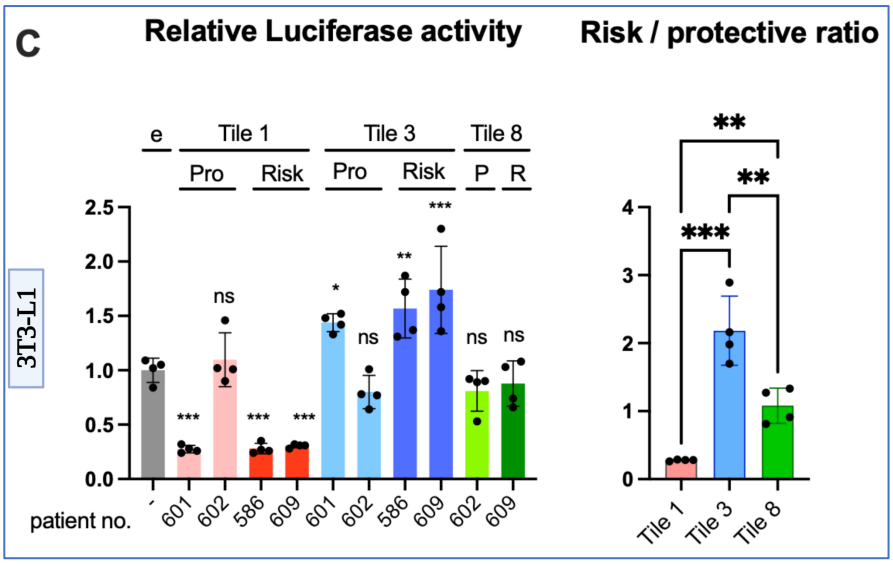

C
Relative Luciferase activity
Risk / protective ratio
3T3-L1
e
Tile 1
Tile 3
Tile 8
Pro
Risk
Pro
Risk
P
R
patient no.
-
601
602
586
609
601
602
586
609
602
609
Tile 1
Tile 3
Tile 8
ns
*
**
***

**Figure 5.13 Luciferase activity of tile 1, 3 and 8.** Luciferase activity in **A)** ME3, **B)** COS-1, and **C)** 3T3-L1 cells transfected with all the constructs of tile 1, 3 and 8. Y-axis of the left panel represents relative luciferase activity compared to the empty reporter plasmid (grey bars). All the light colors represent protective haplotype and darker color represents risk haplotype. The right panel shows the risk/protective ratio of all the cloned tiles from patient V609 (risk) and V602 (protective). All the data shown in both graphs are normalized to the pGL4.23-*luc2*-minP-empty vector. The bars indicating the mean ± SD of the luciferase activity (n=4). One way ANOVA with Holm- Šídák correction for multiple testing was performed to test for significant differences between groups. Left panel: each group compared to the empty vector. Right panel; each group compared to every other group. Ns –not significant * P <0.033, ** P <0.002, and *** P <0.001. Representative results from 2-3 independent experiments (supplementary data figure 10.2 )

Relative luciferase activity of tile 3 constructs in ME3 (beige) cells had very high signal compared to other constructs with luciferase activity about 10 times higher than the empty control vector, which was highly significant (figure 5.13 A, left panel). In contrast, tile 1 constructs showed variable and minor non-significant effects independently of the genotype (figure 5.13A). For tile 8, which had been cloned from only one patient of each genotype, the protective version showed a 50% reduction in activity, although this effect did not achieve statistical significance. Similarly, the risk variant of tile 8 showed a trend towards elevated activity. Thus, when calculating the ratio between risk and protective genotypes, tile 8 showed a highly significant 3.5-fold higher enhancer activity in the risk compared to protective genotype (figure 5.13A, right panel). In contrast, no genotype-dependent differences were observed for tile 1 or 3.

In COS-1 cells, a similar luciferase activity pattern as in ME3 cells was observed, with the tile 3 constructs, being significantly different from the control (figure 15.13 B, left panel). Signals produced by tile 1 constructs (with the exception of tile 1_V609 which produced high signals) and tile 8 constructs were not significantly different from the control. In COS-1 cells the relative risk/protective ratio was roughly 2 for both tiles 1, 3 and 8 (figure 15.13B, right panel. In 3T3-L1 cells, more variation was observed. While most tile 1 tiles showed a significant repressive effect, tile 1 derived from patient 602 was unchanged compared to the control (15.13.C, left panel). The signals produced by tile 8 in both protective and risk type were also unchanged in the 3T3-L1 cells. As for tile 3 constructs, the signal pattern was similar to the other two cell lines with elevated signal, except for patient 602, which was unchanged. In terms of risk/protective ratio, tile 1 showed a reduction, tile 3 an increase and tile 8 no difference between risk and protective tiles) when directly comparing patients 602 and 609 (figure 15.13.C, right panel). Of note, the ratios were closer to 1 for tile 1 and 3 when taking all patients into account (data not shown). In summary, although all the constructs of

tile 3 in all three cell lines had the highest enhancer activity, tile 8 manifested the most convincing genotype-dependent enhancer activity.

## 5.4 Effect of individual SNPs on enhancer activity

We next wanted to investigate possible mechanisms underlying the different enhancer activities between risk and protective haplotype of tile 8. We therefore examined whether any of the SNPs in the LD block were located in this tile. Indeed, 7 SNPs, including rs1799993 were found to be present in tile 8. To investigate the relative contribution of each SNP to the increased enhancer activity in the risk haplotype, we decided to perform site-directed mutagenesis of each individual SNP. As a proof-of-concept, we initially performed mutagenesis on a previously generated shorter version of tile 8; the tile 8a, which covers six of the seven SNPs, figure 5.14. This tile was of particular interest since it had only been cloned from a protective, but not risk donor[88].

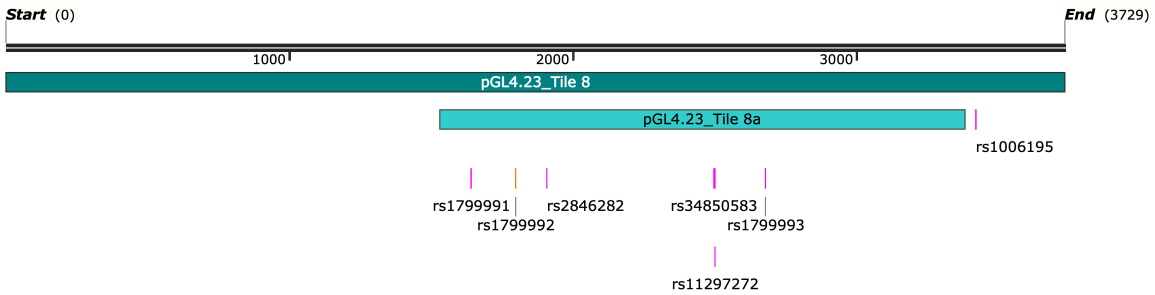



**Figure 5.14 SnapGene image/ map of tile 8 and tile 8a.** Tile 8 contains seven SNPs in total and 6 of these SNPs are situated in tile 8a, table 5.6 represents overview of all the SNPs in tile 8.

**Table 5.3 Overview of individual SNPs in tile 8.**

| Chr | Pos(hg38) | LD ($r^2$) | LD(D') | Variant | Ref | Alt |
|---|---|---|---|---|---|---|
| 11 | 119086381 | 0.88 | 0.95 | rs1799991 | G | A |
| 11 | 119086536 | 0.87 | 0.95 | rs1799992 | T | C |
| 11 | 119086647 | 0.9 | 0.95 | rs2846282 | G | A |
| 11 | 119087238 | 0.97 | 1 | rs34850583 | AAG | A, AA |
| 11 | 119087239 | 0.97 | 0.99 | rs11297272 | AG | A |
| 11 | 119087417 | 1 | 1 | rs1799993 | C | A |
| 11 | 119088159 | 0.95 | 0.99 | rs1006195 | G | T |

Chr stands for chromosome, pos(hg38) is the position of each SNP in the chromosome (GRCh38 Human Reference). Ref stands for "reference sequence" and Alt stands for "alternative sequence". Reference sequence is protective type and alternative sequence is the risk type. This table was adapted from: https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php

### 5.4.1 Site-directed mutagenesis of tile 8a

Site-directed mutagenesis (section 4.3.6) is a method that allows specific changes to a desired DNA sequence. In this study, we attempted to change each protective variant to the risk variant of each the SNPs within tile 8a (figure 5.14). Upon closer inspection of the SNPs, rs34850**83** and rs11297**72** risk variants were both found to be representing a deletion of the same "G" nucleotide. Thus, in this thesis, mutation of rs11297**72** represents both of these SNPs. Therefore, from the protective tile 8a, five mutant plasmids were generated with mutations in a single SNP each.

First, PCR amplification was performed with site-directed mutagenesis primes on the desired double stranded (tile 8a) DNA, followed by digestion of the parental DNA with DpnI. The amplified products were visualized on 1,2% agarose gel. In this study, we edited five SNPs including –rs1799991, rs1799992, rs2846282, rs11297272 and rs1799993 from protective type to risk type according to table 5.3 . Length of the plasmid (4283 bp) and the insert (tile 8a, 1854 bp) together was approximately 6200 bp (figure 5.15). All the site-directed mutagenesis-amplified pGL4.23-minP-tile 8a products had a visible band around 6200 bp, indicating right sized DNA fragments.

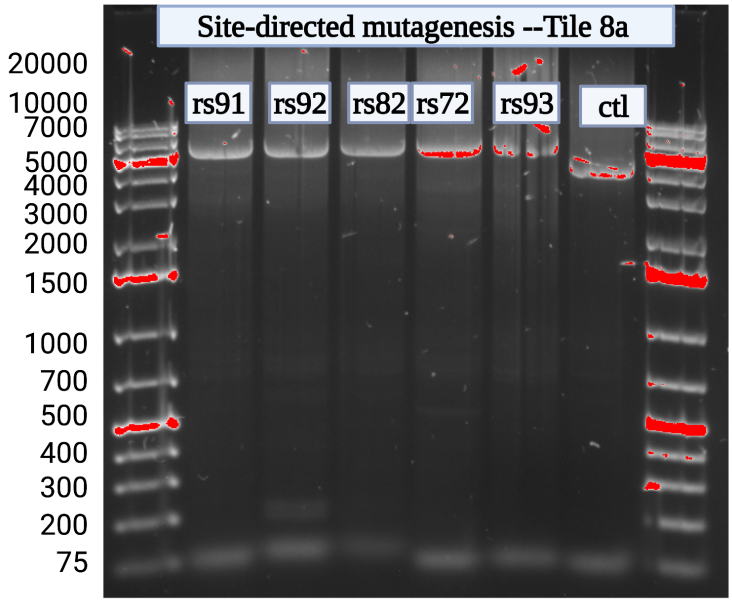



**Figure 5.15 Gel electrophoresis image of site-directed mutagenesis PCR amplified products.** The figure represents the 5 SNPs we intendent to edit through site directed mutagenesis. Ctl stands for control, which was an unrelated positive mutagenesis control provided by the kit. Both the plasmid and the insert of the control were 2000 bp, thus in total 4000 bp.

Based on the result, next we transformed 25 µL of the XL 1-Blue cells with 0.5 µL of site-directed mutagenesis master mix and plated on agar plates containing ampicillin, before incubation at 37º C overnight. As figure 5.16 shows, all the plasmids containing the modifications had colonial growth. Furthermore, positive control (provided by the kit) also had colonial growth. The colonial growth indicated molecular successful site-directed mutagenesis of tile 8a.

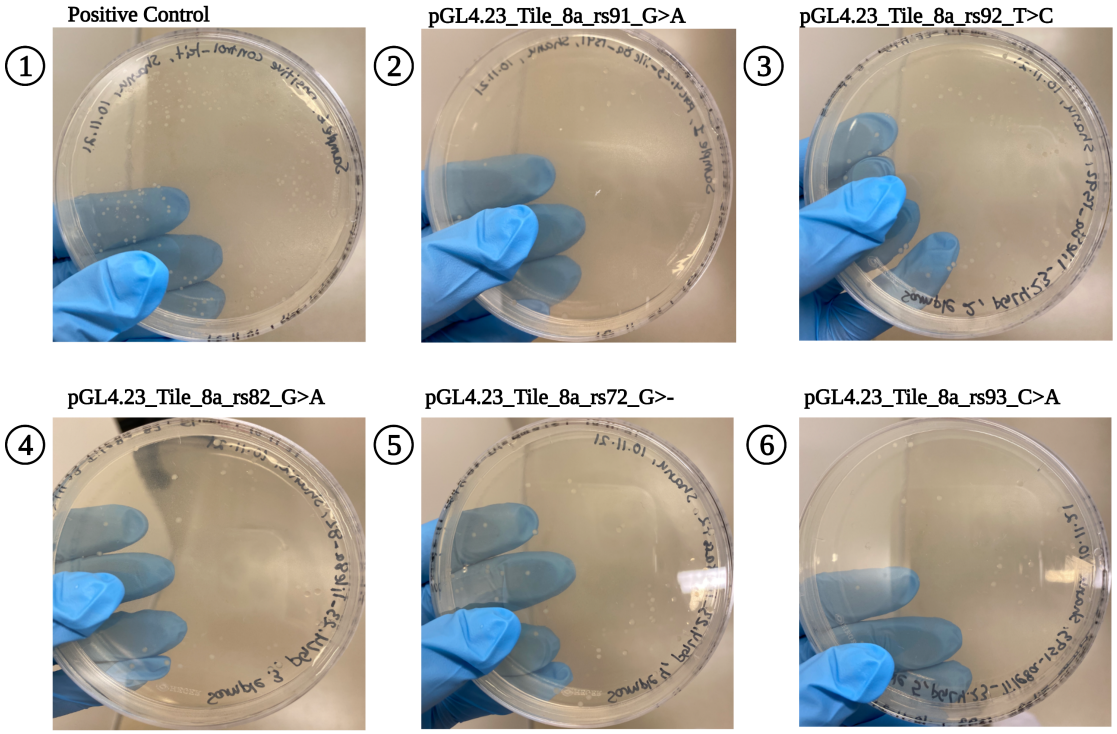



**Figure 5.16 Colony formation of site-directed mutagenesis of tile 8a constructs**. We were able to clone all the site-directed mutagenesis constricts of tile 8a, including the SNPs rs1799991, rs1799992, rs2846282, rs11297272 and rs1799993.

### 5.4.2 Plasmid purification of tile 8a mutants

A single colony from each construct was grown in mini culture at 37º C overnight. The next day 2.5 ml of mini culture was transferred to a bigger flask containing and grown overnight again. A plasmid purification kit (section 4.3.7, subsection C) was used to isolate insert DNA from the cells and concentration was measured by NanoDrop. Table 5.4 shows an overview of all the mutants created in this study. As the table shows, alle the constructs created by this study had very high concentration/ yield.

**Table 5.4 Purified concentration of the mutant 8a tiles, assessed by NanoDrop.**

| Tiles | Cloned tiles | $ng/\mu L$ |
|---|---|---|

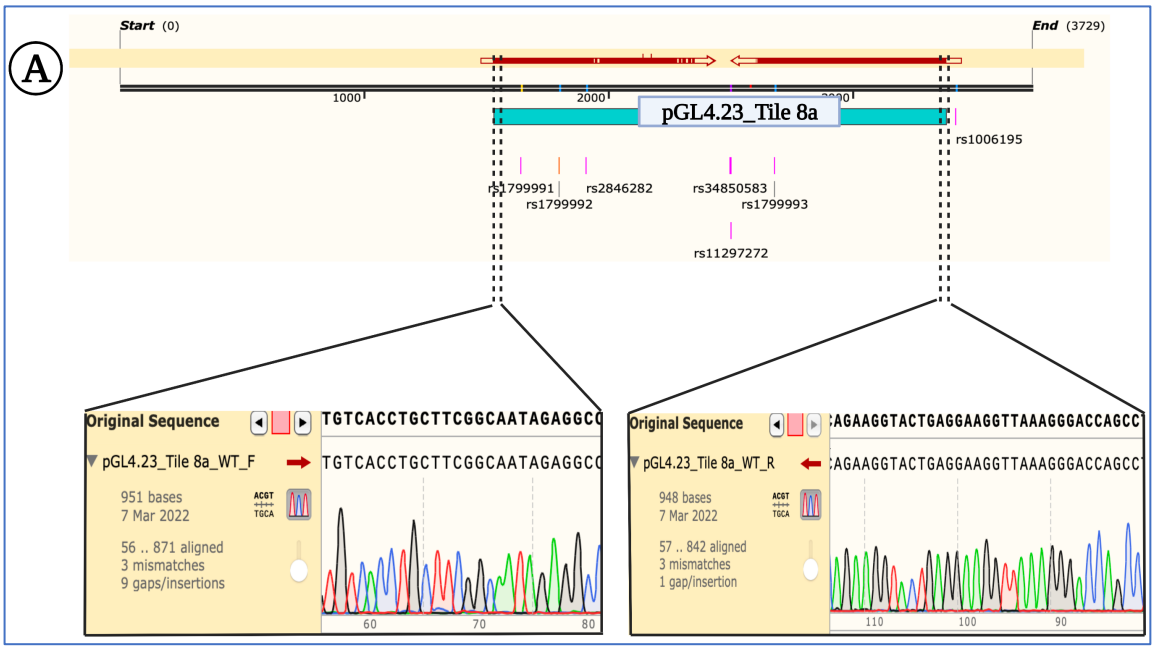
| | | |
|---|---|---|
| Tile 8a | pGL4.23_Tile 8a_WT_protective | 481 |
| Tile 8a_rs91 | pGL4.23_Tile 8a_rs91_G >A_risk | 1374 |
| Tile 8a_rs92 | pGL4.23_Tile 8a_rs92_T > C_risk | 1695 |
| Tile 8a_rs82 | pGL4.23_Tile 8a_rs82_G >A_risk | 1697 |
| Tile 8a_rs72 | pGL4.23_Tile 8a_rs72_G>-_risk | 1343 |
| Tile 8a_rs93 | pGL4.23_Tile 8a_rs93_C >A_risk | 1589 |

WT stands for wild type, Del stands for deletion. All the SNPs presented in this table are abbreviated version of the original /full name of the SNPs.

### 5.4.3 Verification by sanger sequencing

To verify if the editing/modification of the SNPs was correct, sanger sequencing was performed on the purified constructs presented in table 5.4, except for construct pGL4.23_Tile 8a_rs72_G>-_risk. The reason for not performing sequencing PCR on this construct is that the SNP is located outside of the 800 bp range of sanger sequencing in both 5'→3' direction and 3'→5' direction. Due to time constraints, separate sequencing primers for this area were not designed. For the remaining mutants, already available plasmid specific primers were used for sequencing, and the data was analyzed by SnapGene.

The WT tile 8a used as starting point for the mutagenesis was verified to possess a protective genotype at each of the SNPs (figure 5.17A, B and supplementary data, figure 10.5), confirming an intact haplotype. Moreover, all the sequenced tile 8a mutants were found to contain the correct respective mutations (figure 5.17 B and supplementary data figure10.5 ). Thus, sanger sequencing confirmed successful editing and cloning of desired constructs (table 5.4).



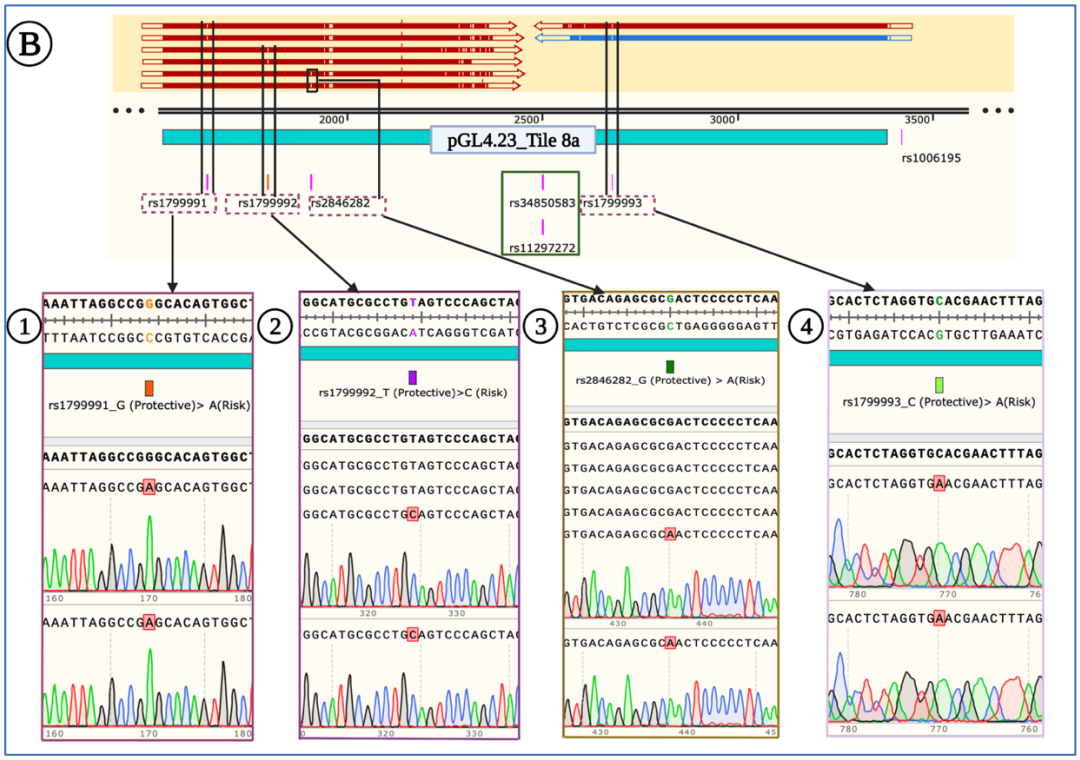



**Figure 5.17 Sanger sequencing of the tile 8a_WT and mutants. A)** Tile 8a_WT protective type and the SNPs located in the region. Red arrows indicate sequencing direction and coverage. The electropherogram below shows sanger sequence data from both the 5'→3' direction and 3'→5' direction. **B)** Sequence of all the mutants aligned to the WT sequence. Each plasmid/SNP was sequenced with 2 parallels, using plasmid specific forward or reverse primers. Red arrows indicate sequencing direction and coverage. Electropherogram peaks, as well as the corresponding sequence surrounding each SNP are shown in boxes. 1, rs.91; 2, rs.92; 3, rs.82; 4, rs.93. Mutated SNPs are highlighted in red. WT sequences are shown directly above each mutant sequence. The rs11297272 was not sequenced and marked with green rectangle.

### 5.4.4 Assessment of genotype-dependent enhancer activity by luciferase assay

Following successful site-directed mutagenesis cloning, we assessed SNP-dependent enhancer activity by luciferase assay (figure 5.18).

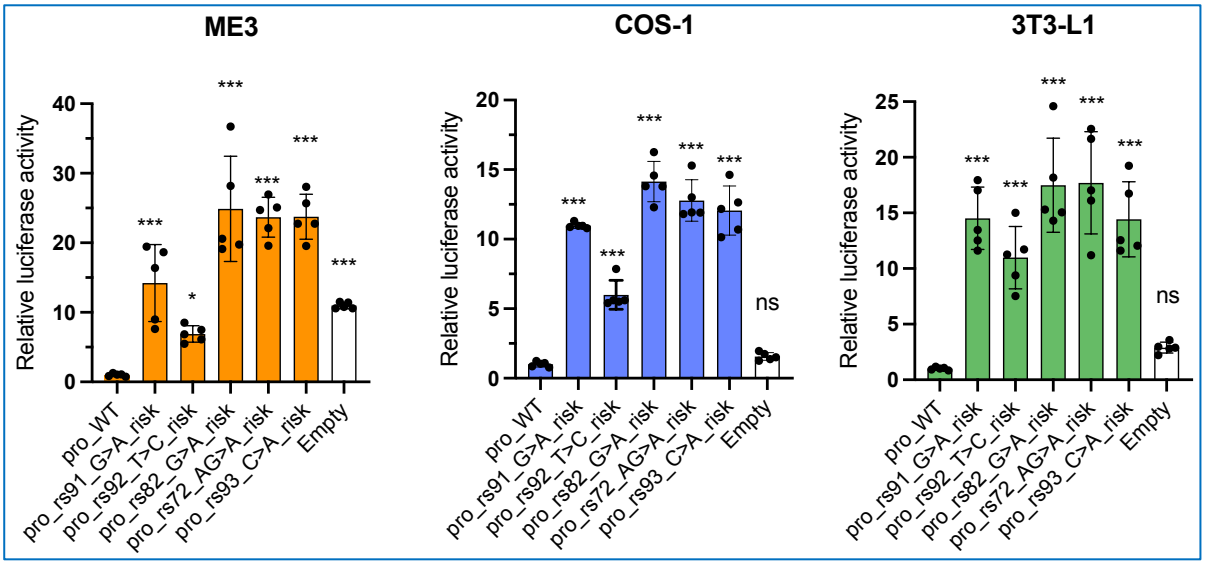


**Figure 5.18 Luciferase activity of tile 8a protective and risk haplotype.** All the data shown are normalized to the pGL4.23-Tile 8a_WT protective plasmid. The bars indicating the mean ± SD of the luciferase activity (n=5). One way ANOVA test was run to determine significancy on luciferase activity between pGL4.23-Tile 8a_WT protective plasmid and all the tiles/ constructs. * P <0.033, and *** P <0.001. This figure represents only one replicate out of independent biological triplicates (supplementary data figure 10.3).

In all cell lines, the protective/WT tile 8a showed clear repression of enhancer activity as observed for the long tile 8 in ME3 cells. Strikingly, all the individual risk mutants showed a strong and highly significant increase in enhancer activity in all cell lines, although the effect was less strong for rs1799992. These results suggest that all these SNPs may be causal by strongly contributing to the risk-dependent increased enhancer activity in tile 8a of the 11q23.3 locus.

### 5.4.5 Investigation of C/EBPb as a mediator of SNP-dependent effects

Next, we decided to investigate how any of the SNPs within tile 8a could alter the enhancer activity so dramatically in a genotype-dependent manner. Since SNP variants likely affect binding sites of TFs, we hypothesized that the risk variants in this case could either introduce a new binding site of a transcriptional activator (gain-of-function) or destroy the binding site of a repressor (loss-of-function). Previously, a search for consensus transcription binding sites overlapping the SNPS of tile 8a using the Promo predictions web tool have been performed[88]. Interestingly, they found the risk sequence of rs2846282 to introduce a binding site for the TF C/EBP-β (figure 5.19), a key TF for adipocyte differentiation. We therefore investigated whether co-expression of C/EBP-β with the reporter tiles could increase the enhancer activity specifically in the 8a rs2846282 risk reporter compared to WT 8a or any of the other mutants (figure 5.20).

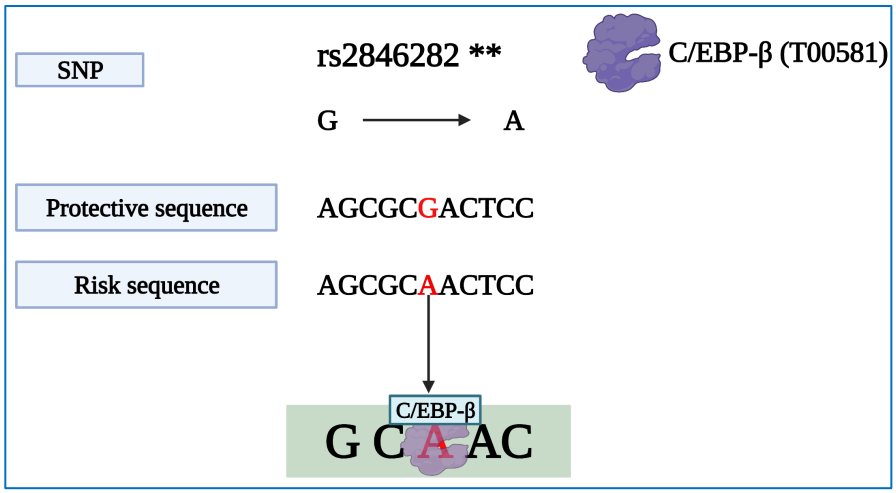


**Figure 5.19 Transcription binding sites overlapping the rs2846282 SNP**. The image shows a binding site for C/EBP-β in the risk, but not protective sequence of rs2846282. This figure was adapted from Nina Tangen Samuelsen's master thesis (Functional dissection of a genetic locus for visceral fat mass (11q23.3), figure 5.1) and created in Biorender.com.

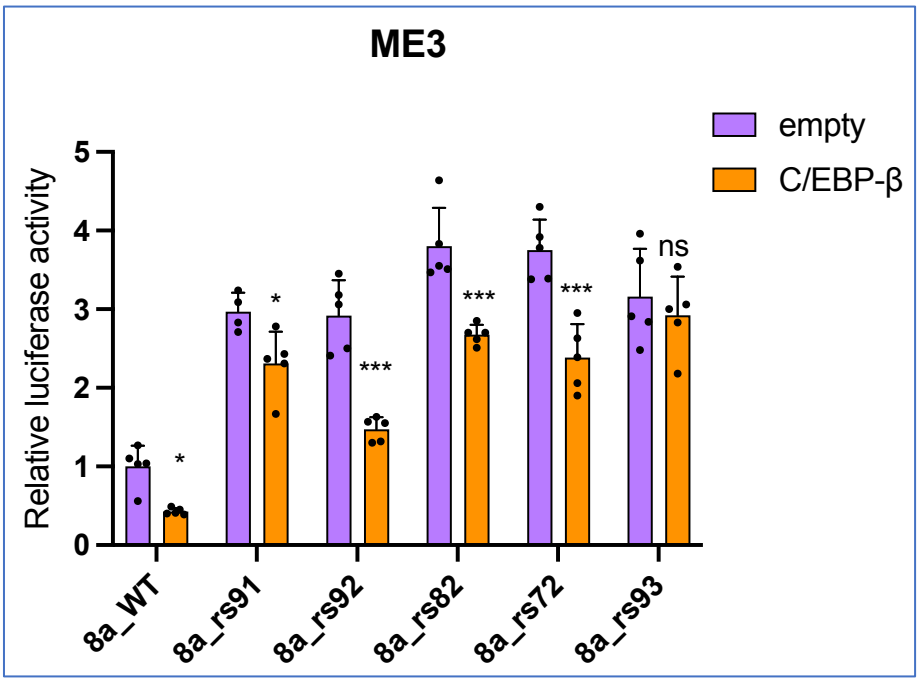



**Figure 5.20 Comparison of luciferase activity between reporter plasmid and overexpression plasmid.** All the data shown are normalized to the pCMV-empty plasmid. The bars indicating the mean ± SD of the luciferase activity (n=5). Two-way ANOVA test was run to determine significancy on luciferase activity between pCMV-empty plasmid and pSPORT-CEBP-$\beta$ within the reporter plasmids. Ns –not significant P <0.12, * P <0.033, and *** P <0.001.

Surprisingly, co-expression with CEBP-β resulted in significantly lower enhancer activity for most of the tiles compared co-expression with empty plasmid. These data clearly show that CEBP-β did not have any activator effect on the reporter plasmids. Thus, we conclude that CEBP-β is not the mediator of the SNP/genotype effect we observed in the previous luciferase assays (figures 5.13, 5.18 and 5.20).

### 5.4.6 Search for additional TF binding motifs overlapping SNPS in tile 8a

Since the predicted CEBP- β binding site identified by PROMO was found to not be functionally relevant, we used a different approach to search for potential other TF binding sites. Using UCSC Genome Browser (chr11:119086268-119088121, GRCh38 Human Reference) and the JASPAR CORE 2022 TF binding tracks, we were able to identify multiple TF binding sites overlapping several of the protective SNPs within tile 8a (figure 5.21 and table 5.5). Having found intact TF binding sites in the protective variants, we next

used the JASPAR database directly to score the predicted binding of each of these TF to the protective and risk genotype of tile 8a (table 5.5).

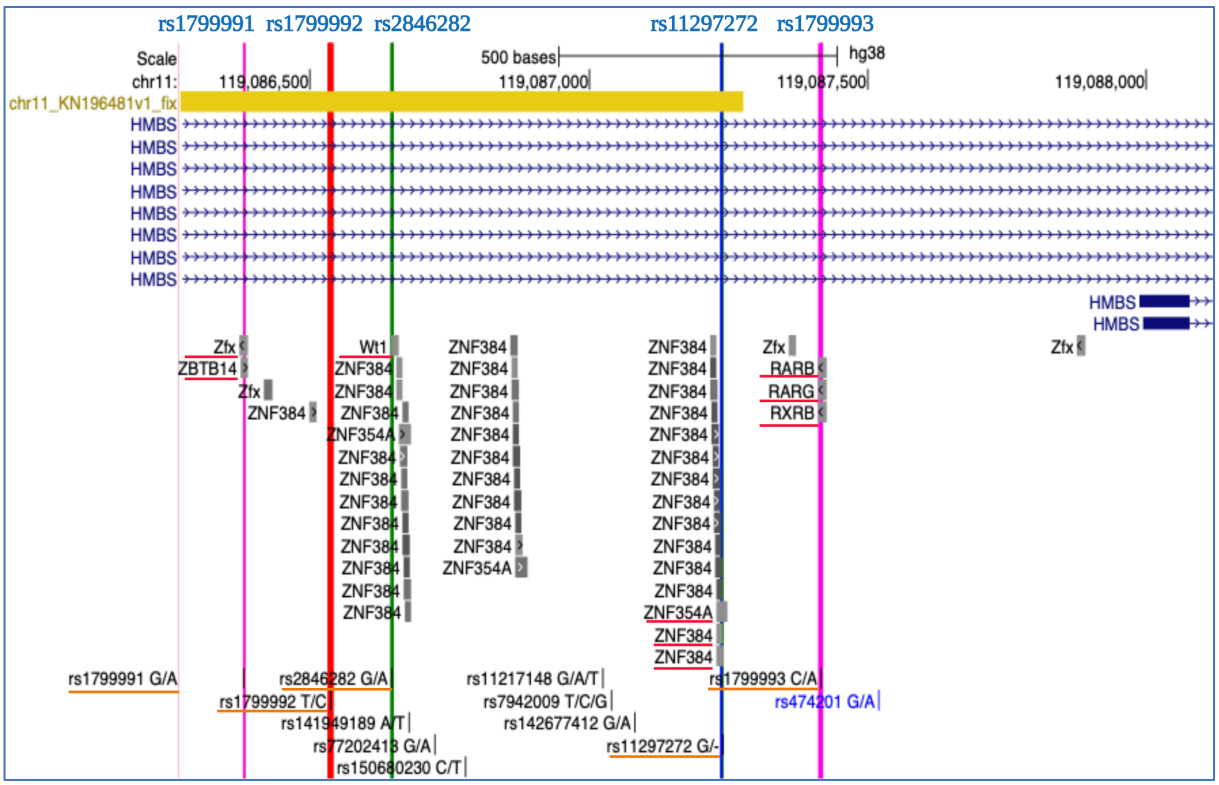



**Figure 5.21 Overview of TF binding sites within the tile 8a SNPs**. The upper navigation bar represents the position within chromosome 11 for the expanded visualization of *HMBS* gene locus below. The next parameter is the JASPAR CORE 2022 profile of genome-wide predicted binding sites of predicted TF. The last parameter represents short genetic variants from dbSNP build 153: single nucleotide variants. The colorful bars along the image representing the SNPs within tile 8a, and in the dbSNP 153 parameter all the SNPs are highlighted with orange bars. Within the JASPAR CORE 2022 parameter all the red colored highlighted TFs are the TFs that bind to the SNPs within tile 8a. This figure was created in UCSC Genome Browser.

**Table 5.5 Overview of TF binding sites tile 8a_WT (wild type/protective).**

| SNP | Ref (P) | Alt (R) | TFs that bind protective genotype | TFs that bind risk genotype | Functions of TFs | Predicted effect of risk SNP |
|---|---|---|---|---|---|---|
| rs1799991 | G | A | Zfx (0.83) | Zfx (0.84) | Transcription co-activator[111] | No effect |
| | | | ZBTB14 (0.81) | ZBTB14 (0.86) | Transcriptional activator [112,113] | Slightly increased transcriptional activity |
| rs1799992 | T | C | - | - | - | - |
| rs2846282 | G | A | Wt1(0.88) | Wt1 (0.86) | Suppressor [114] | Minimum change |

| | | | | | | |
|---|---|---|---|---|---|---|
| rs11297272 | AG | A | ZNF354A (0.83) | - | Transcriptional repressor[115] | Loss of repressor activity |
| | | | ZNF384 (0.81) | ZNF384 (0.94) | Transcriptional activator[112] | Slight increase in transcriptional activity |
| rs1799993 | C | A | RARB (0.83) | - | Repressor of transcriptional activity[116] | Loss of binding and repressor activity |
| | | | RARG (0,85) | - | repressor of transcriptional activity[117] | Loss of binding and repressor activity |
| | | | RXRB (0.84) | - | Repressor of transcriptional activity[116] | Loss of binding and repressor activity |

All the data were retrieved from JASPAR CORE 2022, Website: https://testjaspar.uio.no/. Numbers presented in the bracket are relative numbers retrieved from JASPAR CORE 2022. P stands for protective, and R stands for risk.

Strikingly the three TFs RARB, RARG, and RXRB were found to bind only to the protective, but not to the risk variant of the SNP rs1799993. Since these factors can act as repressors, this would result in loss of repressor activity and thereby increased enhancer activity. Similarly, the repressor ZNF354A was only able to bind to rs11297272 in the protective, but not the risk variant. Other TFs, including Zfx, ZBTB14, Wt1, and ZNF384 were found to bind to both the risk or the protective sequence of either rs1799991, rs2846282, or rs11297272, thereby likely not contributing to the genotype-dependent effect. In summary, we were able to find several additional TFs binding sites overlapping the SNPs in tile 8a and have found loss of repressor binding to rs11297272 and rs1799993 to be potential causal mechanisms underlying the elevated the enhancer activity in the risk haplotype.

# 6. Discussion

The underlying mechanisms that promote selective VAT storage are largely unexplored. This thesis aimed to investigate the VAT associated 11q23.3 locus and identify casual genetic variant(s) in cis-regulatory regions within the locus, with the long-term goal of revealing underlying mechanism of VAT storage and VAT associated diseases. To this end, multiple analyses were performed as discussed in this chapter.

## 6.1 11q23.3 locus associated with visceral adiposity

Fat distribution plays an important role in the risk of developing obesity related diseases, and VAT depots are found to be more harmful compared to other types of adipose depots in the body[39]. As mentioned earlier, GWAS is a population-based powerful method that has identified thousands of genetic variants that are associated with human diseases, including obesity and type 2 diabetes[118,119]. However, very few GWAS have investigated and identified VAT associated variants. A likely explanation is that VAT associated studies require advanced imaging techniques, which are very expensive and time-consuming techniques. Moreover, GWAS cannot identify the exact causal variants nor precise downstream target genes and underlying mechanisms of diseases[119]. However, a recent study by Karlsson et al. was able to identify over 200 new VAT-associated loci, including the 11q23.3 locus[24].
The Karlsson et al. study developed a computational prediction model that allowed VAT mass prediction on a large scale, and a GWAS was performed on the participants with predicted VAT mass. One of the GWAS signals, or VAT-associated tag-SNPs were rs1784461, which is located within the 11q23.3 locus, and shows an eQTL signal for *HMBS* and *VPS11* in VAT. The findings led to Karlsson et al. to perform luciferase assays to compare the regulatory activity between risk and protective alleles within the three regions with predicted regulatory activity of the11q23.3 locus (Karlsson et al. figure 1 b, c, d, and extended figure 5 A). Interestingly, they were able to map rs199993, a SNP in high LD with the tag-SNP rs1784461, to a regulatory region with genotype-dependent promoter activity and suggested that rs199993 is likely a casual variant whose risk variant may lead to increased *HMBS* expression and increased VAT mass via HMBS' known role in regulation of mitochondrial function[24].

However, there are some limitations with the Karlsson et al. study; they did not examine the entire LD block for regulatory activity, which means there might be undetectable casual

variants. Furthermore, Karlsson et al. study didn't investigate for TFs binding sites and target genes. Lasty, tile 3 (Karlsson et al. extended figure 5 A) was cloned into pGL4.10 to assess only promoter activity. The suggested causal SNP rs1799993 is located in the promoter region of the short isoforms of *HMBS* (figure 5.12), which are only expressed in blood cells[87], which is unlikely to be relevant for adipose tissue. Also, as the figure 5 showed, other *HMBS* isoforms (expressed in various tissues, including adipose tissue) contain additional exons further upstream, meaning that rs1799993 is located in the first intron of these variants. Therefore, rather than regulating the promoter of a blood-specific short isoforms of HMBS, we believe that this SNP (if causal), is probably rather regulating an enhancer that may act on the long isoforms of *HMBS*, *VPS11* (genes with eQTL signal) or another unknown gene. In this study we aimed to construct similar sized tiles covering the entire LD block to search for enhancer activity across the entire locus. Unfortunately, due to PCR challenges and limited time, only 3 tiles were successfully made. However, these tiles were cloned from patients with both the risk and protective haplotype, allowing a direct assessment of genotype-dependent enhancer activity in these areas. Of note, the tiles made in this study ended up covering roughly the same regions as those described in the Karlsson study. Thus, although this thesis did not manage to expand the number of investigated regions, it does allow a relatively direct comparison of genotype-dependent promoter versus enhancer activity.

In agreement with the Karlsson study, we found the regions harboring rs1799993 to show the most consistent genotype-dependent effect. However, while Karlsson et al. found the risk/protective ratio to be about 1.5 for promoter activity, we here found it to be ~3.5-20 for enhancer activity. Thus, the effect of the risk genotype was much stronger on enhancer activity. While the previous study only focused on rs1799993, we systematically investigated the contribution of several of the closest SNPs and actually found all of them to have a strong contribution. While our study was not designed to identify downstream target genes (see future perspectives), our identification of causal/contributing SNPs allowed investigation of the upstream regulatory mechanisms through TF binding motif discovery, as discussed further below.

## 6.2 Methodological limitations

### 6.2.1 Primer design and PCR amplification

To investigate the whole LD block, the first and most important step was to design the appropriate primers, since it would affect all the downstream approaches in this study. Multiple factors must be considered when designing primers, including specificity, melting temperature (Tm), GC content and degree of self-complementary and hair-pin structures[120]. Thus, to account for all these factors, primers were design using NCBI primer-BLAST with stringent criteria and validated using the Beacon Designer Premier Biosoft web tool (section 4.2). For each tile, three alternative primer pairs were designed (table 3.3) Although the tiles should ideally be of the same length, designing primers for some of the tiles were challenging. As a result, some of the tiles were larger than the other tiles. A likely explanation can be that the 11q23.3 locus is situated in a complex region of the genome, and a possible solution to primer associated problems would be chemically synthesize the tiles. Chemically synthesis does not require template DNA, rather addition of nucleotide to single stranded DNA, which then serves as template DNA[121]. Thus, this method would have allowed synthesizing identical sizes tiles. However, we did not use this approach as it is time consuming and very costly. Moreover, to be representative, the synthetic DNA would need to reflect actual patient DNA, meaning that comprehensive sequencing would be necessary, which would require design of at least 40 sequencing primers anyways.

To perform molecular cloning, amplification of desired DNA sequence by PCR plays a crucial role. As presented in section 5 (result), some of the amplified DNA products had unspecific bands or no visible band at all on the gel image. One of the reasons could be that annealing temperature was too low, since too low annealing temperature promotes off-target binding. It could be that the temperature in some cases was too high, making primers unable to bind at all, leading to little PCR product[122,123]. Thus, in this study we optimized annealing temperature, cycle numbers, and extension time. In some cases, the alternations were successful, while other cases would require further optimization. For example, testing of a broader range of annealing temperatures could have been performed. Moreover, we didn't optimize other components of the PCR, including divalent cations such as $Mg^{2+}$ (co-factor of Taq-polymerase), pH of buffers, concentration of nucleotidetriphosphates (dNTPs) or DNA polymerase[122,123]. Thus, in the future, for the remaining tiles, these parameters could be tested, so that the right DNA fragments can be amplified.

### 6.2.2 Troubleshooting –In-fusion cloning

In-fusion cloning is a recombinational cloning, based on site-specific recombination, where both the insert and the vector have homologues ends, which is recognized by the cloning enzyme[124]. In-fusion cloning was performed to clone the tiles into the pGL4.23_*Luc-2*_minP vector (section 5.2.3) for further functional assessment. We had difficulties with cloning of the tile 1 protective genotype and tile 3 both the protective and risk genotype, due to large number of colonies obtained with no inserts. It could be due to incomplete linearization of the plasmid (pGL4.23_*Luc-2*_minP empty), even though we have utilized NebCloner to check the compatibility between restriction enzyme and their respective buffer compatibility. Other possible reasons for the troubleshooting could be that in-fusion reaction was contaminated by plasmids with same antibiotic resistance or the agar plate contained wrong antibiotic[125]. Therefore, we digested the plasmid with the chosen restriction enzyme once more, performed in-fusion cloning, transformed competent cells, and added to agar plates with right antibiotics, which led to successful cloning of the remaining tiles.

### 6.2.3 Luciferase assay

Genetic reporter assay such as luciferase assay is utilized to determine whether a protein can activate or repress expression of a target gene, thus investigating gene regulation and transcriptional activity[126]. However, there are some limitations to luciferase assay, including the assay requires exogenous addition of decanal substrate, which can be made endogenously by bacterial system. Another disadvantage with the assay is the requirement of energy from ATP, which can complicate measurements or interpretation of luciferase activity[126]. Furthermore, there is a 4-6 hours delay from the effect of the biological treatment on transfected cells and response from the cells, which then leads to transcription[127]. Reporter gene assays can also exhibit gene expression or transcriptional noise, which can be either extrinsic noise or intrinsic noise[128,129]. Extrinsic noise is gene-independent, meaning the noise varies between cells and affects all genes. In contrast, intrinsic noise is gen-dependent since it effects varies from gene to gene and from cell to cell. Consequently, data collected from one luciferase experiment to another luciferase experiment might not be reproducible[128,129].

Transfection efficiency of the cells dependent on various factors including transfection method, cell line viability, percentage of confluency, passage number, both quality and quantity of the nucleic acid, and presence or absence of the serum in the medium[130]. In this study, among all the luciferase assay performed on the three different cell lines, 3T3-L1 cell line had the lowest value of luciferase signal, and the possible explanation is that low

transfection efficiency (discussed later) or luciferase assay associated limitations. However, we did not optimize the different parameters mentioned above that can affect transfection, the only optimization was different transfection reagents used for different cell lines.
Another limitation of this study was the use of a single-luciferase reporter. A dual-luciferase kit, using co-expression of a constitutively active different luciferase reporter, such as *Renilla* luciferase, could have been used to normalize transfection efficiencies[131,132]. Because a dual kit costs about 5-10X more pr reaction, luciferase assays were first performed using the single kit. Repeating all the luciferase assays with the dual kit was unfortunately not feasible due to time limitations but should be considered for future work.

In this study, luciferase data were normalized to the empty reporter, according to common practice. However, because the size of the different reporters containing tiles was almost double that of the empty reporter, this could affect transfection efficiency and thereby luciferase activity. A possible solution would be to clone a known "regulatory dead" region of similar size as the tiles into the reporter. However, because we are interested in the relative luciferase activity between the cloned tiles (independently of the empty vector), generating this extra control was not prioritized.

Finally, the luciferase assays used in this thesis mostly relied on the interaction between exogenous DNA (plasmid with enhancer elements controlling the expression of the luciferase reporter gene) and *endogenous* TFs that bind to and activate or repress the enhancer[133]. In contrast, luciferase assays are most often used to assess the interaction between two or more exogenous factors (reporter plasmid and a co-expressed TF). Here, the latter approach was only possible after initial screening in the endogenous context, which allowed the identification of putative TFs (such as CEBP-β), that later were used for co-expression. Thus, the endogenous context may play an important role in determining the luciferase activity of the various tiles. Therefore, the choice of cell type may have a strong influence on the observed luciferase activity.

## 6.3 Evaluation of cell lines

Because this thesis investigated the genetic effects of visceral obesity, and because the epigenetic data suggested that the enhancer in i.e. *HMBS* was only active in mesenchymal stem cells (precursors of fat cells), the ideal cells to evaluate luciferase activity in would have

been primary, mesenchymal stem cells from human visceral fat. However, this cell model was not available as it would require fresh collection of VAT from an obese patient undergoing i.e. bariatric surgery, isolation of the stromovascular fraction, fluorescence-activated cell sorting (FACS) based on unique surface markers and cultivation for maximum a few passages[134,135]. Often cell lines or more specifically immortalized cell lines are therefore utilized in research to assess biological processes *in vitro*, instead of primary cells[136]. Even though primary cells resemble the cells within a living tissue, it has been shown that culturing mature adipocytes is technically challenging, since they rapidly dedifferentiate into fibroblast-like cells[137]. Moreover, since mature adipocytes are highly differentiated and thus unable to divide, they cannot be maintained and expanded in culture. Thus, in this study as mentioned earlier (section 4.4.1), we used three cell lines: COS-1, ME3 and 3T3-L1 for the luciferase assays. There are several advantages associated with immortal cell lines include low-priced, easy to use and they provide a pure cell population, which is very important for reproducible results [136].

COS-1 cells are kidney cells and generated from epithelial cells. In contrast, ME3 are beige pre-adipocytes (the most used cell line in this thesis) and 3T3-L1 cells are white pre-adipocytes, both cell lines were generated from mouse embryonic fibroblast cells. COS-1 cell lines are very easy to cultivate and work with, have very high efficiency for transfection[138] and are routinely used for luciferase assays. These cells were used as a non-adipogenic, "control" line. ME3 cells are harder to transfect compared to COS-1 cells but compensates it with very high metabolic activates[101]. Moreover, being isolated from mouse embryonic fibroblasts, they represent a more mesenchymal-like cell type. Even though 3T3-L1 cell are widely used in obesity research, the cell line has very low transcriptions activity and very difficult to transfect/very low efficiency for transfection [139]. The cells also have multiple copies of the genome, thus representing a cell line further from primary cells compared to ME3 cells. In this thesis, we decided to initially use all three cell lines to assess whether the luciferase activity was dependent on the endogenous context or not, as further discussed below.

## 6.4 Comparison of luciferase activity by luciferase constructs

We compared transcriptional activity of the successfully cloned tiles as presented in figure 5.12 by conducting luciferase assay on transfected COS-1, ME3, and 3T3-L1 cells. We

observed higher transcriptional activity of tile 3 luciferase constructs in ME3 cells compared to the other cell lines. This might be due to the variables that effects transfection efficiency (discussed earlier). However, in general tile 3 luciferase constructs had the highest luciferase activity in each cell lines compared to other luciferase constructs. Since each luciferase construct (table 5.2) had regulatory sequence upstream reporter/luciferase gene, thus when the promoter is activated mRNA is transcribed and translated into luciferase protein. Therefore, amount of luciferase signal is dependent on activity of the regulatory element. One possible explanation for such high luciferase activity could be that within tile 3 way more SNPs (12 SNPs, including tag-SNP rs1784461) are located compared to the other tiles. Furthermore, as presented in figure 5.12, WashU Epigenomic Browser data showed in all three cell types (E025-adipose mesenchymal cells, E063-adipose nuclei, and E023 adipose differentiated from mesenchymal cells) very strong TSS (transcription start sites) and RefSeq data shows the *VPS11* gene TSS (exon) is located in tile 3. Also, ENCODE, cCREs, GH Reg Elems (DE), and OregAnno data suggested strong promoter and TBS (transcription binding sites) within tile 3, thus this can be an indication towards the idea that most likely both cis-regulatory and trans-regulatory elements are present in tile 3, which might regulate expression of VAT associated *VPS11* gene. However, the signals between risk and protective tile 3 luciferase constructs did not exhibit huge differences in the luciferase activity, suggesting that the elevated activity was independent of the risk of VAT development. This interpretation is strengthened by the fact that tile 3 demonstrated high, genotype-independent enhancer activity in all three cell lines, thus likely affecting a VAT-independent biological process.

In the ME3 cell line, tile 8 risk luciferase construct had quite high luciferase transcriptional activity compared to protective luciferase construct. In contrast, in COS-1 and 3T3-L1 cells, the luciferase activity was quite low, which can be due to the transfection efficiency or technical issues related to luciferase assay. However, in all the cell lines the signaling pattern of tile 8 constructs are quite similar, which can be an indication that tile 8 protective construct has more repressive effect, while tile 8 risk type construct has more activating effect. It may mean that the tile is responsive to TFs present in several cell types, while only available in mesenchymal stem cells due to the epigenetic signature. Regardless, the risk/protective ratio for tile 8 was most pronounced in the ME3 cells.

As for tile 1 luciferase constructs, they had very low transcriptional activity compared to the other tiles, as supported by the lack of epigenetic enhancer marks in this area and presence of only two SNPs, while tile 8 contained seven SNPs including rs1799993. As mentioned earlier, the Karlsson et al. study suggested that rs1799993 had promoter activity[24], but our investigated data from UCSC Genome Browser and WashU Epigenomic browser (figure 5.12) suggested that tile 8 is more likely involved in enhancer activity, rather than promoter activity. Furthermore, when we compared risk/protective ratio (figure 5.13) of tile 8 compared to other tiles, both in ME3 and 3T3-L1 cell lines tile 8 had genotype-dependent regulatory/enhancer activity. In this study we couldn't clone all the remaining tiles, due to time limitation and misunderstanding (we thought tiles 2, 6 and 7 were successfully cloned by the previous study, while only tile 8a actually was), we only focused on the tiles that were easy to clone. Thus, we can't exclude the possibility that other tiles may harbor additional casual variant(s), although epigenomic data suggest the other tiles are unlikely to have enhancer activity.

## 6.2 Effect of individual SNPs on enhancer activity

In this study we were able to successfully generate the mutant/risk variants of the SNPs located in tile 8a (figure 5.14). Figure 5.18 shows the luciferase data of the five constructs including the protective construct. ME3 cell line had the highest luciferase signal for each mutant constructs and the luciferase activity was significantly different from the wild/protective type. Interestingly, for the other two cell lines, the signaling pattern was highly similar to ME3 cell line. These findings were surprising for two reasons; first the individual SNPs appeared to have a stronger effect alone than combined (in the long tile 8). Secondly, this effect was present even in COS-1 and 3T3-L1 cells that did not respond as strongly to the combined set of risk SNPs (figure 5.13). Thus, our finding suggested that regardless of cell types, mutated SNPs exhibit compound genotype-dependent cis-regulatory or enhancer activity in tile 8a. Moreover, the findings validate the previous claim that tile 8 protective type has a more repressive effect and risk type has an activating affect. As mentioned before, figure 1b from Karlsson et al. study showed the promoter activity of rs1799993 where risk/protective ratio was around 1.5-fold, while the ratio for rs1799993 risk/protective constructs were 25, 15, and 15-fold respectively, which indicates more towards tile 8a being involved in genotype-dependent enhancer activity, rather than promoter activity. As claimed earlier, the findings further suggest that the longer *HMBS* gene isoforms

might be downstream regulatory gene of tile 8. However, if we compare the results of tile 8 constructs from previous section (6.4), one could expect much more higher luciferase signal, based on the results from figure 5.18. One possible explanation for the results could be that tile 8 additionally contains the SNP–rs1006195, which is missing in tile 8a. It is possible that the SNP–rs1006195 has more repressive effect than activating effect, when combined all the SNPs in tile 8 leads to less luciferase transcriptional activity. Importantly, the long tile 8 risk and protective tiles were cloned from 2 different patients, and although they possessed the same theoretical haplotype (the same set of SNPs in LD with the tag SNP), the middle part of the tile was not covered by sequencing. Thus, it is theoretically possible that sequence variation between the patients in that region could have influenced the result. In contrast, the single-SNP mutations in tile 8a were all generated from the same plasmid derived from a single patient. Due to time constrains, mutagenesis was only performed on tile 8a, but should ideally have been done on the long tile 8 as well.

## 6.6 Transcriptional binding motifs overlapping SNPs within tile 8a

Adipogenesis is a process that involves differentiation of preadipocytes into mature adipocytes and the process is divided into two major phases including: the commitment phase and terminal differentiation phase [140]. Adipogenesis is driven by complex signaling cascade involving key transcription factors –PPAR$\gamma$ and multiple members of CCAAT/enhancer-binding family of proteins (C/EBPs). During adipogenesis, the process initiates through expression and activation of C/EBP-$\beta$ and C/EBP-$\delta$, which then activates C/EBP-$\alpha$ and activates several genes required for adipogenesis including PPAR$\gamma$ (master regulator of adipogenesis)[140]. Once activated PPAR$\gamma$ and C/EBP-$\alpha$ stimulates each other's expression, thus forms a feedback loop, and activated PPAR$\gamma$ then activate all other genes required for adipogenesis and fat cell function[140].

A previous study by our lab suggested that the SNP rs2846282 risk variant introduces a binding site for C/EBP-$\beta$ TF (transcription factor). Increased binding of this critical adipogenic factor could explain the increased expansion of VAT in patients with the risk variant. In this study, we therefore functionally tested whether this particular SNP could affect C/EBP-$\beta$ mediated enhancer activity by performing luciferase assay. Surprisingly, as the figure 5.20 shows, co-expression with C/EBP-$\beta$) resulted in significantly lower luciferase activity compared to the empty plasmid. Moreover, this was true for most of the mutant tiles,

suggesting an indirect effect independent of the SNPs. Thus, our findings suggest that C/EBP-$\beta$ didn't have an activator effect and was not the mediator of the observed SNP-dependent activity. The nonspecific effect of the experiment could be due to operator error or outcome of non-specific effects of biological manipulations or cell treatments[141]. For instance, if overexpression of C/EBP-$\alpha$ affected protein translation in general, this would give the observed results. Importantly, the total amount of DNA was equal in all transfected wells (the same amount of empty plasmid was added to wells not receiving the C/EBP-$\alpha$ plasmid), so this could not explain the results.

TFs or sequence specific TFs regulate gene expression through binding to cis-regulatory region within promoter and enhancer sequences[142]. And the existence or affinity of TF binding sites are most likely affected by non-coding variants, which consequently, modify the activity of a regulatory region[74]. Results from section 5.4.5 showed that we were able to identify TFs binding sites (DNA-binding motifs) overlapping SNPs within tile 8a (figure 5.21) from UCSC Genome Browser and JASPAR CORE 2022. Furthermore, we observed from our finding differences in the presence of TF binding sites when comparing the risk and protective haplotype, table 5.5. These TF binding sites were either lost, strengthen, or weakened in a genotype-dependent manner. TFs bind to very specific binding motifs; thus, a single critical nucleotide modification can lead to dramatic changes in the binding affinity of the TFs to the DNA sequence [142]. For example, the three TFs including RARB, RARG, and RXRB bind only to the protective, but not risk variant sequence of the SNP rs1799993. The three TFs can act as repressors, consequently loss of repressor activity and thereby increased enhancer activity may occur in the risk variant of rs1799993. Likewise, for rs11297272, the repressor ZNF354A only binds to protective sequence. Other TFs, including Zfx, ZBTB14, Wt1, and ZNF384 bound to both the risk and the protective sequence of either rs1799991, rs2846282, or rs11297272, thereby likely not contributing to the genotype-dependent effect.

Interestingly, we identified TF Wt1 which during development only expressed in the intermediate mesoderm and is the major regulator of mesenchymal progenitor in the developing kidney and heart[143]. Wt1 binds to both the risk and protective haplotype of SNP rs2846282. Studies suggested Wt1 to be a VAT specific TF that is only expressed in visceral fat depots, but not in subcutaneous or BAT in mice and humans[143,144]. On average, human visceral adipose derived stem cells (ACSs) selectively express Wt1 24-fold more compared to

subcutaneous ASCs[144]. Wt1 has binding site in the promoter region of the human ALDH1A2 and activates the gene, which then leads to production of dehydrogenase enzyme (RDH10 and DHRS3). The dehydrogenase enzyme than oxidizes retinal into (all-trans retinoic acid) RA that has been shown to inhibit adipogenesis[143,144]. RA is the vitamin A metabolite and regulates gene expression by activating nuclear hormone receptors (superfamily of transcription factors)[145]. In agreement, the increased expression of RA-metabolizing genes in VAT is accompanied by higher endogenous RA levels in VAT. Interestingly, increase RA levels leads to activation of RA target RAR family (RARα, RARβ and RARγ) or retinol X receptors (RXRα, RXRβ, and RXRγ). Moreover, out of these six genes, RARB expression is specifically upregulated in visceral ACSs and expression of RARB is inducible by RA[144]. Early stage of adipogenesis of visceral ASCs is inhibited by the excessive RA-mediated activity. A Study by Schwarz et al. reported that RA prevents adipogenesis in mice by preventing C/EBP-$\beta$ induced adipocytes differentiation and PPARγ induction[146]. Inhibition of adipogenesis in human ASCs can be reversed by the antagonism of RA receptors, as well as knockdown of Wt1, which leads to dramatic suppression of ALDH1A2 gene, consequently significant increase in adipogenesis[144]. RXR/RAR heterodimers bind to RAREs (the retinoic acid response elements), which is composed of tandem 5'-AGGTCA-3' sites known as DR1-DR5. RARB can act as repressor when binding to DR1 sites (consensus AGGGCAgAGGTCA) and acts as an activator when bound to DR5 sites (GGTTCAccgaaAGTTCA)[27]. The binding site overlapping rs1799993 is more closely resembling a repressive DR1 site.

Therefore, in visceral obesity risk patients, the increased luciferase enhancer activity in tile 8a with the rs1799993 risk variant may be explained by loss of the RARB/RXR repressor binding site, leading to loss of the anti-adipogenic RA-RARB/RXR signaling and stimulation of visceral adipogenesis. In other words, our findings may have unraveled a potential causal mechanism underlying the association between a risk haplotype in the 11q23.3 locus and elevated risk of increased VAT mass.

# 7. Conclusions

The main goal of this study was to investigate genotype-dependent enhancer activities in the newly discovered VAT associated 11q23.3 locus and identify the causal variant(s) with cis-regulatory activity.

A total of 14 tiles covering 5 major parts of the locus was designed. Using in-fusion cloning, 3 of the 5 regions were successfully cloned from patients homozygous for either the risk or protective haplotype into a luciferase enhancer reporter. Luciferase assays revealed that overall, tile 3 had the highest enhancer activity, while tile 8 demonstrated the strongest genotype-dependent enhancer activity.

A shorter version of tile 8 was found to harbor 5 SNPs, and individual site-directed mutagenesis of each SNP demonstrated that all of these SNPs, including rs2846282 and rs1799993, contributed to the genotype-dependent enhancer activity and can be considered likely causal variants.

While functional luciferase assays showed that the effect of rs2846282 was not mediated via increased binding of C/EBP-β, *in silico* analyses suggested loss of binding of anti-adipogenic RARB/RXR specifically in the risk variant of rs1799993.

Overall, this thesis has strengthened rs1799993 as a likely causal SNP underlying the association between 11q23.3 risk haploblock and increased VAT mass. Additionally, four other SNPs in close vicinity have been identified as likely additional causal contributors.

# 8. Future perspective

Many experiments are needed to find the causal variant(s) and their downstream target gene(s). In this study we were able to only scratch the surface of this vast and complicated matter in the given time. Thus, much more extensive work is needed to understand the underlying mechanisms that may explain selective fat storage in VAT and diseases associated with VAT.

The findings from our study hints at which regions of the 11q23.3 locus might have enhancer activity, however there are still many unanswered questions. For the future, first focus should be designing appropriate primers and amplifying remaining tiles including tile 2, 4, 5, 6 and 7. Also, important step is optimizing PCR conditions, mostly annealing temperature, and touchdown PCR, before testing the amplified product on risk and protective haplotypes, and if it is possible additional patient samples than the four patient samples we tested in this study. After amplifying the right DNA fragment from the risk and protective haplotype, the tiles should be cloned into the linearized pGL4.23_*Luc-2*_minP vector by in-fusion cloning. Furthermore, Sanger sequencing should be to verify that the cloned plasmids contain the right insert. Dual luciferase assay should be performed on each luciferase construct to determine luciferase activity more accurately[147].

Site-directed mutagenesis must be performed in all the SNPs within the long tile 8 and luciferase assays need to be done to assess genotype-dependent enhancer activity. Moreover, if additional tiles show genotype-dependent enhancer activity, site-directed mutagenesis should be performed for all SNPs located in all those tiles, to assess the specific genotype-dependent luciferase activity. Sanger-sequencing must be performed for all these constructs to verify the right insert.

In this thesis, we have identified an RXR/RAR heterodimer binding site that overlaps with the rs1799993 protective, but not risk variant. We hypothesized that this leads to loss of repressor activity[148]. However, this theory must be experimentally tested by overexpression of RARB and/or RXR together with tile 8a protective and tile 8a rs1799993 risk constructs.

# 9. References


1. Guh, D. P. *et al.* The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis. *BMC Public Health* **9**, 1–20 (2009).
2. Pflanz, C.-P. *et al.* Central obesity is selectively associated with cerebral gray matter atrophy in 15,634 subjects in the UK Biobank. *International Journal of Obesity 2022* 1–9 (2022) doi:10.1038/s41366-021-00992-2.
3. Obesity and overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.
4. The Lancet Gastroenterology & Hepatology. Obesity: another ongoing pandemic. *The Lancet Gastroenterology & Hepatology* **6**, 411 (2021).
5. Overweight and obesity in Norway - NIPH. https://www.fhi.no/en/op/hin/health-disease/overweight-and-obesity-in-norway---/.
6. Panuganti, K. K., Nguyen, M. & Kshirsagar, R. K. Obesity. *Antenatal Disorders for the MRCOG and Beyond* 135–138 (2021).
7. Peeters, A. Obesity and the future of food policies that promote healthy diets. *Nature Reviews Endocrinology 2018 14:7* **14**, 430–437 (2018).
8. Australian Dietary Guidelines | NHMRC. https://www.nhmrc.gov.au/adg.
9. Dietry Guidlines for Americans 2015-2020. U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015–2020. *Dietary Guidelines for Americans. 8th Edition.* (2015).
10. Hobbs, M. & Radley, D. Obesogenic environments and obesity: A comment on "Are environmental area characteristics at birth associated with overweight and obesity in school-aged children? Findings from the SLOPE (Studying Lifecourse Obesity PrEdictors) population-based cohort in the south of England." *BMC Medicine* **18**, 1–3 (2020).
11. Hruby, A. & Hu, F. B. The Epidemiology of Obesity: A Big Picture. *Pharmacoeconomics* **33**, 673 (2015).
12. Maes, H. H. M., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**, 325–351 (1997).
13. Zaitlen, N. *et al.* Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLOS Genetics* **9**, e1003520 (2013).
14. Silventoinen, K., Magnusson, P. K. E., Tynelius, P., Kaprio, J. & Rasmussen, F. Heritability of body size and muscle strength in young adulthood: a study of one million Swedish men. *Genet Epidemiol* **32**, 341–349 (2008).
15. Elks, C. E. *et al.* Variability in the Heritability of Body Mass Index: A Systematic Review and Meta-Regression. *Frontiers in Endocrinology* **3**, (2012).
16. Kilpeläinen, T. O. *et al.* Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Med* **8**, (2011).
17. Andreasen, C. H. *et al.* Low Physical Activity Accentuates the Effect of the FTO rs9939609 Polymorphism on Body Fat Accumulation. *Diabetes* **57**, 95–101 (2008).
18. Cooper, J. H. F., Collins, B. E. G., Adams, D. R., Robergs, R. A. & Donges, C. E. Limited Effects of Endurance or Interval Training on Visceral Adipose Tissue and Systemic Inflammation in Sedentary Middle-Aged Men. *Journal of Obesity* **2016**, (2016).

19. Obesity : preventing and managing the global epidemic : report of a WHO Consultation on Obesity, Geneva, 3-5 June 1997. https://apps.who.int/iris/handle/10665/63854.
20. Rothman, K. J. BMI-related errors in the measurement of obesity. *International Journal of Obesity 2008 32:3* **32**, S56–S59 (2008).
21. Madden, A. M. & Smith, S. Body composition and morphological assessment of nutritional status in adults: a review of anthropometric variables. *Journal of Human Nutrition and Dietetics* **29**, 7–25 (2016).
22. Duren, D. L. *et al.* Body Composition Methods: Comparisons and Interpretation. *Journal of diabetes science and technology (Online)* **2**, 1139 (2008).
23. Veghari, G., Salehi, A. & Vaghari, M. The comparison of waist circumference, waist-to-hip ratio, and waist-to-height ratio among rural women adults in the North of Iran, between the years 2004 and 2013. *ARYA Atherosclerosis* **14**, 169 (2018).
24. Karlsson, T. *et al.* Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. *Nature Medicine 2019 25:9* **25**, 1390–1395 (2019).
25. Sakers, A., de Siqueira, M. K., Seale, P. & Villanueva, C. J. Adipose-tissue plasticity in health and disease. *Cell* **185**, 419–446 (2022).
26. Lee, Y. H., Mottillo, E. P. & Granneman, J. G. Adipose tissue plasticity from WAT to BAT and in between. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1842**, 358–369 (2014).
27. Sethi, J. K. & Vidal-Puig, A. J. Thematic review series: adipocyte biology. Adipose tissue function and plasticity orchestrate nutritional adaptation. *J Lipid Res* **48**, 1253–1262 (2007).
28. Funcke, J. B. & Scherer, P. E. Beyond adiponectin and leptin: adipose tissue-derived mediators of inter-organ communication. *J Lipid Res* **60**, 1648–1697 (2019).
29. Sethi, J. K. & Vidal-Puig, A. J. Thematic review series: Adipocyte Biology. Adipose tissue function and plasticity orchestrate nutritional adaptation. *J Lipid Res* **48**, 1253 (2007).
30. Zwick, R. K., Guerrero-Juarez, C. F., Horsley, V. & Plikus, M. v. Anatomical, Physiological, and Functional Diversity of Adipose Tissue. *Cell Metabolism* **27**, 68–83 (2018).
31. Bolsoni-Lopes, A. & Alonso-Vale, M. I. C. Lipolysis and lipases in white adipose tissue - An update. *Arch Endocrinol Metab* **59**, 335–342 (2015).
32. Song, Z., Xiaoli, A. M. & Yang, F. Regulation and Metabolic Significance of De Novo Lipogenesis in Adipose Tissues. *Nutrients* **10**, (2018).
33. Rosen, E. D. & Spiegelman, B. M. What we talk about when we talk about fat. *Cell* **156**, 20–44 (2014).
34. Angueira, A. R. *et al.* Early B Cell Factor Activity Controls Developmental and Adaptive Thermogenic Gene Programming in Adipocytes. (2020) doi:10.1016/j.celrep.2020.02.023.
35. Ouellet, V. *et al.* Brown adipose tissue oxidative metabolism contributes to energy expenditure during acute cold exposure in humans. *J Clin Invest* **122**, 545–552 (2012).
36. Fernández-Verdejo, R., Marlatt, K. L., Ravussin, E. & Galgani, J. E. Contribution of brown adipose tissue to human energy metabolism. *Mol Aspects Med* **68**, 82 (2019).

37. Lee, M. J., Wu, Y. & Fried, S. K. Adipose tissue heterogeneity: Implication of depot differences in adipose tissue for obesity complications. *Molecular Aspects of Medicine* **34**, 1–11 (2013).
38. Tran, T. T., Yamamoto, Y., Gesta, S. & Kahn, C. R. Beneficial effects of subcutaneous fat transplantation on metabolism. *Cell Metab* **7**, 410–420 (2008).
39. Shuster, A., Patlas, M., Pinthus, J. H. & Mourtzakis, M. The clinical importance of visceral adiposity: a critical review of methods for visceral adipose tissue analysis. *Br J Radiol* **85**, 1–10 (2012).
40. Schleinitz, D., Böttcher, Y., Blüher, M. & Kovacs, P. The genetics of fat distribution. *Diabetologia* **57**, 1276–1286 (2014).
41. Donohoe, C. L., Lysaght, J., O'Sullivan, J. & Reynolds, J. v. Emerging Concepts Linking Obesity with the Hallmarks of Cancer. *Trends Endocrinol Metab* **28**, 46–62 (2017).
42. Cawley, J. *et al.* Direct medical costs of obesity in the United States and the most populous states. *Journal of Managed Care and Specialty Pharmacy* **27**, 354–366 (2021).
43. Consensus statements. https://www.idf.org/e-library/consensus-statements/60-idfconsensus-worldwide-definitionof-the-metabolic-syndrome.html.
44. Brahe, L. K., Astrup, A. & Larsen, L. H. Can We Prevent Obesity-Related Metabolic Diseases by Dietary Modulation of the Gut Microbiota? *Adv Nutr* **7**, 90–101 (2016).
45. Bischoff, S. C. & Schweinlin, A. Obesity therapy. *Clinical Nutrition ESPEN* **38**, 9–18 (2020).
46. Apovian, C., Aronne, L., … D. B.-T. J. of & 2015, undefined. Pharmacological management of obesity: an endocrine Society clinical practice guideline. *academic.oup.com*.
47. Hjelmesæth, J., Lund, R. S., Sagen, J. V. & Valderhaug, T. G. Vektreduserende medisiner – for hvem, hvordan, hvor lenge? *Tidsskrift for Den norske legeforening* (2022) doi:10.4045/TIDSSKR.22.0115.
48. Bray, G. A., Frühbeck, G., Ryan, D. H. & Wilding, J. P. H. Management of obesity. *Lancet* **387**, 1947–1956 (2016).
49. Bocarsly, M. E. Pharmacological Interventions for Obesity: Current and Future Targets. *Current Addiction Reports* **5**, 202–211 (2018).
50. Apovian, C., Clinics, N. I.-E. and M. & 2016, undefined. Obesity: guidelines, best practices, new research. *endo.theclinics.com*.
51. Brashier, D., Sharma, A., … N. D.-… of pharmacology & & 2014, undefined. Lorcaserin: A novel antiobesity drug. *ncbi.nlm.nih.gov*.
52. Secher, A. *et al.* The arcuate nucleus mediates GLP-1 receptor agonist liraglutide-dependent weight loss. *J Clin Invest* **124**, 4473–4488 (2014).
53. FDA Approves New Drug Treatment for Chronic Weight Management, First Since 2014 | FDA. https://www.fda.gov/news-events/press-announcements/fda-approves-new-drug-treatment-chronic-weight-management-first-2014.
54. Varnado, P. J. *et al.* Prevalence of binge eating disorder in obese adults seeking weight loss treatment. *Eat Weight Disord* **2**, 117–123 (1997).
55. Yang, P. *et al.* Long-term outcomes of laparoscopic sleeve gastrectomy versus Roux-en-Y gastric bypass for morbid obesity: Results from a meta-analysis of randomized controlled trials. *Surgery for Obesity and Related Diseases* **15**, 546–555 (2019).
56. Chapmon, K., Stoklassa, C. J. & Benson-Davies, S. Nutrition for pregnancy after metabolic and bariatric surgery: literature review and practical guide.

*Surgery for Obesity and Related Diseases* (2022) doi:10.1016/J.SOARD.2022.02.019.
57. Hajer, G. R., van Haeften, T. W. & Visseren, F. L. J. Adipose tissue dysfunction in obesity, diabetes, and vascular diseases. *European Heart Journal* **29**, 2959–2971 (2008).
58. Elks, C. E. *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne)* **3**, (2012).
59. Silventoinen, K., Rokholm, B., Kaprio, J. & Sørensen, T. I. A. The genetic and environmental influences on childhood obesity: a systematic review of twin and adoption studies. *Int J Obes (Lond)* **34**, 29–40 (2010).
60. Larsen, L. H. Obesity: Underlying Mechanisms and the Evolving Influence of Diet. *Current Nutrition Reports* **1**, 205–214 (2012).
61. Vos, N. *et al.* Bariatric Surgery for Monogenic Non-syndromic and Syndromic Obesity Disorders. *Current Diabetes Reports* **20**, 1–10 (2020).
62. Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nature Reviews Genetics 2021 23:2* **23**, 120–133 (2021).
63. Barroso, I. & McCarthy, M. I. The Genetic Basis of Metabolic Disease. *Cell* **177**, 146–161 (2019).
64. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
65. Albuquerque, D., Nóbrega, C., Manco, L. & Padez, C. The contribution of genetics and environment to obesity. *British Medical Bulletin* **123**, 159–173 (2017).
66. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans . *New England Journal of Medicine* **373**, 895–907 (2015).
67. Genetic Variation and Disease: GWAS | Learn Science at Scitable. https://www.nature.com/scitable/topicpage/genetic-variation-and-disease-gwas-682/#:~:text=GWAS%20seek%20to%20identify%20the,are%20distributed%20across%20different%20populations.
68. Belmont, J. W. *et al.* The International HapMap Project. *Nature 2004 426:6968* **426**, 789–796 (2003).
69. Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature 2020 577:7789* **577**, 179–189 (2020).
70. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics 2019 20:8* **20**, 467–484 (2019).
71. Al-Barghouthi, B. M. *et al.* Transcriptome-wide Association Study and eQTL colocalization identify potentially causal genes responsible for bone mineral density GWAS associations. *bioRxiv* 2021.10.12.464046 (2022) doi:10.1101/2021.10.12.464046.
72. Barroso, I. & McCarthy, M. I. The genetic basis of metabolic disease. *Cell* **177**, 146 (2019).
73. Yang, S. *et al.* Prioritizing genetic variants in GWAS with lasso using permutation-assisted tuning. *Bioinformatics* **36**, 3811–3817 (2020).
74. Broekema, R. v., Bakker, O. B. & Jonkers, I. H. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biology* **10**, (2020).
75. Herman, M. A. & Rosen, E. D. Making Biological Sense of GWAS Data: Lessons from the FTO Locus. *Cell Metab* **22**, 538–539 (2015).
76. Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature 2012 491:7422* **491**, 56–65 (2012).

77. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nature Reviews Genetics 2019 20:8* **20**, 437–455 (2019).
78. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *American Journal of Human Genetics* **102**, 717 (2018).
79. Tehranchi, A. *et al.* Fine-mapping cis-regulatory variants in diverse human populations. *Elife* **8**, (2019).
80. Lawrenson, K. *et al.* Cis-eQTL analysis and functional validation of candidate susceptibility genes for high-grade serous ovarian cancer. *Nature Communications 2015 6:1* **6**, 1–14 (2015).
81. Verdugo, R. A., Farber, C. R., Warden, C. H. & Medrano, J. F. Serious limitations of the QTL/Microarray approach for QTL gene discovery. *BMC Biology* **8**, 1–17 (2010).
82. Ohnmacht, J., May, P., Sinkkonen, L. & Krüger, R. Missing heritability in Parkinson's disease: the emerging role of non-coding genetic variation. *Journal of Neural Transmission* **127**, 729–748 (2020).
83. Wittkopp, P. J. & Kalay, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics 2011 13:1* **13**, 59–69 (2011).
84. Battle, A. & Montgomery, S. B. Determining causality and consequence of expression quantitative trait loci. *Hum Genet* **133**, 727 (2014).
85. Claussnitzer, M. *et al.* Leveraging Cross-Species Transcription Factor Binding Site Patterns: From Diabetes Risk Loci to Disease Mechanisms. *Cell* **156**, 343 (2014).
86. Moreno-Navarrete, J. M. *et al.* Heme Biosynthetic Pathway is Functionally Linked to Adipogenesis via Mitochondrial Respiratory Activity. *Obesity* **25**, 1723–1733 (2017).
87. Song, G. *et al.* Structural insight into acute intermittent porphyria. *The FASEB Journal* **23**, 396–404 (2009).
88. Presentation of master's thesis Nina T Samuelsen: Functional dissection of a genetic locus for visceral fat mass (11q23.3) | Department of Biological Sciences (BIO) | UiB. https://www.uib.no/en/bio/145730/presentation-masters-thesis-nina-t-samuelsen-functional-dissection-genetic-locus-visceral.
89. pGL4.23[luc2/minP] Vector Protocol. https://no.promega.com/resources/protocols/product-information-sheets/a/pgl423-vector-protocol/.
90. Buckhout-White, S., Person, C., Medintz, I. L. & Goldman, E. R. Restriction enzymes as a target for DNA-based sensing and structural rearrangement. *ACS Omega* **3**, 495–502 (2018).
91. Lee, P. Y., Costumbrado, J., Hsu, C. Y. & Kim, Y. H. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments : JoVE* 3923 (2012) doi:10.3791/3923.
92. GelRed® Nucleic Acid Gel Stain - Biotium. https://biotium.com/product/gelred-nucleic-acid-gel-stain/.
93. Genomic DNA Extraction and Genomic DNA Isolation | Thermo Fisher Scientific - NO. https://www.thermofisher.com/no/en/home/life-science/dna-rna-purification-analysis/genomic-dna-extraction.html?s_kwcid=AL!3652!3!585290755617!p!!g!!dna%20purification&ef_id=CjwKCAjw9LSSBhBsEiwAKtf0n3K2JXGrdeBGmIJQQPAKjhgJXr7iNnzBeu6nvnCsrMVQDfZBW9peUxoCXLcQAvD_BwE:G:s&s_kwcid=AL!3652!3!585290755617!p!!g!!dna%20purification&cid=bid_sap_rst_r01_co_cp136

2_pjt0000_bid00000_0se_gaw_nt_awa_con&gclid=CjwKCAjw9LSSBhBsEiwAKtf0n3K2JXGrdeBGmIJQQPAKjhgJXr7iNnzBeu6nvnCsrMVQDfZBW9peUxoCXLcQAvD_BwE.
94. In-Fusion Snap Assembly master mixes for seamless DNA cloning. https://www.takarabio.com/products/cloning/in-fusion-seamless-cloning/in-fusion-snap-assembly.
95. QuikChange II Site-Directed Mutagenesis Kits - Details & Specifications | Agilent. https://www.agilent.com/en/quikchange-ii-site-directed-mutagenesis-kits-details.
96. Bacterial Transformation and Competent Cells–A Brief Introduction | Thermo Fisher Scientific - NO. https://www.thermofisher.com/no/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/molecular-cloning/transformation/competent-cell-basics.html.
97. Bergkessel, M. & Guthrie, C. Colony PCR. *Methods in Enzymology* **529**, 299–309 (2013).
98. Tan, S. C. & Yiap, B. C. DNA, RNA, and protein extraction: The past and the present. *Journal of Biomedicine and Biotechnology* **2009**, (2009).
99. Sanger Sequencing Steps & Method. https://www.sigmaaldrich.com/NO/en/technical-documents/protocol/genomics/sequencing/sanger-sequencing.
100. Hancock, J. F. COS Cell Expression. *Methods Mol Biol* **8**, 153–158 (1992).
101. Bjune, J. I. *et al.* The homeobox factor Irx3 maintains adipogenic identity. *Metabolism* **103**, 154014 (2020).
102. Lukas, J., Bartkova, J., Rohde, M., Strauss, M. & Bartek, J. Cyclin D1 Is Dispensable for G 1 Control in Retinoblastoma Gene-Deficient Cells Independently of cdk4 Activity. *MOLECULAR AND CELLULAR BIOLOGY* **15**, 2600–2611 (1995).
103. Green, H. & Meuth, M. An established pre-adipose cell line and its differentiation in culture. *Cell* **3**, 127–133 (1974).
104. Morrison, S. & Mcgee, S. L. 3T3-L1 adipocytes display phenotypic characteristics of multiple adipocyte lineages. (2015) doi:10.1080/21623945.2015.1040612.
105. Morrison, S. & McGee, S. L. 3T3-L1 adipocytes display phenotypic characteristics of multiple adipocyte lineages. *Adipocyte* **4**, 295 (2015).
106. Aseptic Technique | Thermo Fisher Scientific - NO. https://www.thermofisher.com/no/en/home/references/gibco-cell-culture-basics/aseptic-technique.html.
107. CELL CULTURE BASICS Handbook Cell Culture Basics Cell Culture Basics Cell Culture Basics | i.
108. Trypan Blue Stain (0.4%) for use with the Countess™ Automated Cell Counter. https://www.thermofisher.com/order/catalog/product/T10282.
109. Cell Thawing Protocol for Frozen Primary Cells | STEMCELL Technologies. https://www.stemcell.com/how-to-thaw-frozen-primary-cells.html.
110. Carter, M. & Shieh, J. Biochemical Assays and Intracellular Signaling. *Guide to Research Techniques in Neuroscience* 311–343 (2015) doi:10.1016/B978-0-12-800511-8.00015-0.
111. North, M. *et al.* Comparison of ZFY and ZFX gene structure and analysis of alternative 3′ untranslated regions of ZFY. *Nucleic Acids Research* **19**, 2579–2586 (1991).

112. Margolis, R. L. *et al.* CDNAs with long CAG trinucleotide repeats from human brain. *Human Genetics* **100**, 114–122 (1997).
113. Sobek-Klocke, I. *et al.* The human gene ZFP161 on 18p11.21-pter encodes a putative c-myc repressor and is homologous to murine Zfp161 (Chr 17) and Zfp161-rs1 (X Chr). *Genomics* **43**, 156–164 (1997).
114. Sharma, P. M., Bowman, M., Madden, S. L., Rauscher, F. J. & Sukumar, S. RNA editing in the Wilms' tumor susceptibility gene, WT1. *Genes and Development* **8**, 720–731 (1994).
115. ZNF354A. https://www.pharmgkb.org/gene/PA36383/link.
116. Hauksdottir, H., Farboud, B. & Privalsky, M. L. Retinoic acid receptors β and γ do not repress, but instead activate target gene transcription in both the absence and presence of hormone ligand. *Molecular Endocrinology* **17**, 373–385 (2003).
117. Gaudet, P., Livstone, M. S., Lewis, S. E. & Thomas, P. D. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Briefings in Bioinformatics* **12**, 449–462 (2011).
118. Zhang, Q. *et al.* Novel Common Variants Associated with Obesity and Type 2 Diabetes Detected Using a cFDR Method. *Scientific Reports 2017 7:1* **7**, 1–11 (2017).
119. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics* **11**, (2020).
120. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
121. Applications & Technologies for DNA synthesis Gene Synthesis Handbook.
122. Śpibida, M., Krawczyk, B., Olszewski, M. & Kur, J. Modified DNA polymerases for PCR troubleshooting. *Journal of Applied Genetics* **58**, 133–142 (2017).
123. PCR Troubleshooting Guide | Thermo Fisher Scientific - NO. https://www.thermofisher.com/no/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-troubleshooting.html.
124. Park, J., Throop, A. L. & LaBaer, J. Recombinational Cloning Using Gateway and In-Fusion Cloning Schemes. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **110**, 3.20.1 (2015).
125. Laboratories Inc, C. In-Fusion® HD Cloning Kit User Manual. (2011).
126. Botella, E. *et al.* High-resolution temporal analysis of global promoter activity in Bacillus subtilis. *Methods in Microbiology* **39**, 1–26 (2012).
127. Bauer, P. Luciferase Reporter Gene Assays. *Encyclopedia of Cancer* 2077–2081 (2011) doi:10.1007/978-3-642-16483-5_3430.
128. Raser, J. M. & O'Shea, E. K. Noise in Gene Expression: Origins, Consequences, and Control. *Science* **309**, 2010 (2005).
129. Urban, E. A. & Johnston, R. J. Buffering and Amplifying Transcriptional Noise During Cell Fate Specification. *Frontiers in Genetics* **9**, (2018).
130. Factors Influencing Transfection Efficiency | Thermo Fisher Scientific - NO. https://www.thermofisher.com/no/en/home/references/gibco-cell-culture-basics/transfection-basics/factors-influencing-transfection-efficiency.html.
131. McNabb, D. S., Reed, R. & Marciniak, R. A. Dual Luciferase Assay System for Rapid Assessment of Gene Expression in Saccharomyces cerevisiae. *Eukaryotic Cell* **4**, 1539 (2005).

132. Bioluminescent Reporter Genes. https://no.promega.com/resources/pubhub/enotes/bioluminescent-reporter-genes/.
133. Buckley, S. M. K. *et al.* In vivo bioimaging with tissue-specific transcription factor activated luciferase reporters. *Scientific Reports 2015 5:1* **5**, 1–13 (2015).
134. Helmy, M. A., Mohamed, A. F., Rasheed, H. M. & Fayad, A. I. A protocol for primary isolation and culture of adipose-derived stem cells and their phenotypic profile. *https://doi.org/10.1080/20905068.2020.1750863* **56**, 42–50 (2020).
135. Condé-Green, A. *et al.* Shift toward Mechanical Isolation of Adipose-derived Stromal Vascular Fraction: Review of Upcoming Techniques. *Plastic and Reconstructive Surgery Global Open* **4**, (2016).
136. Kaur, G. & Dufour, J. M. Cell lines: Valuable tools or useless artifacts. *Spermatogenesis* **2**, 1 (2012).
137. Lessard, J. *et al.* Characterization of Dedifferentiating Human Mature Adipocytes from the Visceral and Subcutaneous Fat Compartments: Fibroblast-Activation Protein Alpha and Dipeptidyl Peptidase 4 as Major Components of Matrix Remodeling. *PLOS ONE* **10**, e0122065 (2015).
138. Smith, S. L. & Shioda, T. Advantages of COS-1 monkey kidney epithelial cells as packaging host for small-volume production of high-quality recombinant lentiviruses. *Journal of Virological Methods* **157**, 47–54 (2009).
139. Orlicky, D. J. & Schaackt, J. Adenovirus transduction of 3T3-L1 cells. *Journal of Lipid Research* **42**, 460–466 (2001).
140. Muruganandan, S., Roman, A. A. & Sinal, C. J. Adipocyte differentiation of bone marrow-derived mesenchymal stem cells: Cross talk with the osteoblastogenic program. *Cellular and Molecular Life Sciences* **66**, 236–253 (2009).
141. Luciferase Reporters | Thermo Fisher Scientific - NO. https://www.thermofisher.com/no/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/luciferase-reporters.html.
142. Inukai, S., Kock, K. H. & Bulyk, M. L. Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev* **43**, 110–119 (2017).
143. Chau, Y. Y. *et al.* Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source. *Nat Cell Biol* **16**, 367 (2014).
144. Takeda, K. *et al.* Retinoic acid mediates visceral-specific adipogenic defects of human adipose-derived stem cells. *Diabetes* **65**, 1164–1178 (2016).
145. Berry, D. C. & Noy, N. All-trans-Retinoic Acid Represses Obesity and Insulin Resistance by Activating both Peroxisome Proliferation-Activated Receptor β/δ and Retinoic Acid Receptor. *Molecular and Cellular Biology* **29**, 3286 (2009).
146. Schwarz, E. J., Reginato, M. J., Shao, D., Krakow, S. L. & Lazar, M. A. Retinoic Acid Blocks Adipogenesis by Inhibiting C/EBP-Mediated Transcription. *MOLECULAR AND CELLULAR BIOLOGY* **17**, 1552–1561 (1997).
147. Alcaraz-Pérez, F., Mulero, V. & Cayuela, M. L. Application of the dual-luciferase reporter assay to the analysis of promoter activity in Zebrafish embryos. *BMC Biotechnology* **8**, 1–8 (2008).
148. Hauksdottir, H., Farboud, B. & Privalsky, M. L. Retinoic acid receptors β and γ do not repress, but instead activate target gene transcription in both the absence and presence of hormone ligand. *Molecular Endocrinology* **17**, 373–385 (2003).

# 10. Supplementary Data

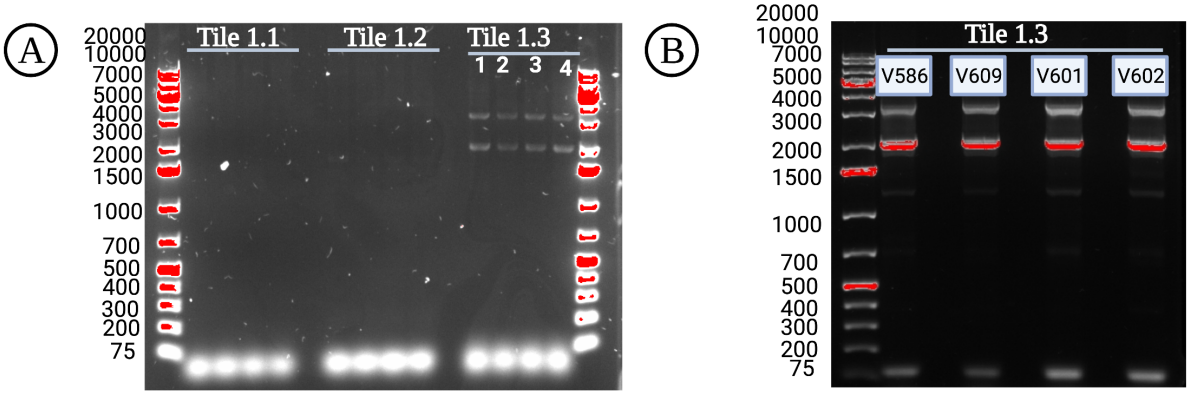



**Figure 10.1: Gel electrophoresis image of tile 1 tested on homozygote risk and protective haplotype. A)** shows PCR amplified products of tile 1,1, 1,2 and 1.3 respectively both on homozygote risk and protective patient. As the image shows there were no visible band for tile 1.1 and 1.2 DNA fragments at annealing temperature at 60°C. PCR amplified products of tile 1.3 had very weak band. **B)** represents PCR amplified products of tile 3.1 on homozygote risk and protective patient gDNA at annealing temperature at 65°C. All the fragments had right band size around 4000 bp, even though all the samples had very weak band around 2000 bp.

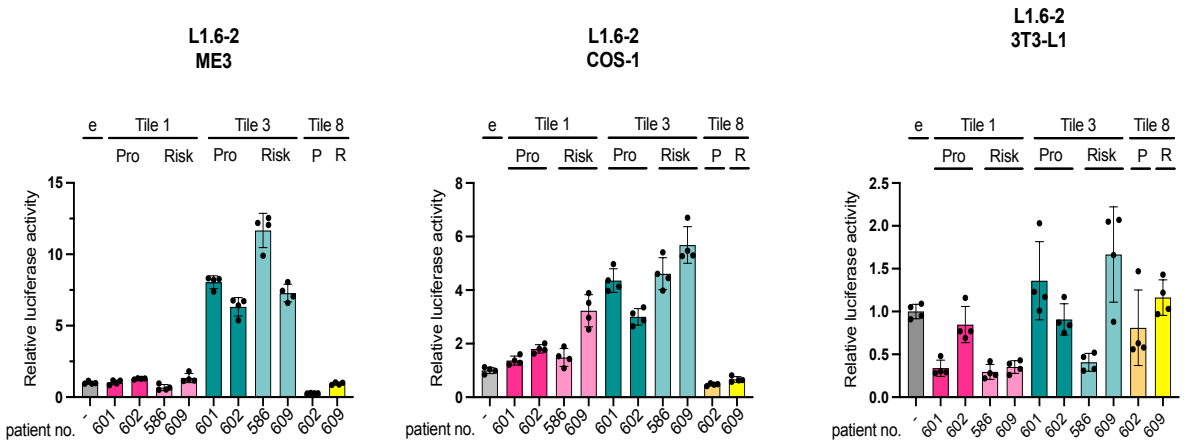



**Figure 10.2 Luciferase activity of tile 1, 3 and 8.** The image representing luciferase transcriptional activity of tile 1. 3 and 8 constructs of ME3, COS-1 and 3T3-L1 transfected cells. Signal pattern of the constructs are similar to signal patterns of the constructs from figure 5.13. ME3 cells had the highest luciferase activity compared to the other two cell lines.

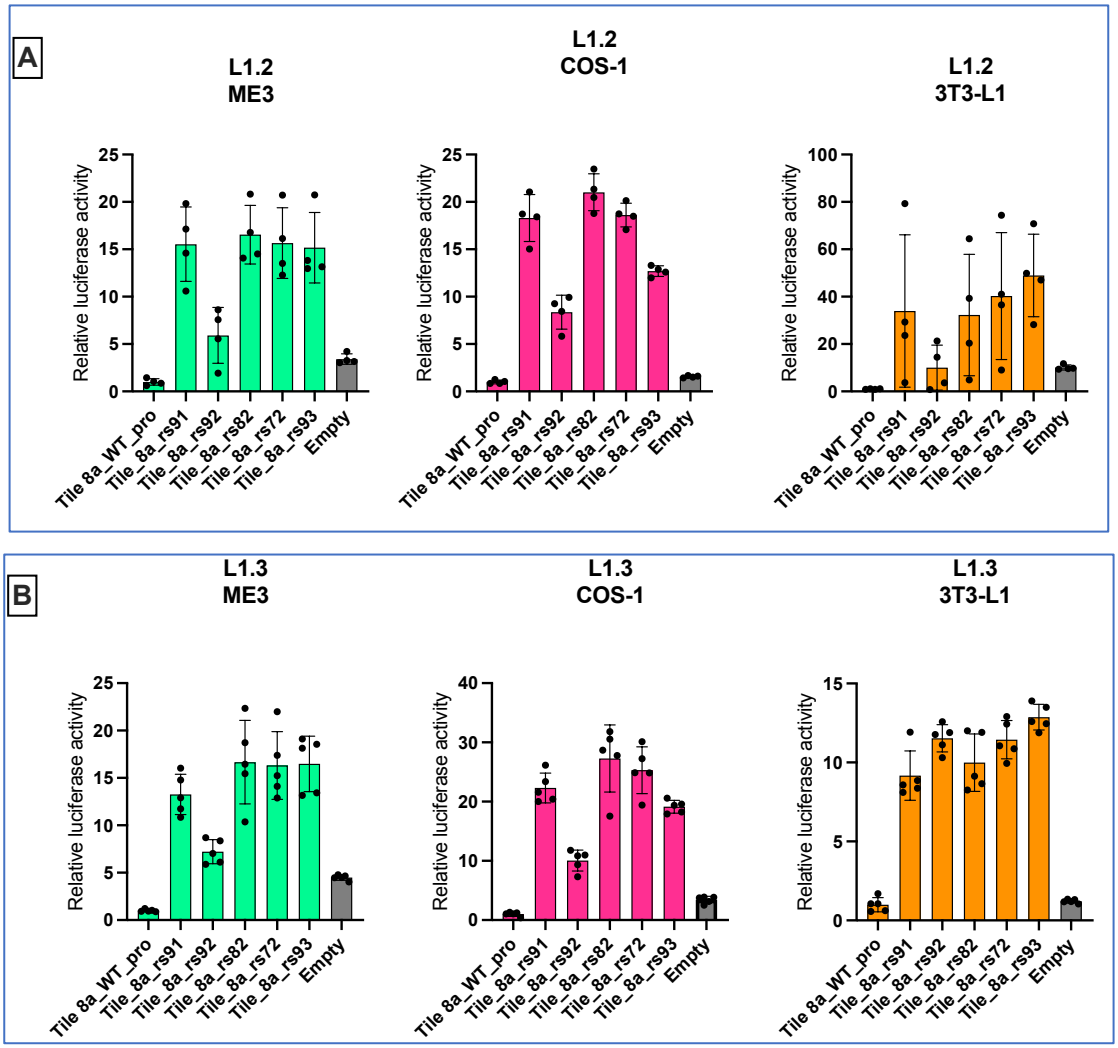



**Figure 10.3 Biological independent duplicates of luciferase activity of tile 8a protective and risk haplotype. A)** L1.2 and **B)** L1.3 represents luciferase data of the five mutant luciferase constructs, including pGL4.23_Tile 8a_WT and pGL4.23_Luc-2_minP_empty construct. All the data shown are normalized to the pGL4.23-Tile 8a_WT protective plasmid. Signal patterns of all the constructs within the cells are very similar in both experiments, except for the signals within 3T3-L1 cell lines of L1.2 experiment, where the signal was almost 30-fold higher than L1.3 experiment.

**Supplementary figure 10.4:**

Genomic sequence of all the tiles this study attempted to clone, including tile 1,3, 4,5 and 8. The sequences was retrieved from SnapGene viewer web tool. Marked in bold is the forward and reverse primers respectively.

**Tile 1.1 (3958 bp)**

**GGGGGAGGCGAAGAAAGAAA**ACACTTAAAACTGGATACCCGGAGTGAAGGAG
ACAGAATCACATCCCAGAACGGAGAGGCTGTAAGATTCATATCTACTTCATTCTT
ACCCAGGGCAGATCAGCCTACTTCTCCCCTTCTCCCTCCTGATGGGTACAAACCA

TTCTCTAATCATCCCAGCAGCGCCCACACAAAGGCCCAAGTGAAAGAGCATGGG
TCCTCGGCTTTGCCACATCCCTAAACAGAGAAGCTAACAGGACAAGAGAAAGGA
GACCCGAGCCTTTTGTCCACCACACAGGCAGTGAGTGGCGCACCGCAGCACTGT
GCCTGTTGTCAGACACCTGGGGTGCGGCCAGGGATCCCCGCCCAGGGCTCAACC
ACCTCCCCACACCTGGAGCTCCCCCGCTGACGGGGAAAGCCAGCTGGGCCTGGA
CGGGAAGAGAAAGGTAATGACGTTGGGGGTAGCTAGGATGAAGGAAAGGCCTA
TCTCCTCCTCGGCATCCCTGCGGGTTGTCCCGCCCCCTCCCCTTTCCCAGCTCACT
CCCCGGCCCCGCTGTGCCCACAGACGCAGTGCCAGGGGCGGTAAGGTTCCCCGC
CCGGCGGGTAGCCGTTACCCGGTGTTCACCTAAACCCCACCTGACGAAGGCGGC
GGCTCCCACTCCCGGAAACGGATCCCGGCCCGCCCCAGCAAGTGGGGACAGACC
GGGGGCCGAGCTGCGTGCCCCCCGGTGCGAGGCCGAGCGCCCGGCGTCCGCGCG
GCCCTCCCTTGCGCTCCCGACCCGCCCCCGGAGCTCGCACGCTGTCTCCCCTAGG
AGGGGTGGCCGGCCCGCACCTCCATCCACCCCCGCTACCTCTTGCCTCCGCTCCT
GCGGCCCCAGCTCTGGGACAGACGCGGCACGAACGTACCCACGAGGCCGGCAGC
GCCCCCCACCCGCGCCCTTCACAACTCCTCTCGGTTTGCAAACTGTTACATTAGC
CACCAACCTCTCGGCGGCGTCTCGCGCACCAGCCGGCCCCGGACGCGGCGCGCG
CTCATTGGAGCCTCGGCCGCCCGGCCCTGCGCTGCGCGCCCGGCCCCGCCCCCGC
CGCCCCGGCGCGCGTACTCATTGGACCACGTCCCAGGCCCCGCCCCCTCCCTTTG
CCAGCCTTGTAATGGGGCGGCGGAGGGGACCAGCCACCGGTTGGGCAGAAGCAG
CCGGCGTCGGCCTGGGATTGGACCACGTCACGGGGCCAGCGCGCTCCACTCCTTC
CTCCCCAGCGCCGGCCCCGGGGCGACCGCGGGACCTTTTCCGGCGCTGGATGCTT
TGCTGCCGGGGTTCGGGAGTGACCTGCGCCTAGAGCAGGTCGGGCGCACTGCAG
TCCCGCACTCGGGGCTCTTCTTCCGGTCACCTGTAGAGCCCGGCCGAGGAATGGA
ATAACAGGCGTCCTGTCCGCAGCCCCGCGGAGGGGGCGCTGGCCTGCCGGTTAG
TATTGAGGTGCCCTAGGCTTTGGATAGACGTTGGAGAATGGCTGTCCCTCGGACT
GAGTCCGTCGCAACCCCCTCTGTTTTACTGAGGAAACTGAAGCCACAGAAAAAG
ATGTTTGCTGATGGGTCAGAGATAACCGATGCATTCGTCGCAGAATCTGGCCAGG
ATTTCTTAGGTTAAGCCCCACTCATTTGTATAAGAAAGAATATTCCAAACTTAAG
TACTACACACAAAAGACACTTTTATTTGTGAGGCTATAGTTGGTTGAGCTTGTGT
GCACCTTCTCCTATTTCTGTTCTATTAAATCTTGTGAACATTTCTCTACTCTTTAGC
TGTTTCTGTCATAGAATGACAGAAGCACCTCACTGCCTTCTCTTTTAACTTACAGC
TCCATAGAAATGTGCGCAGTCGCCCGGGCGCGGTGGCTCACGCCTGTAATCCCA
GCATTTTGGGAGGCCAAGGCGGGCGAATCACTGGAGGTCAGGAGTTCGAGACCA
GCCTGGCCAACATGGTGAAACCCCGTCTCTACCAAAACTACAAAAATTAGCCAG

GCGTGGTGGCGGGTGCCTGTAATCCCAGCTACTCACGAGGCTGAGGCAGGAGAA
TTGCTTGAATCCGGGAGGCAGAGGTTGCTGTGAGCCCAGACTGCGCCACTGCACT
CCAGCCTGGGCGACAACAGCGAGACTCCTTCTCGAAAAAGAAATGTACGCAGTC
CTTGGAGACTGCATTTTCCTTAAGTGTGGATGGAAATAAGTCATCTGTTTTCTCTC
CTACTAAGTTGATGCCTGGCCACCCTTCACAGAAAGGCCCGCTTGAAGACCCTGG
CTGTCAGGGCCAGGGTTGTGTCAAGCCTGAACTTGTTTAAGGAGCTCTTAGCTCT
TCCTTTCTCTACATTTTTCGTGATGCTTAGGTGGAGAGTGGTCCGGTGTCTATGCT
TAGAAGGAAAAATTATTGGCAGTCCCAGAGAGAGTCTATTCTTCCTTTTTTTTTTT
TTTTTTTTTTTATTGATCATTCTTGGGTGTTTCTGGCAGAGGGGGATTTGGCAGGG
TCATAGGACAATAGTGGAGGGAAGGTCAGCAGATAAACAAGTGAACAAAGGTC
TCTGGCTTTCCTAGGCAGAGGACCCTGCGGCCTTCCGCAGTGTTTGTGTCCCTGG
GTACTTGAGATTAGGGAGTGGTGATGACTCCTAACGAGCATGCTGCCTTCAAGCA
TCTGTTTAACAAAGCACATCTTGCACCGCCCTTAATCCATTTAACCCTGAGTGGA
CACAGCACATGTTTCAGAGAGCACAGGGTTGGGGGTAAGGTCATAGATCAACAG
CATCCCAAGGCAGAAGAATTTTTCTTAGTACAGAACAAAATGAAGTCTCCCATGT
CTACTTCTTTCCACACAGACACAGCAACAATCCGATCTCTATCTTTTCCCCACCTT
TCCCCCTTTTCTATTCCACAAAACCACCATTGTCATCATGGCCCGTTCTCAATGAG
CTGTTGGGTACACCTCCTAGACGGGGTGGTGGCCGGGCAGAGGGGCTCCTCACC
TCCCAGAAGGGGCGGCCGGGCAGAGGCGCCCCCCACCTCCCGGACGGGGCGGCG
GCCGGGCGGAGGCGCCCCCCCACCTCCCTCCCGGACGGGGCGGCTGGCCGTGCG
GGGGCTGGCCCCCCACCTCCCTCCCGGACGGGGCGGCTGGCGGGGTGGCTACGG
CGGGGCAGAGGCGCTCCCCACATCTCAGATGATGGGCGGCCGGGCAGAGACGCT
CCTCACTTCTTAGACGGGATGGCGGCCGGGAAGAGGCGCTCCTCACTTCCCAGAC
TGGGCAGCCGGGCAGAGGGGCTCCTCACATCCCAGACGATGGGCAGCCAGGCAG
AGACGCTCCTCACTTCCCAGACGGGGTGGCGGCCGGGCAGAGGCTGCAATCTCG
GCACTTTGGGAGGCCAAGGCAGGCGGCTGGGAGGTGGAGGTTGTAGCTAGCCGA
GATCACGCCACTGCACTCCAGCCTGGGCAACATTGAGCACTGAGTGAACGAGAC
TCCGTCTGCAATCCCGGCACCTCGGGAGGCCGAGGCTGGCGGATCACTCGCGGTT
AGGAGCTGGAGACCAGCCCGGCCAACACAGCGAAACCCCGTCTCCACCAAAAAA
ACACAGGCGTGGCGGCGCGCGCCCGCAATCGCAGGCACTCGGCAGGCTGAGGCA
GGAGAATCAGGCAGGGAGGTTGCAGTGAGCCGAGATGGCAGCAGTACAGTCTA
GCTTCGGCTGGGCATCAGAGGGAGACCGTGGAAAGAGAGGGAGAGGGAGACCC
TGGGGAAAGGGGGAGGGGGAGGGAGAGCTATTCTTCCTTTTAAGAGTAGTTATG
TATTCAGGCAGTAGACAAATTGGTAGAAATAGAAATTCAGTTAAAATTTTAACA

GAATTTAATTACACAAATAATACACAAATGTAATCTTTAAAAAATTCATTCTACA
CAAAGGCAAATTT**CCTTGATGCCAACCCAAACC**

**Tile 1.2 (4084 bp)**

**TTTGCTCCCTCTACTGGGGT**ACATACCACTCAGTGCCAACAGGTCTGCCAAGA
GCACAGCCACCAAGGCCCAACAGACTCGCCTCCTCGGCCTCTGCCTCCTAACTTT
GTCTGCCATAGTGCCCCTGGGGGAGGCGAAGAAAGAAAACACTTAAAACTGGAT
ACCCGGAGTGAAGGAGACAGAATCACATCCCAGAACGGAGAGGCTGTAAGATTC
ATATCTACTTCATTCTTACCCAGGGCAGATCAGCCTACTTCTCCCCTTCTCCCTCC
TGATGGGTACAAACCATTCTCTAATCATCCCAGCAGCGCCCACACAAAGGCCCA
AGTGAAAGAGCATGGGTCCTCGGCTTTGCCACATCCCTAAACAGAGAAGCTAAC
AGGACAAGAGAAAGGAGACCCGAGCCTTTTGTCCACCACACAGGCAGTGAGTGG
CGCACCGCAGCACTGTGCCTGTTGTCAGACACCTGGGGTGCGGCCAGGGATCCC
CGCCCAGGGCTCAACCACCTCCCCACACCTGGAGCTCCCCCGCTGACGGGGAAA
GCCAGCTGGGCCTGGACGGGAAGAGAAAGGTAATGACGTTGGGGGTAGCTAGG
ATGAAGGAAAGGCCTATCTCCTCCTCGGCATCCCTGCGGGTTGTCCCGCCCCCTC
CCCTTTCCCAGCTCACTCCCCGGCCCCGCTGTGCCCACAGACGCAGTGCCAGGGG
CGGTAAGGTTCCCCGCCCGGCGGGTAGCCGTTACCCGGTGTTCACCTAAACCCCA
CCTGACGAAGGCGGCGGCTCCCACTCCCGGAAACGGATCCCGGCCCGCCCCAGC
AAGTGGGGACAGACCGGGGGCCGAGCTGCGTGCCCCCCGGTGCGAGGCCGAGC
GCCCGGCGTCCGCGCGGCCCTCCCTTGCGCTCCCGACCCGCCCCCGGAGCTCGCA
CGCTGTCTCCCCTAGGAGGGGTGGCCGGCCCGCACCTCCATCCACCCCCGCTACC
TCTTGCCTCCGCTCCTGCGGCCCCAGCTCTGGGACAGACGCGGCACGAACGTACC
CACGAGGCCGGCAGCGCCCCCCACCCGCGCCCTTCACAACTCCTCTCGGTTTGCA
AACTGTTACATTAGCCACCAACCTCTCGGCGGCGTCTCGCGCACCAGCCGGCCCC
GGACGCGGCGCGCGCTCATTGGAGCCTCGGCCGCCCGGCCCTGCGCTGCGCGCC
CGGCCCCGCCCCCGCCGCCCCGGCGCGCGTACTCATTGGACCACGTCCCAGGCCC
CGCCCCCTCCCTTTGCCAGCCTTGTAATGGGGCGGCGGAGGGGACCAGCCACCG
GTTGGGCAGAAGCAGCCGGCGTCGGCCTGGGATTGGACCACGTCACGGGGCCAG
CGCGCTCCACTCCTTCCTCCCCAGCGCCGGCCCCGGGGCGACCGCGGGACCTTTT
CCGGCGCTGGATGCTTTGCTGCCGGGGTTCGGGAGTGACCTGCGCCTAGAGCAG
GTCGGGCGCACTGCAGTCCCGCACTCGGGGCTCTTCTTCCGGTCACCTGTAGAGC
CCGGCCGAGGAATGGAATAACAGGCGTCCTGTCCGCAGCCCCGCGGAGGGGGCG
CTGGCCTGCCGGTTAGTATTGAGGTGCCCTAGGCTTTGGATAGACGTTGGAGAAT

GGCTGTCCCTCGGACTGAGTCCGTCGCAACCCCCTCTGTTTTACTGAGGAAACTG
AAGCCACAGAAAAAGATGTTTGCTGATGGGTCAGAGATAACCGATGCATTCGTC
GCAGAATCTGGCCAGGATTTCTTAGGTTAAGCCCCACTCATTTGTATAAGAAAGA
ATATTCCAAACTTAAGTACTACACACAAAAGACACTTTTATTTGTGAGGCTATAG
TTGGTTGAGCTTGTGTGCACCTTCTCCTATTTCTGTTCTATTAAATCTTGTGAACA
TTTCTCTACTCTTTAGCTGTTTCTGTCATAGAATGACAGAAGCACCTCACTGCCTT
CTCTTTTAACTTACAGCTCCATAGAAATGTGCGCAGTCGCCCGGGCGCGGTGGCT
CACGCCTGTAATCCCAGCATTTTGGGAGGCCAAGGCGGGCGAATCACTGGAGGT
CAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACCAAAAC
TACAAAAATTAGCCAGGCGTGGTGGCGGGTGCCTGTAATCCCAGCTACTCACGA
GGCTGAGGCAGGAGAATTGCTTGAATCCGGGAGGCAGAGGTTGCTGTGAGCCCA
GACTGCGCCACTGCACTCCAGCCTGGGCGACAACAGCGAGACTCCTTCTCGAAA
AAGAAATGTACGCAGTCCTTGGAGACTGCATTTTCCTTAAGTGTGGATGGAAATA
AGTCATCTGTTTTCTCTCCTACTAAGTTGATGCCTGGCCACCCTTCACAGAAAGG
CCCGCTTGAAGACCCTGGCTGTCAGGGCCAGGGTTGTGTCAAGCCTGAACTTGTT
TAAGGAGCTCTTAGCTCTTCCTTTCTCTACATTTTTCGTGATGCTTAGGTGGAGAG
TGGTCCGGTGTCTATGCTTAGAAGGAAAAATTATTGGCAGTCCCAGAGAGAGTCT
ATTCTTCCTTTTTTTTTTTTTTTTTTTTTATTGATCATTCTTGGGTGTTTCTGGCAG
AGGGGGATTTGGCAGGGTCATAGGACAATAGTGGAGGGAAGGTCAGCAGATAA
ACAAGTGAACAAAGGTCTCTGGCTTTCCTAGGCAGAGGACCCTGCGGCCTTCCGC
AGTGTTTGTGTCCCTGGGTACTTGAGATTAGGGAGTGGTGATGACTCCTAACGAG
CATGCTGCCTTCAAGCATCTGTTTAACAAAGCACATCTTGCACCGCCCTTAATCC
ATTTAACCCTGAGTGGACACAGCACATGTTTCAGAGAGCACAGGGTTGGGGGTA
AGGTCATAGATCAACAGCATCCCAAGGCAGAAGAATTTTTCTTAGTACAGAACA
AAATGAAGTCTCCCATGTCTACTTCTTTCCACACAGACACAGCAACAATCCGATC
TCTATCTTTTCCCCACCTTTCCCCCTTTTCTATTCCACAAAACCACCATTGTCATCA
TGGCCCGTTCTCAATGAGCTGTTGGGTACACCTCCTAGACGGGGTGGTGGCCGGG
CAGAGGGGCTCCTCACCTCCCAGAAGGGGCGGCCGGGCAGAGGCGCCCCCCACC
TCCCGGACGGGGCGGCGGCCGGGCGGAGGCGCCCCCCCACCTCCCTCCCGGACG
GGGCGGCTGGCCGTGCGGGGGCTGGCCCCCCACCTCCCTCCCGGACGGGGCGGC
TGGCGGGGTGGCTACGGCGGGGCAGAGGCGCTCCCCACATCTCAGATGATGGGC
GGCCGGGCAGAGACGCTCCTCACTTCTTAGACGGGATGGCGGCCGGGAAGAGGC
GCTCCTCACTTCCCAGACTGGGCAGCCGGGCAGAGGGGCTCCTCACATCCCAGA
CGATGGGCAGCCAGGCAGAGACGCTCCTCACTTCCCAGACGGGGTGGCGGCCGG

GCAGAGGCTGCAATCTCGGCACTTTGGGAGGCCAAGGCAGGCGGCTGGGAGGTG
GAGGTTGTAGCTAGCCGAGATCACGCCACTGCACTCCAGCCTGGGCAACATTGA
GCACTGAGTGAACGAGACTCCGTCTGCAATCCCGGCACCTCGGGAGGCCGAGGC
TGGCGGATCACTCGCGGTTAGGAGCTGGAGACCAGCCCGGCCAACACAGCGAAA
CCCCGTCTCCACCAAAAAAACACAGGCGTGGCGGCGCGCGCCCGCAATCGCAGG
CACTCGGCAGGCTGAGGCAGGAGAATCAGGCAGGGAGGTTGCAGTGAGCCGAG
ATGGCAGCAGTACAGTCTAGCTTCGGCTGGGCATCAGAGGGAGACCGTGGAAAG
AGAGGGAGAGGGAGACCCTGGGGAAAGGGGGAGGGGGAGGGAGAGCTATTCTT
CCTTTTAAGAGTAGTTATGTATTCAGGCAGTAGACAAATTGGTAGAAATAGAAAT
TCAGTTAAAATTTTAACAGAATTTAATTACACAAATAATACACAAATGTAATCTT
TAAAAAATTCATTCTACACAAAGGCAAATTT**CCTTGATGCCAACCCAAACC**

**Tile 1.3 (4003 bp)**

**AAGGAGACCCGAGCCTTTTG**TCCACCACACAGGCAGTGAGTGGCGCACCGCAG
CACTGTGCCTGTTGTCAGACACCTGGGGTGCGGCCAGGGATCCCCGCCCAGGGCT
CAACCACCTCCCCACACCTGGAGCTCCCCCGCTGACGGGGAAAGCCAGCTGGGC
CTGGACGGGAAGAGAAAGGTAATGACGTTGGGGGTAGCTAGGATGAAGGAAAG
GCCTATCTCCTCCTCGGCATCCCTGCGGGTTGTCCCGCCCCCTCCCCTTTCCCAGC
TCACTCCCCGGCCCCGCTGTGCCCACAGACGCAGTGCCAGGGGCGGTAAGGTTC
CCCGCCCGGCGGGTAGCCGTTACCCGGTGTTCACCTAAACCCCACCTGACGAAG
GCGGCGGCTCCCACTCCCGGAAACGGATCCCGGCCCGCCCCAGCAAGTGGGGAC
AGACCGGGGGCCGAGCTGCGTGCCCCCCGGTGCGAGGCCGAGCGCCCGGCGTCC
GCGCGGCCCTCCCTTGCGCTCCCGACCCGCCCCCGGAGCTCGCACGCTGTCTCCC
CTAGGAGGGGTGGCCGGCCCGCACCTCCATCCACCCCCGCTACCTCTTGCCTCCG
CTCCTGCGGCCCCAGCTCTGGGACAGACGCGGCACGAACGTACCCACGAGGCCG
GCAGCGCCCCCCACCCGCGCCCTTCACAACTCCTCTCGGTTTGCAAACTGTTACA
TTAGCCACCAACCTCTCGGCGGCGTCTCGCGCACCAGCCGGCCCCGGACGCGGC
GCGCGCTCATTGGAGCCTCGGCCGCCCGGCCCTGCGCTGCGCGCCCGGCCCCGCC
CCCGCCGCCCCGGCGCGCGTACTCATTGGACCACGTCCCAGGCCCCGCCCCCTCC
CTTTGCCAGCCTTGTAATGGGGCGGCGGAGGGGACCAGCCACCGGTTGGGCAGA
AGCAGCCGGCGTCGGCCTGGGATTGGACCACGTCACGGGGCCAGCGCGCTCCAC
TCCTTCCTCCCCAGCGCCGGCCCCGGGGCGACCGCGGGACCTTTTCCGGCGCTGG
ATGCTTTGCTGCCGGGGTTCGGGAGTGACCTGCGCCTAGAGCAGGTCGGGCGCA
CTGCAGTCCCGCACTCGGGGCTCTTCTTCCGGTCACCTGTAGAGCCCGGCCGAGG

AATGGAATAACAGGCGTCCTGTCCGCAGCCCCGCGGAGGGGGCGCTGGCCTGCC
GGTTAGTATTGAGGTGCCCTAGGCTTTGGATAGACGTTGGAGAATGGCTGTCCCT
CGGACTGAGTCCGTCGCAACCCCCTCTGTTTTACTGAGGAAACTGAAGCCACAGA
AAAAGATGTTTGCTGATGGGTCAGAGATAACCGATGCATTCGTCGCAGAATCTG
GCCAGGATTTCTTAGGTTAAGCCCCACTCATTTGTATAAGAAAGAATATTCCAAA
CTTAAGTACTACACACAAAAGACACTTTTATTTGTGAGGCTATAGTTGGTTGAGC
TTGTGTGCACCTTCTCCTATTTCTGTTCTATTAAATCTTGTGAACATTTCTCTACTC
TTTAGCTGTTTCTGTCATAGAATGACAGAAGCACCTCACTGCCTTCTCTTTTAACT
TACAGCTCCATAGAAATGTGCGCAGTCGCCCGGGCGCGGTGGCTCACGCCTGTA
ATCCCAGCATTTTGGGAGGCCAAGGCGGGCGAATCACTGGAGGTCAGGAGTTCG
AGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACCAAAACTACAAAAATT
AGCCAGGCGTGGTGGCGGGTGCCTGTAATCCCAGCTACTCACGAGGCTGAGGCA
GGAGAATTGCTTGAATCCGGGAGGCAGAGGTTGCTGTGAGCCCAGACTGCGCCA
CTGCACTCCAGCCTGGGCGACAACAGCGAGACTCCTTCTCGAAAAAGAAATGTA
CGCAGTCCTTGGAGACTGCATTTTCCTTAAGTGTGGATGGAAATAAGTCATCTGT
TTTCTCTCCTACTAAGTTGATGCCTGGCCACCCTTCACAGAAAGGCCCGCTTGAA
GACCCTGGCTGTCAGGGCCAGGGTTGTGTCAAGCCTGAACTTGTTTAAGGAGCTC
TTAGCTCTTCCTTTCTCTACATTTTTCGTGATGCTTAGGTGGAGAGTGGTCCGGTG
TCTATGCTTAGAAGGAAAAATTATTGGCAGTCCCAGAGAGAGTCTATTCTTCCTT
TTTTTTTTTTTTTTTTTTTTATTGATCATTCTTGGGTGTTTCTGGCAGAGGGGGATT
TGGCAGGGTCATAGGACAATAGTGGAGGGAAGGTCAGCAGATAAACAAGTGAA
CAAAGGTCTCTGGCTTTCCTAGGCAGAGGACCCTGCGGCCTTCCGCAGTGTTTGT
GTCCCTGGGTACTTGAGATTAGGGAGTGGTGATGACTCCTAACGAGCATGCTGCC
TTCAAGCATCTGTTTAACAAAGCACATCTTGCACCGCCCTTAATCCATTTAACCCT
GAGTGGACACAGCACATGTTTCAGAGAGCACAGGGTTGGGGGTAAGGTCATAGA
TCAACAGCATCCCAAGGCAGAAGAATTTTTCTTAGTACAGAACAAAATGAAGTC
TCCCATGTCTACTTCTTTCCACACAGACACAGCAACAATCCGATCTCTATCTTTTC
CCCACCTTTCCCCCTTTTCTATTCCACAAAACCACCATTGTCATCATGGCCCGTTC
TCAATGAGCTGTTGGGTACACCTCCTAGACGGGGTGGTGGCCGGGCAGAGGGGC
TCCTCACCTCCCAGAAGGGGCGGCCGGGCAGAGGCGCCCCCCACCTCCCGGACG
GGGCGGCGGCCGGGCGGAGGCGCCCCCCCACCTCCCTCCCGGACGGGGCGGCTG
GCCGTGCGGGGGCTGGCCCCCCACCTCCCTCCCGGACGGGGCGGCTGGCGGGGT
GGCTACGGCGGGGCAGAGGCGCTCCCCACATCTCAGATGATGGGCGGCCGGGCA
GAGACGCTCCTCACTTCTTAGACGGGATGGCGGCCGGGAAGAGGCGCTCCTCAC

TTCCCAGACTGGGCAGCCGGGCAGAGGGGCTCCTCACATCCCAGACGATGGGCA
GCCAGGCAGAGACGCTCCTCACTTCCCAGACGGGGTGGCGGCCGGGCAGAGGCT
GCAATCTCGGCACTTTGGGAGGCCAAGGCAGGCGGCTGGGAGGTGGAGGTTGTA
GCTAGCCGAGATCACGCCACTGCACTCCAGCCTGGGCAACATTGAGCACTGAGT
GAACGAGACTCCGTCTGCAATCCCGGCACCTCGGGAGGCCGAGGCTGGCGGATC
ACTCGCGGTTAGGAGCTGGAGACCAGCCCGGCCAACACAGCGAAACCCCGTCTC
CACCAAAAAAACACAGGCGTGGCGGCGCGCCCGCAATCGCAGGCACTCGGCA
GGCTGAGGCAGGAGAATCAGGCAGGGAGGTTGCAGTGAGCCGAGATGGCAGCA
GTACAGTCTAGCTTCGGCTGGGCATCAGAGGGAGACCGTGGAAAGAGAGGGAG
AGGGAGACCCTGGGGAAAGGGGGAGGGGGAGGGAGAGCTATTCTTCCTTTTAAG
AGTAGTTATGTATTCAGGCAGTAGACAAATTGGTAGAAATAGAAATTCAGTTAA
AATTTTAACAGAATTTAATTACACAAATAATACACAAATGTAATCTTTAAAAAAT
TCATTCTACACAAAGGCAAATTTCCTTGATGCCAACCCAAACCCCCACTCTCTCT
TGTGGAGGCACCTGCTATTGTCAGTGTGGTGTATAAGTTTCCAGGTCTTTCCCTCT
TTTACACATATAGATATCCTAGAAACATGTAGCTTTTATCTTTTCATAATATACAT
TTTTTATTTTTTGCAACCAACACCTTGTTTGCCTTTTATCTTTTTTAAAACCATAAA
ATGTTACACAATTACATTGTTTGCAATTTTTTTAAACTAATAATAAAGTGTGGAG
CATTTTCCATGTAACTGTGGTTCATTTCCTCGTTTACTGCTGTAGAACATTCCA**TA**
**GCATGGGGTTCGTTCTGC**

**Tile 3.1 (4761 bp)**

**TGCATTTGGGACAAGAAGTGT**TCCTGTTTTGTTCCAGTGGATGTAACTGTTTTA
TATCATGGGCAGCAGTTTGCTAAAGGGAAAAGAAAATACTGCAAGACAATTCTT
GCATTCTTGATTCTGAGTTATGACCTCTCTAATTCGAAACTAAAGACTTGTATTTC
TTCTTTATCTTTAGTCCATTCAGAACTCGAGTTCTCAGTATGCAGAACTAACTGTA
AGTGAACAGTCATTTTCTGTGTAATTTGCTTGGTTATATTCCTCACTATCCTGAAA
TTCCTGGGTTCAGACCAGTGTAACTAAGAGACAAACCCTGGTGTATGTTCTGACA
AGCATAATCTAGCTTTACACCCATCAAACTATCTAATTAGAGGGTAGGGGACTGA
ATTACTCATTTTATTTCCAAAGAACTTTGAAGTTTTTGCCCAGAGAGAGCCACTTT
AAGTTAAACTCTTTGAGACTCCATTTCTCATCAGGGATAATGATACCTCTTCTGC
AGATGGTTGCTACGAAGATTCCATGAAAAGCGAAGGCACTTTATTAGATATAATT
TACTACATGAAATTAGGGGCTTGTTACTGTCAGTTCTTTTTATGCCTGTCTCTAGG
GTGGTGGTTGAGGACCATGTCTTGTTCATTTATATTCTGGGCATTTGGTACAATTC
CTGGACTAGGAGTGCTCAATACATATTTGTCAAAAGAAGGAAAGGAAAGAAGGT

AGTAAGTCTGAAGAGGTGCTGTGGGTCTCTCTGGCAGGCCAATCTCGGCCTTTGT
AACCCTCATCAATAAGAGGATCCAATGGGCACAAGTAGAACCAGATTCAGCCCT
CAGAGGAAGACGTGGAGCTTCTGCCATTGGACAACAGGGGGAGCACCCAAGCA
ACAAAGATGCACTCTATCTCTTGTTCACATCTCTGCTTTTACATTTCTTCTTTTTTG
AAGCCTGTTGCCCACGCTGGTCTTGAACTCCTGGCCTCAAGGGATCCTCCTGCCT
CAGCCTCTTAAAGCGCTGGGATTACAAGCATGAGCCAGCTTGCCTGGCCACATTG
CTTCTTTTAACATATATTCTTCATAGGCTTAACAACTGACATATTAACATACTATC
TAAACATTCTGATGACCAGTATGGAATATGCTCAGTAGGTGAAAGTGGGTTCAAC
CCGCTCCCATTTGTGCTTCTGCCCAACAGTTGGTATCCTGAAACAGTTGCAATGA
ATTGGAGATATCCTGTGTGATGAAGAGGAAAGAAACTTGAGTTGGAAGTCAGGA
AATCTGTGTGCATGGTGGTGGCTAGAGTTTCTTCCTATCATTTACCTCTTCAGACA
AATGGACGCCACCTGCCCATCAACCAGATTGAGGAGAGGAGTGGATAAGAGGA
GAGGCATCCACACCACATTCATTCCTTCAACCAAACATGGAACTTCTACCAGGCT
CTGCTAGGCCTCGTTCACAGTTACGGTACAAGATGTAGCCCCCGCTCTCAGGGAA
CTCTGGCTGGTGGAAATACAGCATCAAATATTCTGAAATAAGTTGGGGGAGGGC
CGGGTGCGGTGGCTCACGCCTGTAATTCCAGTACTTGGGAGGCCGAGGCGGGTG
GATCACCCGAGGTCAGGAGTTCGAGACCAGCCTGGCCAAAATGGCAAAACCCCA
ACTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGCAGGCAGCTGTAATCC
CAGCTACTCAGGAGGCTGAGGCATAAGAATCGCTTGAACTCGGGAGGCGGAGGT
TGCAGTCAGCTGAGATTGTGCCACTGCGCTTCAGCCTGGGGGATAGAGAGAGAC
TCTGTCTAAAAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAGGAAGTTGG
GGGAGTGCAAACAATAAACTTCACAAATACCACAGGGTCACTTTTCCATATGTCC
CTACCATCTCAAAATGATGCTATGTCACTTCTGTCTATAGGTTCTCATGGCACAC
AAACAGTAAACCTGCAGTAGTTGTCCATGAGGATAGATCTTGCTTGGATTCAAAG
ATGGTTTAGCCAGAGTTGGGAGTGGAGAATGATCATGTAATATCTTGGACACCTG
CCACTTCTCATTAAAACACTCAGCTCTCAAAGCTGCCACTTCCAGGTTTTTGCTCA
GTTCTATGTCCCCTACCCTGCCACTATGAAAGGGGGAGAAAGCCCAGGGCCTTG
GGAATTCCTGCCCCACCAGTATTTCCCAGTTTCCTTTTTTTTTTTTTGAGACAAGG
TCTTGCTCTGTCGCCCATGCTGGAGTGCAGCGGCTCGATCACAGCTCACTGCAGC
TTCCAACTCCGGGGCTCAATCGATCCTCCCACTTCAGCCTCCCAAGTAGCTGGGA
CCACAGGTGTCCACCACCATGCCAGGCTCATTTTTGTATTTTTTGTAGAGATGGG
GTTTCACCATGTTGCCCAGGCTGGTCTCGAACACCTGGGCTCAAGTGATTCCCCC
ACCTCGGCCTCCCAAAGTGCTGGGATTACAGACAGGCATGAGCCATCGCGCCCA
GCCCGTTTCCTTAAGAAACGTTTTGAGCGTCTGTTAGGTGTCATACGCTCTGACG

GGGGAAACAAAAACGAATATAAAACATTGACTATGGTTCACAATCTGCGTGGTG
GCAGACACAAACTTCTAATTTTATTTCTCCTCTCATACAGCCGAAAACCTGGTTT
AACATACACTGTTTCATTTATGCCCCAAAACAGTCCTGTGAGGTCAAAACGTCAT
TCACTTACGCACGGCGGGTGACAGTTCACAAGGCACATGACATCCCTCATCTTGA
TCTCCCGGTATCCGCGCATTTTAGGAAGAAACATGCTCAAAGGGATGAAGGGTC
CGCTCTCAAGCATTTTAGCTAGCGAAGTTACAGTCACATCCGGCAAACTCGCACA
CTGCAGATCTAGGACTACCCCGCGCGGCCCCGCCCACTTTCCCAGCAGCGGGAC
GCTGTCACCCCGACTCAGGAGCTCCTGGGCCCGCGGGCTCCCGGAAGCTGCCGA
CCACGTGATTCGCTGGCTCAGCTCACGTGACAAAGCTCCCGGAGGTGGGAGCCC
TGGGCCAAAATGGCGGCCTACCTGCAGTGGCGGCGCTTCGTTTTCTTCGACAAGG
AGCTGGTGAAGGAGCCGCTGAGCAATGATGGGGCCGCTCCCGGGGCCACACCTG
CTTCTGGATCCGCTGCTTCCAAGTTCCTTTGCCTCCCTCCTGGCATCACTGTCTGC
GACTCAGGCCGAGGGAGCCTGGTCTTTGGAGATATCCTTCGTTTGGAGCTGTCTT
TTCCCTCCCGGGATCCCGAAGAGATCGAGTTAGGATGAAATCTGTTTGTCGGAGG
GGTCCGTGCCGCGCGCCTCTTTGGTTTAGCTGGTCATCCAGAGTTGTTCTGTGGTC
TACGGGAAGAAAGAAGGAAGTCCATCTCCTAACTTGTTATCAACTGAGCAATCC
TGGGCAAGTCATTTAACCTCTCTGGGTTTTAATTTCCTTATCTGTAGGAGATGTGT
GCATGCCAATTTGACGAGGTTAATTTGAGATCAATAAGAAAAAGAATGCAAGTT
GGCTTTAAAAGATGAATATGTAAAGTGACATGATTTCCCCTGCCCTAATATTATC
GTTGGCGAGGAGGTAAATGAGCCTAGCGGTTTGTTGCTTTCATCTTCATGTTGCT
GCTACTAAATTCTGGCCTTTTTACCTTTTTTTTTTTTTTTTTTTTTGATACGGAGT
CTTACTCTGTTGCCCAGGCTGGAGTGCAGTGGCGGGATCTGGGCTCACTGTAACC
GTCTCTTGAGTTCAAGCGATTCTCCTGCCACAGCCTCCCGAGTAGCTGGGACTAC
AGGCGCTTGCCACCACACCCGGCTAATTTTTTACATTTTCAGTAGAGACGGGGTT
TCACCATATTGGCCAGGCTGGTCTCGAACTCCTGACCTCGTGATCCACCCGCCTC
GGCCTCCCAAAGTGCTGGGATTACAGGAGTGAGCCATCGCGCCCGGCGCACTAA
ATTTTTTTCTTTTTTTTGAGACGGACTTTCACTCTTGTTGCACAGGCTGGAGTGCA
ATGGCATGATCTGGGCACACTGCAACCTCCGCCTCCCGGGTTCAAGCAATTCTCC
TGCCTCAACCTCCTGAGTAGCTGAAATTACAGGCACGCACCATTATACCTGGCTA
ATTTTTGTATTTTTAGTAGAGACAGGTTTCACCATGTTGGCCAGGCTGGTCTCGAA
CCCCTGACCTCAGGTGATCCGCACCCCCCCCCCCCCCCGGCCTCCCAAAGTGCTG
GGATTACAGGTGTGAGTCACCGCGCCTGGCCTCACTCTTTTTAAGATAAGGAGTA
TGTGGGAAAATCCTTGAAAATTTTAGAGTTATATACATGTAATTGTTATCTGAGT
TCTGGTCTGAATGTAAGTATCTCTCCTTGGAAAGGAGGCTCCTTGACTACCCTGC

ACATATGGAAGGCCAGATCTGGTTCTTGCCACGTTCCCTACAGCTTACAGGCTTC
CAAGCCTACAAACTACGGGTGACACACCTGTACCAACTGAAGCAGCACAATATT
CTGGCATCTGTTGGAGAAGATGAAGAGGGCATCAACCCCTTGGTGAGTCCCAGC
AGGGAAATGGGAAAGATCCAGAAGCCTAGGAATGATTTTTTGTTGGAGGATGAC
TAGCATTTACACTTCTGAGGTCTGTCCACAGGTTAAGATCTGGAACCTGGAGAAG
AGAGATGGTGGCAATCCACTCT**GCACTCGAATCTTCCCTGCT**

**Tile 3.2 (4604 bp)**

**TGCATTTGGGACAAGAAGTGTT**CCTGTTTTGTTCCAGTGGATGTAACTGTTTTA
TATCATGGGCAGCAGTTTGCTAAAGGGAAAAGAAAATACTGCAAGACAATTCTT
GCATTCTTGATTCTGAGTTATGACCTCTCTAATTCGAAACTAAAGACTTGTATTTC
TTCTTTATCTTTAGTCCATTCAGAACTCGAGTTCTCAGTATGCAGAACTAACTGTA
AGTGAACAGTCATTTTCTGTGTAATTTGCTTGGTTATATTCCTCACTATCCTGAAA
TTCCTGGGTTCAGACCAGTGTAACTAAGAGACAAACCCTGGTGTATGTTCTGACA
AGCATAATCTAGCTTTACACCCATCAAACTATCTAATTAGAGGGTAGGGGACTGA
ATTACTCATTTTATTTCCAAAGAACTTTGAAGTTTTTGCCCAGAGAGAGCCACTTT
AAGTTAAACTCTTTGAGACTCCATTTCTCATCAGGGATAATGATACCTCTTCTGC
AGATGGTTGCTACGAAGATTCCATGAAAAGCGAAGGCACTTTATTAGATATAATT
TACTACATGAAATTAGGGGCTTGTTACTGTCAGTTCTTTTTATGCCTGTCTCTAGG
GTGGTGGTTGAGGACCATGTCTTGTTCATTTATATTCTGGGCATTTGGTACAATTC
CTGGACTAGGAGTGCTCAATACATATTTGTCAAAAGAAGGAAAGGAAAGAAGGT
AGTAAGTCTGAAGAGGTGCTGTGGGTCTCTCTGGCAGGCCAATCTCGGCCTTTGT
AACCCTCATCAATAAGAGGATCCAATGGGCACAAGTAGAACCAGATTCAGCCCT
CAGAGGAAGACGTGGAGCTTCTGCCATTGGACAACAGGGGGAGCACCCAAGCA
ACAAAGATGCACTCTATCTCTTGTTCACATCTCTGCTTTTACATTTCTTCTTTTTTG
AAGCCTGTTGCCCACGCTGGTCTTGAACTCCTGGCCTCAAGGGATCCTCCTGCCT
CAGCCTCTTAAAGCGCTGGGATTACAAGCATGAGCCAGCTTGCCTGGCCACATTG
CTTCTTTTAACATATATTCTTCATAGGCTTAACAACTGACATATTAACATACTATC
TAAACATTCTGATGACCAGTATGGAATATGCTCAGTAGGTGAAAGTGGGTTCAAC
CCGCTCCCATTTGTGCTTCTGCCCAACAGTTGGTATCCTGAAACAGTTGCAATGA
ATTGGAGATATCCTGTGTGATGAAGAGGAAAGAAACTTGAGTTGGAAGTCAGGA
AATCTGTGTGCATGGTGGTGGCTAGAGTTTCTTCCTATCATTTACCTCTTCAGACA
AATGGACGCCACCTGCCCATCAACCAGATTGAGGAGAGGAGTGGATAAGAGGA
GAGGCATCCACACCACATTCATTCCTTCAACCAAACATGGAACTTCTACCAGGCT

CTGCTAGGCCTCGTTCACAGTTACGGTACAAGATGTAGCCCCCGCTCTCAGGGAA
CTCTGGCTGGTGGAAATACAGCATCAAATATTCTGAAATAAGTTGGGGGAGGGC
CGGGTGCGGTGGCTCACGCCTGTAATTCCAGTACTTGGGAGGCCGAGGCGGGTG
GATCACCCGAGGTCAGGAGTTCGAGACCAGCCTGGCCAAAATGGCAAAACCCCA
ACTCTACTAAAAATACAAAAATTAGCCAGGCGTGGTGGCAGGCAGCTGTAATCC
CAGCTACTCAGGAGGCTGAGGCATAAGAATCGCTTGAACTCGGGAGGCGGAGGT
TGCAGTCAGCTGAGATTGTGCCACTGCGCTTCAGCCTGGGGGATAGAGAGAGAC
TCTGTCTAAAAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAGGAAGTTGG
GGGAGTGCAAACAATAAACTTCACAAATACCACAGGGTCACTTTTCCATATGTCC
CTACCATCTCAAAATGATGCTATGTCACTTCTGTCTATAGGTTCTCATGGCACAC
AAACAGTAAACCTGCAGTAGTTGTCCATGAGGATAGATCTTGCTTGGATTCAAAG
ATGGTTTAGCCAGAGTTGGGAGTGGAGAATGATCATGTAATATCTTGGACACCTG
CCACTTCTCATTAAAACACTCAGCTCTCAAAGCTGCCACTTCCAGGTTTTTGCTCA
GTTCTATGTCCCCTACCCTGCCACTATGAAAGGGGGAGAAAGCCCAGGGCCTTG
GGAATTCCTGCCCCACCAGTATTTCCCAGTTTCCTTTTTTTTTTTTTGAGACAAGG
TCTTGCTCTGTCGCCCATGCTGGAGTGCAGCGGCTCGATCACAGCTCACTGCAGC
TTCCAACTCCGGGGCTCAATCGATCCTCCCACTTCAGCCTCCCAAGTAGCTGGGA
CCACAGGTGTCCACCACCATGCCAGGCTCATTTTTGTATTTTTTGTAGAGATGGG
GTTTCACCATGTTGCCCAGGCTGGTCTCGAACACCTGGGCTCAAGTGATTCCCCC
ACCTCGGCCTCCCAAAGTGCTGGGATTACAGACAGGCATGAGCCATCGCGCCCA
GCCCGTTTCCTTAAGAAACGTTTTGAGCGTCTGTTAGGTGTCATACGCTCTGACG
GGGGAAACAAAAACGAATATAAAACATTGACTATGGTTCACAATCTGCGTGGTG
GCAGACACAAACTTCTAATTTTATTTCTCCTCTCATACAGCCGAAAACCTGGTTT
AACATACACTGTTTCATTTATGCCCCAAAACAGTCCTGTGAGGTCAAAACGTCAT
TCACTTACGCACGGCGGGTGACAGTTCACAAGGCACATGACATCCCTCATCTTGA
TCTCCCGGTATCCGCGCATTTTAGGAAGAAACATGCTCAAAGGGATGAAGGGTC
CGCTCTCAAGCATTTTAGCTAGCGAAGTTACAGTCACATCCGGCAAACTCGCACA
CTGCAGATCTAGGACTACCCCGCGCGGCCCCGCCCACTTTCCCAGCAGCGGGAC
GCTGTCACCCCGACTCAGGAGCTCCTGGGCCCGCGGGCTCCCGGAAGCTGCCGA
CCACGTGATTCGCTGGCTCAGCTCACGTGACAAAGCTCCCGGAGGTGGGAGCCC
TGGGCCAAAATGGCGGCCTACCTGCAGTGGCGGCGCTTCGTTTTCTTCGACAAGG
AGCTGGTGAAGGAGCCGCTGAGCAATGATGGGGCCGCTCCCGGGGCCACACCTG
CTTCTGGATCCGCTGCTTCCAAGTTCCTTTGCCTCCCTCCTGGCATCACTGTCTGC
GACTCAGGCCGAGGGAGCCTGGTCTTTGGAGATATCCTTCGTTTGGAGCTGTCTT

TTCCCTCCCGGGATCCCGAAGAGATCGAGTTAGGATGAAATCTGTTTGTCGGAGG
GGTCCGTGCCGCGCGCCTCTTTGGTTTAGCTGGTCATCCAGAGTTGTTCTGTGGTC
TACGGGAAGAAAGAAGGAAGTCCATCTCCTAACTTGTTATCAACTGAGCAATCC
TGGGCAAGTCATTTAACCTCTCTGGGTTTTAATTTCCTTATCTGTAGGAGATGTGT
GCATGCCAATTTGACGAGGTTAATTTGAGATCAATAAGAAAAAGAATGCAAGTT
GGCTTTAAAAGATGAATATGTAAAGTGACATGATTTCCCCTGCCCTAATATTATC
GTTGGCGAGGAGGTAAATGAGCCTAGCGGTTTGTTGCTTTCATCTTCATGTTGCT
GCTACTAAATTCTGGCCTTTTTACCTTTTTTTTTTTTTTTTTTTTTGATACGGAGT
CTTACTCTGTTGCCCAGGCTGGAGTGCAGTGGCGGGATCTGGGCTCACTGTAACC
GTCTCTTGAGTTCAAGCGATTCTCCTGCCACAGCCTCCCGAGTAGCTGGGACTAC
AGGCGCTTGCCACCACACCCGGCTAATTTTTTACATTTTCAGTAGAGACGGGGTT
TCACCATATTGGCCAGGCTGGTCTCGAACTCCTGACCTCGTGATCCACCCGCCTC
GGCCTCCCAAAGTGCTGGGATTACAGGAGTGAGCCATCGCGCCCGGCGCACTAA
ATTTTTTTCTTTTTTTTGAGACGGACTTTCACTCTTGTTGCACAGGCTGGAGTGCA
ATGGCATGATCTGGGCACACTGCAACCTCCGCCTCCCGGGTTCAAGCAATTCTCC
TGCCTCAACCTCCTGAGTAGCTGAAATTACAGGCACGCACCATTATACCTGGCTA
ATTTTTGTATTTTTAGTAGAGACAGGTTTCACCATGTTGGCCAGGCTGGTCTCGAA
CCCCTGACCTCAGGTGATCCGCACCCCCCCCCCCCCCCGGCCTCCCAAAGTGCTG
GGATTACAGGTGTGAGTCACCGCGCCTGGCCTCACTCTTTTTAAGATAAGGAGTA
TGTGGGAAAATCCTTGAAAATTTTAGAGTTATATACATGTAATTGTTATCTGAGT
TCTGGTCTGAATGTAAGTATCTCTCCTTGGAAAGGAGGCTCCTTGACTACCCTGC
ACATATGGAAGGCCAGATCTGGTTCTTGCCACGTTCCCTACAGCTTACAGGCTTC
CAAGCCTACAAACTACGGGTGACACACCTGTACCAACTGAAGCAGCACAATATT
CTGGCATCTGTTGGAGAAGATGAAGAGG**GCATCAACCCCTTGGTGAGT**

**Tile 3.3 (6025 bp)**

**AGGTTTGACTTTGCTGTAACCC**TAGGCAGGAGAGAGCCCATTTTCAGCAGGAG
ATAAACCAAGATAAATTTTTTCCTACTTCTATAACAATATTTTTAGAATTACAAA
AGAAATATGTGGCTAGGGCCAGGCACGGTGGCTCACAGTACTTTGGGAGGCCCA
GGCGGGTGGGTCATTTGTGGTCAGGAGTTTGAGACCAGCCTGGACAACATGGTG
AAACCCTGTCTCTAGTAAAAATACAAAAATTAACTGGGTGGTAGTGGTGTGTGCC
TGTAATCCCAGATACTTGAGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAG
GCAGAGGTTGGGGTGAGCTGAGATCGCGCCATTGCACTCCAGCCTGGGCGACAG
AGTGAGACCCTGTCTCCCCGGCCCCCGCCTCCAAAAAAAAGAAAGAAATATGGT

GGCTTACACTTGTAATCCTAGCACTTTGGAAGGCTGAGGCGTGTGGATTGCTTGA
GCCTAGCAGTTTGTGACTAGCCTGGGCAACAGGGCAGAACACTGTCTCTACTAA
AAGTACAACAACAACAAAATTAACTGGGCATGGTGGCATGAGCCTGTAGTCTCA
GCTACTGGGGAGGCTGAGGTGGGAGTATCTCGTGAGCCTAGGAGGTGGAGACTG
TGGTGAGCCGTGATTGTGCCACTGCACTCTAGCCGTGTCTCAAAAAACAAAACA
AAACAAAAAAGAACGAAAGAAATATGCGGTTATTGACGAAAACTGGAAGAACA
TTAAAACAACAACAGGCCGGGCGTGGTGGCTCATGCCTGTAATCCTAGCACTTTG
GGAGGCAGAGGCGAGTGGATCACTTGAGGCCAGGAGTTTGAGACCAACCGGGG
AAACATGTCAAAACCATGTCTCTACAAAAAATACAAAAACTAGCCGGGCATGAT
GGCGTGCGCCTGTAGTCTCAGCTACTTGGGTGGCTGAGGCACAAGAATCGCTTGA
ACCTGGGAGACGAGACAGAGGTTGCAGTGAGTCGAGATCGTGCCACTGCACTCC
AGCCTGGGCAATGGAGGGAGACTCTATCTCAAAAAGAAAAGAAGGTAGTTTCTA
CCTGGAGAGCCAGATGATAAAAACGGAAAATTAAAGTTTGCTGAGGCAACTACA
CAGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGGACGTTGCAGTGA
GCCGAGATCACGCCACTGCACTCAAGCCTGACGACTGGCGACAGAGCGAGACTC
CGTCTCAAAAATAAAATAAAATAAGATAAAAATAAATTAAAGTTTGCATTTGGG
ACAAGAAGTGTTCCTGTTTTGTTCCAGTGGATGTAACTGTTTTATATCATGGGCA
GCAGTTTGCTAAAGGGAAAAGAAAATACTGCAAGACAATTCTTGCATTCTTGATT
CTGAGTTATGACCTCTCTAATTCGAAACTAAAGACTTGTATTTCTTCTTTATCTTT
AGTCCATTCAGAACTCGAGTTCTCAGTATGCAGAACTAACTGTAAGTGAACAGTC
ATTTTCTGTGTAATTTGCTTGGTTATATTCCTCACTATCCTGAAATTCCTGGGTTC
AGACCAGTGTAACTAAGAGACAAACCCTGGTGTATGTTCTGACAAGCATAATCT
AGCTTTACACCCATCAAACTATCTAATTAGAGGGTAGGGGACTGAATTACTCATT
TTATTTCCAAAGAACTTTGAAGTTTTTGCCCAGAGAGAGCCACTTTAAGTTAAAC
TCTTTGAGACTCCATTTCTCATCAGGGATAATGATACCTCTTCTGCAGATGGTTGC
TACGAAGATTCCATGAAAAGCGAAGGCACTTTATTAGATATAATTTACTACATGA
AATTAGGGGCTTGTTACTGTCAGTTCTTTTTATGCCTGTCTCTAGGGTGGTGGTTG
AGGACCATGTCTTGTTCATTTATATTCTGGGCATTTGGTACAATTCCTGGACTAGG
AGTGCTCAATACATATTTGTCAAAAGAAGGAAAGGAAAGAAGGTAGTAAGTCTG
AAGAGGTGCTGTGGGTCTCTCTGGCAGGCCAATCTCGGCCTTTGTAACCCTCATC
AATAAGAGGATCCAATGGGCACAAGTAGAACCAGATTCAGCCCTCAGAGGAAG
ACGTGGAGCTTCTGCCATTGGACAACAGGGGGAGCACCCAAGCAACAAAGATGC
ACTCTATCTCTTGTTCACATCTCTGCTTTTACATTTCTTCTTTTTTGAAGCCTGTTG
CCCACGCTGGTCTTGAACTCCTGGCCTCAAGGGATCCTCCTGCCTCAGCCTCTTA

AAGCGCTGGGATTACAAGCATGAGCCAGCTTGCCTGGCCACATTGCTTCTTTTAA
CATATATTCTTCATAGGCTTAACAACTGACATATTAACATACTATCTAAACATTCT
GATGACCAGTATGGAATATGCTCAGTAGGTGAAAGTGGGTTCAACCCGCTCCCA
TTTGTGCTTCTGCCCAACAGTTGGTATCCTGAAACAGTTGCAATGAATTGGAGAT
ATCCTGTGTGATGAAGAGGAAAGAAACTTGAGTTGGAAGTCAGGAAATCTGTGT
GCATGGTGGTGGCTAGAGTTTCTTCCTATCATTTACCTCTTCAGACAAATGGACG
CCACCTGCCCATCAACCAGATTGAGGAGAGGAGTGGATAAGAGGAGAGGCATCC
ACACCACATTCATTCCTTCAACCAAACATGGAACTTCTACCAGGCTCTGCTAGGC
CTCGTTCACAGTTACGGTACAAGATGTAGCCCCCGCTCTCAGGGAACTCTGGCTG
GTGGAAATACAGCATCAAATATTCTGAAATAAGTTGGGGGAGGGCCGGGTGCGG
TGGCTCACGCCTGTAATTCCAGTACTTGGGAGGCCGAGGCGGGTGGATCACCCG
AGGTCAGGAGTTCGAGACCAGCCTGGCCAAAATGGCAAAACCCCAACTCTACTA
AAAATACAAAAATTAGCCAGGCGTGGTGGCAGGCAGCTGTAATCCCAGCTACTC
AGGAGGCTGAGGCATAAGAATCGCTTGAACTCGGGAGGCGGAGGTTGCAGTCAG
CTGAGATTGTGCCACTGCGCTTCAGCCTGGGGGATAGAGAGAGACTCTGTCTAA
AAAAAAAAAAAAAAAAAAAGAAAAAAGAAAAAAAAGGAAGTTGGGGGAGTGC
AAACAATAAACTTCACAAATACCACAGGGTCACTTTTCCATATGTCCCTACCATC
TCAAAATGATGCTATGTCACTTCTGTCTATAGGTTCTCATGGCACACAAACAGTA
AACCTGCAGTAGTTGTCCATGAGGATAGATCTTGCTTGGATTCAAAGATGGTTTA
GCCAGAGTTGGGAGTGGAGAATGATCATGTAATATCTTGGACACCTGCCACTTCT
CATTAAAACACTCAGCTCTCAAAGCTGCCACTTCCAGGTTTTTGCTCAGTTCTATG
TCCCCTACCCTGCCACTATGAAAGGGGGAGAAAGCCCAGGGCCTTGGGAATTCC
TGCCCCACCAGTATTTCCCAGTTTCCTTTTTTTTTTTTTGAGACAAGGTCTTGCTCT
GTCGCCCATGCTGGAGTGCAGCGGCTCGATCACAGCTCACTGCAGCTTCCAACTC
CGGGGCTCAATCGATCCTCCCACTTCAGCCTCCCAAGTAGCTGGGACCACAGGTG
TCCACCACCATGCCAGGCTCATTTTTGTATTTTTTGTAGAGATGGGGTTTCACCAT
GTTGCCCAGGCTGGTCTCGAACACCTGGGCTCAAGTGATTCCCCCACCTCGGCCT
CCCAAAGTGCTGGGATTACAGACAGGCATGAGCCATCGCGCCCAGCCCGTTTCCT
TAAGAAACGTTTTGAGCGTCTGTTAGGTGTCATACGCTCTGACGGGGGAAACAA
AAACGAATATAAAACATTGACTATGGTTCACAATCTGCGTGGTGGCAGACACAA
ACTTCTAATTTTATTTCTCCTCTCATACAGCCGAAAACCTGGTTTAACATACACTG
TTTCATTTATGCCCCAAAACAGTCCTGTGAGGTCAAAACGTCATTCACTTACGCA
CGGCGGGTGACAGTTCACAAGGCACATGACATCCCTCATCTTGATCTCCCGGTAT
CCGCGCATTTTAGGAAGAAACATGCTCAAAGGGATGAAGGGTCCGCTCTCAAGC

ATTTTAGCTAGCGAAGTTACAGTCACATCCGGCAAACTCGCACACTGCAGATCTA
GGACTACCCCGCGCGGCCCCGCCCACTTTCCCAGCAGCGGGACGCTGTCACCCCG
ACTCAGGAGCTCCTGGGCCCGCGGGCTCCCGGAAGCTGCCGACCACGTGATTCG
CTGGCTCAGCTCACGTGACAAAGCTCCCGGAGGTGGGAGCCCTGGGCCAAAATG
GCGGCCTACCTGCAGTGGCGGCGCTTCGTTTTCTTCGACAAGGAGCTGGTGAAGG
AGCCGCTGAGCAATGATGGGGCCGCTCCCGGGGCCACACCTGCTTCTGGATCCG
CTGCTTCCAAGTTCCTTTGCCTCCCTCCTGGCATCACTGTCTGCGACTCAGGCCGA
GGGAGCCTGGTCTTTGGAGATATCCTTCGTTTGGAGCTGTCTTTTCCCTCCCGGGA
TCCCGAAGAGATCGAGTTAGGATGAAATCTGTTTGTCGGAGGGGTCCGTGCCGC
GCGCCTCTTTGGTTTAGCTGGTCATCCAGAGTTGTTCTGTGGTCTACGGGAAGAA
AGAAGGAAGTCCATCTCCTAACTTGTTATCAACTGAGCAATCCTGGGCAAGTCAT
TTAACCTCTCTGGGTTTTAATTTCCTTATCTGTAGGAGATGTGTGCATGCCAATTT
GACGAGGTTAATTTGAGATCAATAAGAAAAAGAATGCAAGTTGGCTTTAAAAGA
TGAATATGTAAAGTGACATGATTTCCCCTGCCCTAATATTATCGTTGGCGAGGAG
GTAAATGAGCCTAGCGGTTTGTTGCTTTCATCTTCATGTTGCTGCTACTAAATTCT
GGCCTTTTTACCTTTTTTTTTTTTTTTTTTTTTTGATACGGAGTCTTACTCTGTTGC
CCAGGCTGGAGTGCAGTGGCGGGATCTGGGCTCACTGTAACCGTCTCTTGAGTTC
AAGCGATTCTCCTGCCACAGCCTCCCGAGTAGCTGGGACTACAGGCGCTTGCCAC
CACACCCGGCTAATTTTTTACATTTTCAGTAGAGACGGGGTTTCACCATATTGGC
CAGGCTGGTCTCGAACTCCTGACCTCGTGATCCACCCGCCTCGGCCTCCCAAAGT
GCTGGGATTACAGGAGTGAGCCATCGCGCCCGGCGCACTAAATTTTTTTCTTTTT
TTTGAGACGGACTTTCACTCTTGTTGCACAGGCTGGAGTGCAATGGCATGATCTG
GGCACACTGCAACCTCCGCCTCCCGGGTTCAAGCAATTCTCCTGCCTCAACCTCC
TGAGTAGCTGAAATTACAGGCACGCACCATTATACCTGGCTAATTTTTGTATTTTT
AGTAGAGACAGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACCCCTGACCTCAG
GTGATCCGCACCCCCCCCCCCCCCCGGCCTCCCAAAGTGCTGGGATTACAGGTGT
GAGTCACCGCGCCTGGCCTCACTCTTTTTAAGATAAGGAGTATGTGGGAAAATCC
TTGAAAATTTTAGAGTTATATACATGTAATTGTTATCTGAGTTCTGGTCTGAATGT
AAGTATCTCTCCTTGGAAAGGAGGCTCCTTGACTACCCTGCACATATGGAAGGCC
AGATCTGGTTCTTGCCACGTTCCCTACAGCTTACAGGCTTCCAAGCCTACAAACT
ACGGGTGACACACCTGTACCAACTGAAGCAGCACAATATTCTGGCATCTGTTGG
AGAAGATGAAGAGGGCATCAACCCCTTGGTGAGTCCCAGCAGGGAAATGGGAA
AGATCCAGAAGCCTAGGAATGATTTTTTGTTGGAGGATGACTAGCATTTACACTT

CTGAGGTCTGTCCACAGGTTAAGATCTGGAACCTGGAGA**AGAGAGATGGTGGC**
**AATCCA**

**Tile 4.1 (4762 bp)**

**GGTTCTTGCCACGTTCCCTA**CAGCTTACAGGCTTCCAAGCCTACAAACTACGGG
TGACACACCTGTACCAACTGAAGCAGCACAATATTCTGGCATCTGTTGGAGAAG
ATGAAGAGGGCATCAACCCCTTGGTGAGTCCCAGCAGGGAAATGGGAAAGATCC
AGAAGCCTAGGAATGATTTTTTGTTGGAGGATGACTAGCATTTACACTTCTGAGG
TCTGTCCACAGGTTAAGATCTGGAACCTGGAGAAGAGAGATGGTGGCAATCCAC
TCTGCACTCGAATCTTCCCTGCTATTCCAGGAACAGAGCCAACTGTTGTATCTTGT
TTGACTGTCCATGAAAATCTCAACTTTATGGCCATTGGTAAACAGAAGGCAAAAC
TAACCCTCCTAGATTTGTTATAGATTTTCTTCAGAGTTGCTTCTGCCTCTCATCTTT
ACTGTTTTCAGGAGACCACTTTGTACTGAACTGAGGCCTTTGCGATATTAAATTTT
GGAATGGGGGCCAGGCGCGGTGGCTCATGCCTGTAATCCCAGCACCTTGGGAGG
CTGAGGTGGGCGGATCACCTGAGATTGGGAGTTTGAGATCAGCCTGACCAACAT
GGAGAAACCCCATCTCTACTAAAAATACAAAATTAGTTGGGTGTAGTGGCACAT
GCCTGTAATCCCAGCTACTAGGGAGGCTGCGGCAGGAAAATCACTTGAACCTGG
GAGGCGGAGTTTGCAGTGAGCCAAGATCGCGCCATTGCACTGTAGCCTGGGCAA
CAAGAGCAAAACTCTGTCTCAAAATAAATAAATAAATAAATAAATAAATAAAAT
AAATAAAGTTTGGAATGGGAAGGCATTTAATTTTGTTCTTTGAAAATGTGTCAGT
GAGTACCAATATTCAGGTTAACCATTAGAGAGGTTGGGTATATGGAGAAAGCTT
GTCACTTCTGAGTGCCTCAAGTGGTCTTTTTTCCCCTCTCCACTTCCTGATCTTTTC
AGCCTTACTGAAGTAATTTTCTTGTTTTCTTACAGGTTTCACAGATGGCAGTGTTA
CATTGAACAAAGGAGACATCACCCGGGACCGGCATAGCAAGACCCAGATTTTGC
ACAAGGGCAACTATCCTGTAACTGGATTGGCCTTTCGCCAAGCAGGAAAGACCA
CTCACTTGTTTGTTGTGACAACAGAGAACGTCCAGGTATGACCAAGGCCTCCACT
CTTAGGAGCAGGCAGGGAGGGCTTCTCCATTGTTCAGGGGGATAGGGTAATGAA
GTGACAGGAAAGGTGGGAGTGTTAAAGTTTTAATCATATAATTCGAATCATTTGC
CTAGCTTTGTATTTTATTTTTTTGCCTAGGAAGCTGGAAAGAGTTAGACTCTGGGC
TTGTAGTAAAAAGATGTGAATTTTCTTTCTTTTTTTTTTTTGAGACCGAGTCTCAC
TGTGTTGCTCAGGCTGGAGTGCAGTGGCGAGATCTTGACTCACTGTAACCTCAGT
CTCCCAGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAG
GCACACACCAGCATGTCCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCA
CCATGTTGGCCAGTCTGGTCTCCAACTCCTGACCTCAGGTGATCCACCCACCTTG

GCCTTCCAAAGTGCTGAGATTACAGGCGTGAGCCACCGCACTCGGCCTTGAATTT
TATTTCTAACTAGCTTTTTTTTTTTTTTTCTTTGAGACAGGATCTTGCTGTGTAACC
CAGGCTAGAGTGCAATGGTGCTGTCTCAGCTCACTGCAACCTCTGCCTCCTGGGC
TCAAGTGATCCTCCCACCTCAGCCTCCTGAGTACGTGGGATTACAGGCATGTGTC
ACCATGCCTGGCTAATTTTTGTATTTTTTCTAGATATGGGGGTTTCACCATGTTGC
CCAGGCTGATCTCGAACTCCTGGGCCCAAGCAATCCACCTGCCTCAACCACCCAA
AGTGCTGGAATTACAGGCATGAGCCACTGCACCGGCCTATTTTCTAACTAACTTT
TTAAATGACGTTGGACAAGTATCTTTACCTTTCTTGTTCCTGGTGCCCCAGATTTA
CAACAGTCTACTCTCTCAGGAATACTCTTGATAAAGTAGTAGATTATTACATTTTT
CTTACTTGCAGAATTTCTGTACTAGATAGTGATAATATATACAGAGAAAGAAATT
GTGATCTCTGGTTATATATTTAATAGATGGGTAATGGTTAATGAAGAAGGCCAGC
AGCCAAGAGAAAGACTTTAGAATGCCAAGAGGGAAAAAAGAGCCAGTATGGAG
ATGGGAATACCTTTGTGAAGGCTTAGAGTTACCTCCTCAAAGGAGCCTGTTAACC
CCTTTGTTGCTTCTGGCAGTCCTATATAGTTTCTGGAAAAGACTACCCTCGCGTGG
AGTTGGACACCCATGGTTGTGGCCTGCGCTGCTCAGCCCTAAGTGACCCTTCTCA
GGACCTGCAGTTCATTGTGGCCGGGGATGAGTGTGTCTACTTGTACCAGCCTGAT
GAACGTGGGCCCTGCTTCGCCTTTGAGGGCCATAAGCTCATTGCCCACTGGTTTA
GAGGCTACCTTATCATTGTCTCCCGTGACCGGAAGGTTTCTCCCAAGTAAGGACT
CAGTGAGAAGGGACAGGGAGAGGGCTGGACTTGTTCCCCAGAATCCGCCTGATT
TTAAAATCCTAATGCCTGGATACTGCCAATCTCCTACTCCTACTCCGGTGTTATGT
AGGTAACATTTCTGTTTTTTTTTTTTGAGATGGAGTCTCGCTCTGTCACCCAGGAG
TGCAGTGGCGTGATCTCGGCTCACTGCAGCCTCGCCTCCTGGGTTCCAGTGATAC
TCCTGCCTCAGCCTCCTGAATAGCTGGGATTACAGGCATGTGCCACCACGCCTGG
CTACTTTTTGTATTTTCAGTAGAGACGAGGTTTCACCATGTTGGCCAGGCTGGTGT
CAGACTCCTGACCTCAAGTGATCTACCCGCCTCGGCCTCCCAAAGTACTGGGATT
ACAGGTGTGAGCCACTGCACCCAGCCAGTGACACTTATAGCTAGAGTTGGTTAC
AATGCTTTTTTTAACTGAAAAATGGATGTTTTTACTGCAGGTATTTGCTACATAGC
CCTAGCATACTAGAATGCTCCAGGGAATTTGGACCTGCTTTTTTTGTTCTGTTTTG
TTTGAGATGGAGTCTCGCTCTATTGCCCAGGCTGGAGTGCAGTGGCGCGATCTCG
GCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCC
GAGCAACTGGGACTACAGGTGCCCGCCCCCATGCCCGGCTAATTTTTTGTATTTT
TATTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTC
GTGATCTGCCTGCCTCGTCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTG
CGCCCGGCCTTGTTTTGTTTTTTAACTCTTTTGTCAAAATAAATTGGGCAGAGGAA

CAAGGGATTCAGCAGTGGAGAAATTAGAAAGTGCTGGGAAACGTAGGATGCAA
GTTTTCTGTTTTGTTTTGGTTTTATTTTTTTAAGGGGTAAGTGGGGTCTTCTGCATG
ATGAAAGTACTCCCTGTTAGATGCTGGTTTGTTTTGCTTTCATTCAGTTCCAGGCA
CAGTGTTTTACATATACCATGTATTGGATTATTATTTTCTAAATGAATGGATATGT
TAGTGTACTTAATTTTCTAATCTGTTACTTGGGTTAATTGTAGGATCCTAAAGGGT
AAGATCAGGTTTATGTGTTCTCTATGTACAACTACCGGCACAGCGCTGCATGTTA
TGGGATGAGAGGGACCAGAGAGGTGATTTGTGCTGGGTGAAATGTGCACCAAAG
TTATCATGGGAAAGGAAATAGGGGAGATAAGACAGAGATGTCCAAACATCTGGT
GCTTTTTCTTTCCTATGCAGGTCAGAGTTTACCAGCAGGGATTCACAGAGCTCCG
ACAAGCAGATTCTAAACATCTATGACCTGTGCAACAAGTTCATAGCCTATAGCAC
CGTCTTTGAGGATGTAGTGGATGTGCTTGCTGAGTGGGGCTCCCTGTACGTGCTG
ACGCGGGATGGGCGGGTCCACGCACTGCAGGAGAAGGACACACAGACCAAACT
GGAGGCAAGGCCACCAGGCTCGCAGAGCTGGCCACAGGCACCTAGAAGCAGCT
GCAGGAGCAGGTTCCCCAAGTCATATTGGCCAGGGTGTTTCCCTCCTTGTGGGTC
ACAGCCTTACTCAGCTTCCTTAGGGTAGGATTATAAGGTAAATGGCCTGGGATGG
GGCTTTAGAGGGAAGTAGTGGCTGGGAGCCTGAGCCCACTCTAAGGGTTCACTTT
GTTTCCAGAGAACCTTATGAAATAAGATGATACTGATCTTGGGGATATCTGGTAG
CTCTTGGTTTGTCTAGGATCTAGGCAGATCAGAAATCCAGGACTTTCTGTTTTGTG
TGTTGTGCGCACATTTTGGGAAAGTGTCTTCCCAGGACCACTTCTACCAATTTTCT
AATCTCTCTTCCCTTCTAGATGCTGTTTAAGAAGAACCTATTTGAGATGGCGATTA
ACCTTGCCAAGAGCCAGCATCTGGACAGTGATGGGCTGGCCCAGATTTTCATGCA
GTATGGAGACCATCTCTACAGCAAGGGCAACCACGATGGGGCTGTCCAGCAATA
TATCCGGTCA**GTCTGGAGGCACTTTGGGAT**

**Tile 4.2 (4714 bp)**

**GCATCAACCCCTTGGTGAGT**CCCAGCAGGGAAATGGGAAAGATCCAGAAGCCT
AGGAATGATTTTTTGTTGGAGGATGACTAGCATTTACACTTCTGAGGTCTGTCCA
CAGGTTAAGATCTGGAACCTGGAGAAGAGAGATGGTGGCAATCCACTCTGCACT
CGAATCTTCCCTGCTATTCCAGGAACAGAGCCAACTGTTGTATCTTGTTTGACTGT
CCATGAAAATCTCAACTTTATGGCCATTGGTAAACAGAAGGCAAAACTAACCCT
CCTAGATTTGTTATAGATTTTCTTCAGAGTTGCTTCTGCCTCTCATCTTTACTGTTT
TCAGGAGACCACTTTGTACTGAACTGAGGCCTTTGCGATATTAAATTTTGGAATG
GGGGCCAGGCGCGGTGGCTCATGCCTGTAATCCCAGCACCTTGGGAGGCTGAGG
TGGGCGGATCACCTGAGATTGGGAGTTTGAGATCAGCCTGACCAACATGGAGAA

ACCCCATCTCTACTAAAAATACAAAATTAGTTGGGTGTAGTGGCACATGCCTGTA
ATCCCAGCTACTAGGGAGGCTGCGGCAGGAAAATCACTTGAACCTGGGAGGCGG
AGTTTGCAGTGAGCCAAGATCGCGCCATTGCACTGTAGCCTGGGCAACAAGAGC
AAAACTCTGTCTCAAAATAAATAAATAAATAAATAAATAAATAAAATAAATAAA
GTTTGGAATGGGAAGGCATTTAATTTTGTTCTTTGAAAATGTGTCAGTGAGTACC
AATATTCAGGTTAACCATTAGAGAGGTTGGGTATATGGAGAAAGCTTGTCACTTC
TGAGTGCCTCAAGTGGTCTTTTTTCCCCTCTCCACTTCCTGATCTTTTCAGCCTTAC
TGAAGTAATTTTCTTGTTTTCTTACAGGTTTCACAGATGGCAGTGTTACATTGAAC
AAAGGAGACATCACCCGGGACCGGCATAGCAAGACCCAGATTTTGCACAAGGGC
AACTATCCTGTAACTGGATTGGCCTTTCGCCAAGCAGGAAAGACCACTCACTTGT
TTGTTGTGACAACAGAGAACGTCCAGGTATGACCAAGGCCTCCACTCTTAGGAG
CAGGCAGGGAGGGCTTCTCCATTGTTCAGGGGGATAGGGTAATGAAGTGACAGG
AAAGGTGGGAGTGTTAAAGTTTTAATCATATAATTCGAATCATTTGCCTAGCTTT
GTATTTTATTTTTTTGCCTAGGAAGCTGGAAAGAGTTAGACTCTGGGCTTGTAGT
AAAAAGATGTGAATTTTCTTTCTTTTTTTTTTTTGAGACCGAGTCTCACTGTGTTG
CTCAGGCTGGAGTGCAGTGGCGAGATCTTGACTCACTGTAACCTCAGTCTCCCAG
GTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCACACA
CCAGCATGTCCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTG
GCCAGTCTGGTCTCCAACTCCTGACCTCAGGTGATCCACCCACCTTGGCCTTCCA
AAGTGCTGAGATTACAGGCGTGAGCCACCGCACTCGGCCTTGAATTTTATTTCTA
ACTAGCTTTTTTTTTTTTTTCTTTGAGACAGGATCTTGCTGTGTAACCCAGGCTA
GAGTGCAATGGTGCTGTCTCAGCTCACTGCAACCTCTGCCTCCTGGGCTCAAGTG
ATCCTCCCACCTCAGCCTCCTGAGTACGTGGGATTACAGGCATGTGTCACCATGC
CTGGCTAATTTTTGTATTTTTTCTAGATATGGGGGTTTCACCATGTTGCCCAGGCT
GATCTCGAACTCCTGGGCCCAAGCAATCCACCTGCCTCAACCACCCAAAGTGCTG
GAATTACAGGCATGAGCCACTGCACCGGCCTATTTTCTAACTAACTTTTTAAATG
ACGTTGGACAAGTATCTTTACCTTTCTTGTTCCTGGTGCCCCAGATTTACAACAGT
CTACTCTCTCAGGAATACTCTTGATAAAGTAGTAGATTATTACATTTTTCTTACTT
GCAGAATTTCTGTACTAGATAGTGATAATATATACAGAGAAAGAAATTGTGATCT
CTGGTTATATATTTAATAGATGGGTAATGGTTAATGAAGAAGGCCAGCAGCCAA
GAGAAAGACTTTAGAATGCCAAGAGGGAAAAAAGAGCCAGTATGGAGATGGGA
ATACCTTTGTGAAGGCTTAGAGTTACCTCCTCAAAGGAGCCTGTTAACCCCTTTG
TTGCTTCTGGCAGTCCTATATAGTTTCTGGAAAAGACTACCCTCGCGTGGAGTTG
GACACCCATGGTTGTGGCCTGCGCTGCTCAGCCCTAAGTGACCCTTCTCAGGACC

TGCAGTTCATTGTGGCCGGGGATGAGTGTGTCTACTTGTACCAGCCTGATGAACG
TGGGCCCTGCTTCGCCTTTGAGGGCCATAAGCTCATTGCCCACTGGTTTAGAGGC
TACCTTATCATTGTCTCCCGTGACCGGAAGGTTTCTCCCAAGTAAGGACTCAGTG
AGAAGGGACAGGGAGAGGGCTGGACTTGTTCCCCAGAATCCGCCTGATTTTAAA
ATCCTAATGCCTGGATACTGCCAATCTCCTACTCCTACTCCGGTGTTATGTAGGTA
ACATTTCTGTTTTTTTTTTTTGAGATGGAGTCTCGCTCTGTCACCCAGGAGTGCAG
TGGCGTGATCTCGGCTCACTGCAGCCTCGCCTCCTGGGTTCCAGTGATACTCCTG
CCTCAGCCTCCTGAATAGCTGGGATTACAGGCATGTGCCACCACGCCTGGCTACT
TTTTGTATTTTCAGTAGAGACGAGGTTTCACCATGTTGGCCAGGCTGGTGTCAGA
CTCCTGACCTCAAGTGATCTACCCGCCTCGGCCTCCCAAAGTACTGGGATTACAG
GTGTGAGCCACTGCACCCAGCCAGTGACACTTATAGCTAGAGTTGGTTACAATGC
TTTTTTTAACTGAAAAATGGATGTTTTTACTGCAGGTATTTGCTACATAGCCCTAG
CATACTAGAATGCTCCAGGGAATTTGGACCTGCTTTTTTTGTTCTGTTTTGTTTGA
GATGGAGTCTCGCTCTATTGCCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCA
CTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGCA
ACTGGGACTACAGGTGCCCGCCCCCATGCCCGGCTAATTTTTTGTATTTTTATTAG
AGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATC
TGCCTGCCTCGTCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCCG
GCCTTGTTTTGTTTTTTAACTCTTTTGTCAAAATAAATTGGGCAGAGGAACAAGG
GATTCAGCAGTGGAGAAATTAGAAAGTGCTGGGAAACGTAGGATGCAAGTTTTC
TGTTTTGTTTTGGTTTTATTTTTTTAAGGGGTAAGTGGGGTCTTCTGCATGATGAA
AGTACTCCCTGTTAGATGCTGGTTTGTTTTGCTTTCATTCAGTTCCAGGCACAGTG
TTTTACATATACCATGTATTGGATTATTATTTTCTAAATGAATGGATATGTTAGTG
TACTTAATTTTCTAATCTGTTACTTGGGTTAATTGTAGGATCCTAAAGGGTAAGAT
CAGGTTTATGTGTTCTCTATGTACAACTACCGGCACAGCGCTGCATGTTATGGGA
TGAGAGGGACCAGAGAGGTGATTTGTGCTGGGTGAAATGTGCACCAAAGTTATC
ATGGGAAAGGAAATAGGGGAGATAAGACAGAGATGTCCAAACATCTGGTGCTTT
TTCTTTCCTATGCAGGTCAGAGTTTACCAGCAGGGATTCACAGAGCTCCGACAAG
CAGATTCTAAACATCTATGACCTGTGCAACAAGTTCATAGCCTATAGCACCGTCT
TTGAGGATGTAGTGGATGTGCTTGCTGAGTGGGGCTCCCTGTACGTGCTGACGCG
GGATGGGCGGGTCCACGCACTGCAGGAGAAGGACACACAGACCAAACTGGAGG
CAAGGCCACCAGGCTCGCAGAGCTGGCCACAGGCACCTAGAAGCAGCTGCAGGA
GCAGGTTCCCCAAGTCATATTGGCCAGGGTGTTTCCCTCCTTGTGGGTCACAGCC
TTACTCAGCTTCCTTAGGGTAGGATTATAAGGTAAATGGCCTGGGATGGGGCTTT

AGAGGGAAGTAGTGGCTGGGAGCCTGAGCCCACTCTAAGGGTTCACTTTGTTTCC
AGAGAACCTTATGAAATAAGATGATACTGATCTTGGGGATATCTGGTAGCTCTTG
GTTTGTCTAGGATCTAGGCAGATCAGAAATCCAGGACTTTCTGTTTTGTGTGTTGT
GCGCACATTTTGGGAAAGTGTCTTCCCAGGACCACTTCTACCAATTTTCTAATCTC
TCTTCCCTTCTAGATGCTGTTTAAGAAGAACCTATTTGAGATGGCGATTAACCTTG
CCAAGAGCCAGCATCTGGACAGTGATGGGCTGGCCCAGATTTTCATGCAGTATG
GAGACCATCTCTACAGCAAGGGCAACCACGATGGGGCTGTCCAGCAATATATCC
GGTCAGTCTGGAGGCACTTTGGGATATAGCTGTGAATGCAGGGACAGGCAGCTA
GATAGCTAAAGCCCATCCATGCT**CCAGGTAGGGGCTAGAA**

**Tile 4.3 (4732 bp)**

**GGTTCTTGCCACGTTCCCTA**CAGCTTACAGGCTTCCAAGCCTACAAACTACGGG
TGACACACCTGTACCAACTGAAGCAGCACAATATTCTGGCATCTGTTGGAGAAG
ATGAAGAGGGCATCAACCCCTTGGTGAGTCCCAGCAGGGAAATGGGAAAGATCC
AGAAGCCTAGGAATGATTTTTTGTTGGAGGATGACTAGCATTTACACTTCTGAGG
TCTGTCCACAGGTTAAGATCTGGAACCTGGAGAAGAGAGATGGTGGCAATCCAC
TCTGCACTCGAATCTTCCCTGCTATTCCAGGAACAGAGCCAACTGTTGTATCTTGT
TTGACTGTCCATGAAAATCTCAACTTTATGGCCATTGGTAAACAGAAGGCAAAAC
TAACCCTCCTAGATTTGTTATAGATTTTCTTCAGAGTTGCTTCTGCCTCTCATCTTT
ACTGTTTTCAGGAGACCACTTTGTACTGAACTGAGGCCTTTGCGATATTAAATTTT
GGAATGGGGGCCAGGCGCGGTGGCTCATGCCTGTAATCCCAGCACCTTGGGAGG
CTGAGGTGGGCGGATCACCTGAGATTGGGAGTTTGAGATCAGCCTGACCAACAT
GGAGAAACCCCATCTCTACTAAAAATACAAAATTAGTTGGGTGTAGTGGCACAT
GCCTGTAATCCCAGCTACTAGGGAGGCTGCGGCAGGAAAATCACTTGAACCTGG
GAGGCGGAGTTTGCAGTGAGCCAAGATCGCGCCATTGCACTGTAGCCTGGGCAA
CAAGAGCAAAACTCTGTCTCAAAATAAATAAATAAATAAATAAATAAATAAAAT
AAATAAAGTTTGGAATGGGAAGGCATTTAATTTTGTTCTTTGAAAATGTGTCAGT
GAGTACCAATATTCAGGTTAACCATTAGAGAGGTTGGGTATATGGAGAAAGCTT
GTCACTTCTGAGTGCCTCAAGTGGTCTTTTTTCCCCTCTCCACTTCCTGATCTTTTC
AGCCTTACTGAAGTAATTTTCTTGTTTTCTTACAGGTTTCACAGATGGCAGTGTTA
CATTGAACAAAGGAGACATCACCCGGGACCGGCATAGCAAGACCCAGATTTTGC
ACAAGGGCAACTATCCTGTAACTGGATTGGCCTTTCGCCAAGCAGGAAAGACCA
CTCACTTGTTTGTTGTGACAACAGAGAACGTCCAGGTATGACCAAGGCCTCCACT
CTTAGGAGCAGGCAGGGAGGGCTTCTCCATTGTTCAGGGGGATAGGGTAATGAA

GTGACAGGAAAGGTGGGAGTGTTAAAGTTTTAATCATATAATTCGAATCATTTGC
CTAGCTTTGTATTTTATTTTTTTGCCTAGGAAGCTGGAAAGAGTTAGACTCTGGGC
TTGTAGTAAAAAGATGTGAATTTTCTTTCTTTTTTTTTTTTGAGACCGAGTCTCAC
TGTGTTGCTCAGGCTGGAGTGCAGTGGCGAGATCTTGACTCACTGTAACCTCAGT
CTCCCAGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAG
GCACACACCAGCATGTCCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCA
CCATGTTGGCCAGTCTGGTCTCCAACTCCTGACCTCAGGTGATCCACCCACCTTG
GCCTTCCAAAGTGCTGAGATTACAGGCGTGAGCCACCGCACTCGGCCTTGAATTT
TATTTCTAACTAGCTTTTTTTTTTTTTTTCTTTGAGACAGGATCTTGCTGTGTAACC
CAGGCTAGAGTGCAATGGTGCTGTCTCAGCTCACTGCAACCTCTGCCTCCTGGGC
TCAAGTGATCCTCCCACCTCAGCCTCCTGAGTACGTGGGATTACAGGCATGTGTC
ACCATGCCTGGCTAATTTTTGTATTTTTTCTAGATATGGGGGTTTCACCATGTTGC
CCAGGCTGATCTCGAACTCCTGGGCCCAAGCAATCCACCTGCCTCAACCACCCAA
AGTGCTGGAATTACAGGCATGAGCCACTGCACCGGCCTATTTTCTAACTAACTTT
TTAAATGACGTTGGACAAGTATCTTTACCTTTCTTGTTCCTGGTGCCCCAGATTTA
CAACAGTCTACTCTCTCAGGAATACTCTTGATAAAGTAGTAGATTATTACATTTTT
CTTACTTGCAGAATTTCTGTACTAGATAGTGATAATATATACAGAGAAAGAAATT
GTGATCTCTGGTTATATATTTAATAGATGGGTAATGGTTAATGAAGAAGGCCAGC
AGCCAAGAGAAAGACTTTAGAATGCCAAGAGGGAAAAAAGAGCCAGTATGGAG
ATGGGAATACCTTTGTGAAGGCTTAGAGTTACCTCCTCAAAGGAGCCTGTTAACC
CCTTTGTTGCTTCTGGCAGTCCTATATAGTTTCTGGAAAAGACTACCCTCGCGTGG
AGTTGGACACCCATGGTTGTGGCCTGCGCTGCTCAGCCCTAAGTGACCCTTCTCA
GGACCTGCAGTTCATTGTGGCCGGGGATGAGTGTGTCTACTTGTACCAGCCTGAT
GAACGTGGGCCCTGCTTCGCCTTTGAGGGCCATAAGCTCATTGCCCACTGGTTTA
GAGGCTACCTTATCATTGTCTCCCGTGACCGGAAGGTTTCTCCCAAGTAAGGACT
CAGTGAGAAGGGACAGGGAGAGGGCTGGACTTGTTCCCCAGAATCCGCCTGATT
TTAAAATCCTAATGCCTGGATACTGCCAATCTCCTACTCCTACTCCGGTGTTATGT
AGGTAACATTTCTGTTTTTTTTTTTTGAGATGGAGTCTCGCTCTGTCACCCAGGAG
TGCAGTGGCGTGATCTCGGCTCACTGCAGCCTCGCCTCCTGGGTTCCAGTGATAC
TCCTGCCTCAGCCTCCTGAATAGCTGGGATTACAGGCATGTGCCACCACGCCTGG
CTACTTTTTGTATTTTCAGTAGAGACGAGGTTTCACCATGTTGGCCAGGCTGGTGT
CAGACTCCTGACCTCAAGTGATCTACCCGCCTCGGCCTCCCAAAGTACTGGGATT
ACAGGTGTGAGCCACTGCACCCAGCCAGTGACACTTATAGCTAGAGTTGGTTAC
AATGCTTTTTTTAACTGAAAAATGGATGTTTTTACTGCAGGTATTTGCTACATAGC

CCTAGCATACTAGAATGCTCCAGGGAATTTGGACCTGCTTTTTTTGTTCTGTTTTG
TTTGAGATGGAGTCTCGCTCTATTGCCCAGGCTGGAGTGCAGTGGCGCGATCTCG
GCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCC
GAGCAACTGGGACTACAGGTGCCCGCCCCCATGCCCGGCTAATTTTTTGTATTTT
TATTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTC
GTGATCTGCCTGCCTCGTCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTG
CGCCCGGCCTTGTTTTGTTTTTTAACTCTTTTGTCAAAATAAATTGGGCAGAGGAA
CAAGGGATTCAGCAGTGGAGAAATTAGAAAGTGCTGGGAAACGTAGGATGCAA
GTTTTCTGTTTTGTTTTGGTTTTATTTTTTTAAGGGGTAAGTGGGGTCTTCTGCATG
ATGAAAGTACTCCCTGTTAGATGCTGGTTTGTTTTGCTTTCATTCAGTTCCAGGCA
CAGTGTTTTACATATACCATGTATTGGATTATTATTTTCTAAATGAATGGATATGT
TAGTGTACTTAATTTTCTAATCTGTTACTTGGGTTAATTGTAGGATCCTAAAGGGT
AAGATCAGGTTTATGTGTTCTCTATGTACAACTACCGGCACAGCGCTGCATGTTA
TGGGATGAGAGGGACCAGAGAGGTGATTTGTGCTGGGTGAAATGTGCACCAAAG
TTATCATGGGAAAGGAAATAGGGGAGATAAGACAGAGATGTCCAAACATCTGGT
GCTTTTTCTTTCCTATGCAGGTCAGAGTTTACCAGCAGGGATTCACAGAGCTCCG
ACAAGCAGATTCTAAACATCTATGACCTGTGCAACAAGTTCATAGCCTATAGCAC
CGTCTTTGAGGATGTAGTGGATGTGCTTGCTGAGTGGGGCTCCCTGTACGTGCTG
ACGCGGGATGGGCGGGTCCACGCACTGCAGGAGAAGGACACACAGACCAAACT
GGAGGCAAGGCCACCAGGCTCGCAGAGCTGGCCACAGGCACCTAGAAGCAGCT
GCAGGAGCAGGTTCCCCAAGTCATATTGGCCAGGGTGTTTCCCTCCTTGTGGGTC
ACAGCCTTACTCAGCTTCCTTAGGGTAGGATTATAAGGTAAATGGCCTGGGATGG
GGCTTTAGAGGGAAGTAGTGGCTGGGAGCCTGAGCCCACTCTAAGGGTTCACTTT
GTTTCCAGAGAACCTTATGAAATAAGATGATACTGATCTTGGGGATATCTGGTAG
CTCTTGGTTTGTCTAGGATCTAGGCAGATCAGAAATCCAGGACTTTCTGTTTTGTG
TGTTGTGCGCACATTTTGGGAAAGTGTCTTCCCAGGACCACTTCTACCAATTTTCT
AATCTCTCTTCCCTTCTAGATGCTGTTTAAGAAGAACCTATTTGAGATGGCGATTA
ACCTTGCCAAGAGCCAGCATCTGGACAGTGATGGGCTGGCCCAGATTTTCATGCA
GTATGGAGACCATCTCTACAGCAAGGGCAACCAC**GATGGGGCTGTCCAGCAAT**
**A**

**Tile 5.1 (3696 bp)**

**CTCTACAGCAAGGGCAACCA**CGATGGGGCTGTCCAGCAATATATCCGGTCAGT
CTGGAGGCACTTTGGGATATAGCTGTGAATGCAGGGACAGGCAGCTAGATAGCT

AAAGCCCATCCATGCTCCAGGTAGGGGCTAGAATTCTACCAGGAACTGTTTTTGT
TTTTGGTTTTTTGTTTGTTTTTGTTAGTTTTTTTTTGAGATGGAGTTTCGCTCTTGTT
GTCCAAGCTGGAGTGCAATGGCGCGATCTCGACTTACTGCAACCTTCCCTCCTGG
GTTCACATGATTCTCCTGCTTCAGTCTCCCGAGTAGCTGGGATTACAGGCGCGTG
CCACCACACCTGGCTAATTTTTTGTATTTTTAGTAGAAACGGGGTTTCACCATGTT
AGCCAGGCTGGTCATGAACTCCTGACCTCAGGTGATCTGCCCGCTTCAGCCTCCC
AAAGTGCTGGGATTACAGGCATGAGCCACTGCGCCCGGTGTTTTTTTCTTTGCTTT
TTTTGTTTGTTTGTTTGTTTTTGAGACAGATTCTCGCTCTTTCACCCAGGCTGGAGT
GCAATGGTGTGATCTCGGCTCACTGCAACCTCTGCCTCGGGTTCAAGCAATTCTC
CTGTCTCAGCCTCCCAAGTAGCTGGGATTACAGGATCATGCTGCCATGCCTGGCT
AATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCACGTTCTCTAGGCTGGTCTC
GAACTCCTGAGCTCAGGCAATCTGCCTGCCTCGGCCTCCCAAAGTGTTAGGATTA
CAGGTGTGAGCCACAGTGCCTGGCCTTTTTAATTTTCTTTATAAGAACTATGATGC
CTTAAATGAGAGAACAAGATTCCTTGTTGAAAAGGGAATTTTAGGGTCAAATATT
ATATGACCAGTATGCAGTTGAGCCGATGATTACCTCTCATTTTAACCTTTAGCTTG
GCCTTTGTAAATATTCACTTTGACCTGGTTCTTTAAGGGGGAAAAGGTATTCATCT
CTTGCTTCCATAATTCCTTCCATCAGTACTAATTCTTGGAATTTTCTGCCTTAAGA
AGGGGTTTTAGAAATCTGTCTAAGGTGGCCGGGCGCGGTAGCTCACGCATGTAA
TCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACGAGGTCAGGAGATCGAGA
CCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAAATACAAAAAATTA
GCTGGGTGTCGTGGCGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAG
GAGAATGGCCTGAACCCCAGAGGCGGAGCTTGCAGTGAGCCGAGATCGCACCAC
TGCACTCCAGCCTGGGTATCAGAGCAGGACTGTCTCAAAAAAAAAAAGTTTGTC
TAAGGCTGGGGTGCAGTGGCATGTGCCTGTGATCCTATCCCTTTGGGAGACCGAG
GTGGGAGGAATGCATGAGCCCAGTCTGAGCAACATAGCAAGACCCCATTTCTAA
CAATAAATAAATAAATAAAATTTAAAAATATGGCTGGGCGCCGTGGCTCATTCCT
GTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGATCACTTGAGGTTAGGAGT
TTGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCCACTAAAAAAAAAAAA
AAAAAAAAAATACAAGGCCAGGCGTGGTGGCTCACGCCTGTAATCTCAGCACTT
TGGGAGGCTGAGGCAGGTGGATCACCTGAGGTCAAGAGTTCAAGACCAGCCTGG
CCAACATGGCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGCATGG
TGGCGGGCGCCTGTAATCCCATCTTAAAAAAAAAAATTAGCTGGGTGTGGTGGC
ATATGCCTGTAATCCCAGCCACTCGCAAAGCTGAGGCAGGAGATTCACTTGAAC
CTGGGAAGCAGAGGCTATAGTGAGCCAAGATCGTGCCACTGCACTCCAGCCTGG

GCAACAGAGCAAGACTCTATCAAAAAAATAATAGTAGATAATAATAATAATAAT
AAATAAATAAATTGCCAGGGGGCAGTGACTCACGCCTGTAATCCAAACACTTTG
GGAGGCTGAGGTGGGCGGATCACGAGGTCAAGAGATTGAGACCATCCTGGCCAA
CATGGTGAAACCCCGTTTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGC
ATGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATAGCTTGAAC
CCAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGTGCCCTGCACTCCAGCCTGGC
GAAATAGCGAAACTTCGTCTCAAAAAAATAATAATAAAAAAAGGCCGGGTGTGG
TGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGAGTGGATCACCTG
AGGTCGGGAGTTCGAAACTAGCCTGACCAACATGGAGAAACCCTGCCTCTACTA
AAAATACAAAACTAGCTGGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCA
GGAGGCTGAGGCAGAAGAATCGCTTGATCCTGGGAGGCAGAGGTTGCGGTGGGC
TGAGATCGTGCCACTGCACTCCAGCCTGGGCAACAGGAGCGAACCTCCCTCTCA
AAAAAAAAAAAATTAATTAAATTTTAAAATAATGTAAAAGAAAGAGAACAAAA
ATACAACAATAAAACAACAACAAAATAGTAATAATATAAAAAGAAACCTGTTTG
GCCTCTTCCCCGTTCTCTTTAGAATTCCTATTTTTAGCCAAATTTAAGAAGCAGTG
CTCTGGAGCCTTGTTACCCAGTGTGGTCCACAGTCTAGTAGCACCGATATTACCT
GGGACTTTGTTAGAAATGTAGAATCTCAGGCCCCACTCCAGATCCACTGAAATAG
AATTTGCATTTTGAGAAGATCTGCAAGTGATTCCTGTGTACATGAGCTGGAGAAG
TACACTGATTCAGATGCTATCGGGTGCACATGTCTCCCTAGAACCATTGGAAAGT
TGGAGCCATCCTACGTGATCCGCAAGTTTCTGGATGCCCAGCGCATTCACAACCT
GACTGCCTACCTGCAGACCCTGCACCGACAATCCCTGGCCAATGCCGACCATACC
ACCCTGCTCCTCAACTGCTATACCAAGCTCAAGGACAGCTCGAAGCTGGAGGAG
TTCATCAAGGTGCAGGATGTTGTTGGTGGGGAAGTCTTGGAGGCCCCACTGAGC
ATGAGGTGGCTCCACCGGGACTTGTTCGTTTCAGCAAGACATAGGGAGAGGAAA
GGGCTGACCTTATTTAGAGGAGTGGCTGTTCCTGTTAGTTCGGTGCCATCTCCCCT
CTTGTTTTATGATTCTAGAAATTCCCATGGTATCTAAGCCCAGGGTGGTGGCACT
GGGGAAGTCCTGTCCCTAGCGGTGTCCCTCTAATGGTAGGGGCAGAGGAAAACT
TTAGTCAGAAACTCCCACAGGCTGAGTAGCATAATATTTTCAAAGTGTGAACGTT
ATGATATCACCTTAGGAATCTGAAAGTCAGGTGCCTGTTTCCTCTCCCTTCTCTAT
CTCCCTCTGTGTAAATGATTTTGTCTCATGTCTCCCTGGCAGAAAAAGAGTGAGA
GTGAAGTCCACTTTGATGTGGAGACAGCCATCAAGGTCCTCCGGCAGGCTGGCT
ACTACTCCCATGC**CCTGTATCTGGCGGAGAACC**

**Tile 5.2 (3990 bp)**

**GTCTGGAGGCACTTTGGGAT**ATAGCTGTGAATGCAGGGACAGGCAGCTAGATA
GCTAAAGCCCATCCATGCTCCAGGTAGGGGCTAGAATTCTACCAGGAACTGTTTT
TGTTTTTGGTTTTTTGTTTGTTTTTGTTAGTTTTTTTTTGAGATGGAGTTTCGCTCTT
GTTGTCCAAGCTGGAGTGCAATGGCGCGATCTCGACTTACTGCAACCTTCCCTCC
TGGGTTCACATGATTCTCCTGCTTCAGTCTCCCGAGTAGCTGGGATTACAGGCGC
GTGCCACCACACCTGGCTAATTTTTTGTATTTTTAGTAGAAACGGGGTTTCACCAT
GTTAGCCAGGCTGGTCATGAACTCCTGACCTCAGGTGATCTGCCCGCTTCAGCCT
CCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCGCCCGGTGTTTTTTTCTTTG
CTTTTTTTGTTTGTTTGTTTGTTTTTGAGACAGATTCTCGCTCTTTCACCCAGGCTG
GAGTGCAATGGTGTGATCTCGGCTCACTGCAACCTCTGCCTCGGGTTCAAGCAAT
TCTCCTGTCTCAGCCTCCCAAGTAGCTGGGATTACAGGATCATGCTGCCATGCCT
GGCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCACGTTCTCTAGGCTGG
TCTCGAACTCCTGAGCTCAGGCAATCTGCCTGCCTCGGCCTCCCAAAGTGTTAGG
ATTACAGGTGTGAGCCACAGTGCCTGGCCTTTTTAATTTTCTTTATAAGAACTATG
ATGCCTTAAATGAGAGAACAAGATTCCTTGTTGAAAAGGGAATTTTAGGGTCAA
ATATTATATGACCAGTATGCAGTTGAGCCGATGATTACCTCTCATTTTAACCTTTA
GCTTGGCCTTTGTAAATATTCACTTTGACCTGGTTCTTTAAGGGGGAAAAGGTAT
TCATCTCTTGCTTCCATAATTCCTTCCATCAGTACTAATTCTTGGAATTTTCTGCCT
TAAGAAGGGGTTTTAGAAATCTGTCTAAGGTGGCCGGGCGCGGTAGCTCACGCA
TGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATCACGAGGTCAGGAGAT
CGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAAATACAAAA
AATTAGCTGGGTGTCGTGGCGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGA
GGCAGGAGAATGGCCTGAACCCCAGAGGCGGAGCTTGCAGTGAGCCGAGATCGC
ACCACTGCACTCCAGCCTGGGTATCAGAGCAGGACTGTCTCAAAAAAAAAAAGT
TTGTCTAAGGCTGGGGTGCAGTGGCATGTGCCTGTGATCCTATCCCTTTGGGAGA
CCGAGGTGGGAGGAATGCATGAGCCCAGTCTGAGCAACATAGCAAGACCCCATT
TCTAACAATAAATAAATAAATAAAATTTAAAAATATGGCTGGGCGCCGTGGCTC
ATTCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGGATCACTTGAGGTTA
GGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCCACTAAAAAAA
AAAAAAAAAAAAAAAATACAAGGCCAGGCGTGGTGGCTCACGCCTGTAATCTCAG
CACTTTGGGAGGCTGAGGCAGGTGGATCACCTGAGGTCAAGAGTTCAAGACCAG
CCTGGCCAACATGGCGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGG
CATGGTGGCGGGCGCCTGTAATCCCATCTTAAAAAAAAAAATTAGCTGGGTGTG
GTGGCATATGCCTGTAATCCCAGCCACTCGCAAAGCTGAGGCAGGAGATTCACTT

GAACCTGGGAAGCAGAGGCTATAGTGAGCCAAGATCGTGCCACTGCACTCCAGC
CTGGGCAACAGAGCAAGACTCTATCAAAAAAATAATAGTAGATAATAATAATAA
TAATAAATAAATAAATTGCCAGGGGGCAGTGACTCACGCCTGTAATCCAAACAC
TTTGGGAGGCTGAGGTGGGCGGATCACGAGGTCAAGAGATTGAGACCATCCTGG
CCAACATGGTGAAACCCCGTTTCTACTAAAAATACAAAAATTAGCTGGGTGTGGT
GGCATGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATAGCTTG
AACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGTGCCCTGCACTCCAGCCT
GGCGAAATAGCGAAACTTCGTCTCAAAAAAATAATAATAAAAAAAGGCCGGGTG
TGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGAGTGGATCAC
CTGAGGTCGGGAGTTCGAAACTAGCCTGACCAACATGGAGAAACCCTGCCTCTA
CTAAAAATACAAAACTAGCTGGGCATGGTGGCACATGCCTGTAATCCCAGCTAC
TCAGGAGGCTGAGGCAGAAGAATCGCTTGATCCTGGGAGGCAGAGGTTGCGGTG
GGCTGAGATCGTGCCACTGCACTCCAGCCTGGGCAACAGGAGCGAACCTCCCTC
TCAAAAAAAAAAAAATTAATTAAATTTTAAAATAATGTAAAAGAAAGAGAACAA
AAATACAACAATAAAACAACAACAAAATAGTAATAATATAAAAAGAAACCTGTT
TGGCCTCTTCCCCGTTCTCTTTAGAATTCCTATTTTTAGCCAAATTTAAGAAGCAG
TGCTCTGGAGCCTTGTTACCCAGTGTGGTCCACAGTCTAGTAGCACCGATATTAC
CTGGGACTTTGTTAGAAATGTAGAATCTCAGGCCCCACTCCAGATCCACTGAAAT
AGAATTTGCATTTTGAGAAGATCTGCAAGTGATTCCTGTGTACATGAGCTGGAGA
AGTACACTGATTCAGATGCTATCGGGTGCACATGTCTCCCTAGAACCATTGGAAA
GTTGGAGCCATCCTACGTGATCCGCAAGTTTCTGGATGCCCAGCGCATTCACAAC
CTGACTGCCTACCTGCAGACCCTGCACCGACAATCCCTGGCCAATGCCGACCATA
CCACCCTGCTCCTCAACTGCTATACCAAGCTCAAGGACAGCTCGAAGCTGGAGG
AGTTCATCAAGGTGCAGGATGTTGTTGGTGGGGAAGTCTTGGAGGCCCCACTGA
GCATGAGGTGGCTCCACCGGGACTTGTTCGTTTCAGCAAGACATAGGGAGAGGA
AAGGGCTGACCTTATTTAGAGGAGTGGCTGTTCCTGTTAGTTCGGTGCCATCTCC
CCTCTTGTTTTATGATTCTAGAAATTCCCATGGTATCTAAGCCCAGGGTGGTGGC
ACTGGGGAAGTCCTGTCCCTAGCGGTGTCCCTCTAATGGTAGGGGCAGAGGAAA
ACTTTAGTCAGAAACTCCCACAGGCTGAGTAGCATAATATTTTCAAAGTGTGAAC
GTTATGATATCACCTTAGGAATCTGAAAGTCAGGTGCCTGTTTCCTCTCCCTTCTC
TATCTCCCTCTGTGTAAATGATTTTGTCTCATGTCTCCCTGGCAGAAAAAGAGTG
AGAGTGAAGTCCACTTTGATGTGGAGACAGCCATCAAGGTCCTCCGGCAGGCTG
GCTACTACTCCCATGCCCTGTATCTGGCGGAGAACCATGCACATCATGAGTGGTA
CCTGAAGATCCAGCTAGAAGACATTAAGGTAGGTGAGACTGTCTTTTTTCCCCAA

GAGACAGGGTCTCACTATGTTGCCCAGGCTGGAGTTGAATTCCTGGGCTCAGGTG
ATCCTCCCACCCCATCCTCCTGAGTGGAGGTGAGACTCTCAACTTGGGCAGAGCC
CTGTCGTTTTTGAAGACATCTCCAGAATGGTGAAAGTGCTTCCTGGTGGCTGAGG
CAGGAGTATACACTATTCTGCCCTTTACTTTTCCACAGAATTATCAGGAAGCCCT
TCGATACATCGGCAAGCTGCCTTTTGAGCAG**GCAGAGAGCAACATGAAGCG**

**Tile 8.1 (3096 bp)**

**CCCCTTTGGGAAGACACGTT**CCACTTTTGATTCATAGGAGAGAGTATCAGCCA
AGCCTCCGAACTGCACACAAACGTCTTAGAAGTGCGCCTTCTTTTTGTGTTATAG
TGGTCTCCCAGCCACAGCCAACGCTCCAAGTCCCCAGCTGTGACACACCTACTGA
ATTACTACCGTGGGTGGGAGGCCGCCGTGGGCCTTTCCATTACGAGCCTGCTTGC
CGAGCCCTGGGCTTGTGCACAGACAAACTGCAGAGCTGGTGGAGGCCACTGCCA
GGCCGAGATAAGAAAGAGATGGGGAGCTGCTAATCTCCCCCTGTCCAGCCTGTT
GGTGAGGGCTGGGATCTTTGCTCTTGCAGTCATTCCAGAGCCCTGGACTAGGAGT
AGGAAGATCTGAATTGTGGCCCCAACTCTCTTTCGGTTATTAGCTCTGTGACCCT
AGGCAAGTCACCTCATCCCTTGATGCCACCCGTTGCTTCTGTAACATGGTCCCAA
AGGTGCCTGTCTTGTCCACCTGATAGGATTTTTGAGACGACAACAATATGCAAAA
GCAATAGCTTCAACATAGAAGTGCTCAGTGTTTTATTTTTTAATGAAACGGTTTG
ACTTGGATATGCTGTGCACATTCAATGAACTTAAGGAATTGTTTGAACCTAGTAG
TTCTGGGACCTTAGAGTCCTTTCTGTGGGCTCCCTGTGGCCCAGAATTTTGGTGGC
CACGTTTAATATCAAGCCTAGCCTAATTTGCAAAGGGTCTCCCAGGGTTAATTTA
TTGGAGTGATCACATGGAGTAGACCAGAGTCTGAGGGCAGAAAGCTGTCACCTG
CTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCATAGCACCCC
GACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAGGCCGGGCA
CAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAGGTGGATTGC
CTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACCCCGTTTCTA
CTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGTAGTCCCAGCTA
CTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGAGGTTGCACT
GAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGCGACTCCCCC
TCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACACACACACATA
TTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTTGGGAGGCT
GAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAGCTGAGAAC
ATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCAAAAAAAAA
AAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATGTTGACTGTCG

ACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAATGGGATCAG
GCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGGGGCAG
GAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCCGTCTCTAC
TAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTCCAGCTTGA
ACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACTCCAGCCTG
AGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAGAGAGAAATGG
GACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTGAATACTTT
TGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTCTTGACTTG
GGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCACTCTAGGT
GCACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGAGGCAGAA
GAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCTGGTGGGA
CAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACGTCAGATC
CGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACCCTTAATA
CTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGTAGCCAGA
CATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACTCTAGTCT
ACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTCTGTTGTA
GATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGGTCCTGGC
TGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCTGATGGGC
CTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACTGCCTTGGT
CAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCAACAGCAG
GTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGAGGGCAG
AAGGTACTGAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCCACTCTGAG
ACTCAGCTGGCCACAGGCCAGGTTCTGAAGTTCCTTTCTTCCAAGCCAGTGATTC
TGGTTCTTGGACAAGGTGTTGAGGAACACTAGAAACAGAGGGGACTGTGACCTG
GGGACTTTTTCTGCAGGAAGAAAACAGCCCAAAGATGAGAGTGATTCGCGTGGG
TACCCGCAAGAGCCAGGTGGGTGCAGGAGCCGGGGTGGAGGAGGTTTGTCAGAA
CAGTTATGATGCTCACAGCATCACAAATTGGGGGACTCAGAGGGTTAGTTCCTAG
TATGAAGGAGATGGGGTGGCTGGGCGTTAAGTTCCCCGGGAAATGGCAGATTAC
ATTCTATGGCAAGATCATCCCTAGGCTGGGAAAATTGTTGGAGTGCAGAGGGCT
CCCAAGCCCCTTCTCATGCCCAGATGGAAATTCCAGTCCCTTCA**GGATCTGCCT**
**AACCTGTGAC**

**Tile 8.2 (3729 bp)**

**AGAGCAAAGGAAGCGCCATA**GAAGCTGCACTACTTGCTCATGTCACAGCTGGG
GAATGGGGTGGTCGAATGGGGAGGTCCACTGTCGCAATGTTCCAATTCCCGCCC
AGAGGGAGGGACCTCCCCTTCGAGGGAGGGCGCCGGAAGTGACGCGAGGCTCTG
CGGAGACCAGGAGTCAGACTGTAGGACGACCTCGGGTCCCACGTGTCCCCGGTA
CTCGCCGGCCGGAGCCCCCGGCTTCCCGGGGCCGGGGGACCTTAGCGGCACCCA
CACACAGCCTACTTTCCAAGCGGAGCCATGTCTGGTAACGGCAATGCGGCTGCA
ACGGCGGTGAGTGCTGAGCCGGTGACCAGCACACTTTGGGCTTCTGGACGAGCC
GTGCAGCGATTGGCCCCAGGTTGCCATCCTCAGTCGTCTATTGGTCAGAACGGCT
ATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTCCGAGTAGCTTTTAAAGGGCCAG
TAGCTCGGTTGCCCTCCGGAAGGAATGGGGAAATCAGAGAGCGGTGATACTGGG
TTAAGAGTGGAAGGATTGTTTGGAACGGAACTCCGGTCCCTGCGGGCATCTGGG
TGGGATTCCCATCAGGCCTGGGATGCACGGCTCTAGATTTAGTGACCCAGACCAA
GAACGTTCGTCTACACAGACGGGGTCCTTTCATTCGAGGCTGGGCTGAGGCGGAT
GCAGATACGGCCCCTTTGGGAAGACACGTTCCACTTTTGATTCATAGGAGAGAGT
ATCAGCCAAGCCTCCGAACTGCACACAAACGTCTTAGAAGTGCGCCTTCTTTTTG
TGTTATAGTGGTCTCCCAGCCACAGCCAACGCTCCAAGTCCCCAGCTGTGACACA
CCTACTGAATTACTACCGTGGGTGGGAGGCCGCCGTGGGCCTTTCCATTACGAGC
CTGCTTGCCGAGCCCTGGGCTTGTGCACAGACAAACTGCAGAGCTGGTGGAGGC
CACTGCCAGGCCGAGATAAGAAAGAGATGGGGAGCTGCTAATCTCCCCCTGTCC
AGCCTGTTGGTGAGGGCTGGGATCTTTGCTCTTGCAGTCATTCCAGAGCCCTGGA
CTAGGAGTAGGAAGATCTGAATTGTGGCCCCAACTCTCTTTCGGTTATTAGCTCT
GTGACCCTAGGCAAGTCACCTCATCCCTTGATGCCACCCGTTGCTTCTGTAACAT
GGTCCCAAAGGTGCCTGTCTTGTCCACCTGATAGGATTTTTGAGACGACAACAAT
ATGCAAAAGCAATAGCTTCAACATAGAAGTGCTCAGTGTTTTATTTTTTAATGAA
ACGGTTTGACTTGGATATGCTGTGCACATTCAATGAACTTAAGGAATTGTTTGAA
CCTAGTAGTTCTGGGACCTTAGAGTCCTTTCTGTGGGCTCCCTGTGGCCCAGAAT
TTTGGTGGCCACGTTTAATATCAAGCCTAGCCTAATTTGCAAAGGGTCTCCCAGG
GTTAATTTATTGGAGTGATCACATGGAGTAGACCAGAGTCTGAGGGCAGAAAGC
TGTCACCTGCTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCAT
AGCACCCCGACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAG
GCCGGGCACAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAG
GTGGATTGCCTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACC
CCGTTTCTACTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGTAGT
CCCAGCTACTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGAG

GTTGCACTGAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGCG
ACTCCCCCTCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACACA
CACACATATTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTT
GGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAG
CTGAGAACATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCA
AAAAAAAAAAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATGT
TGACTGTCGACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAAT
GGGATCAGGCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
TGGGGCAGGAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCC
GTCTCTACTAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTC
CAGCTTGAACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACT
CCAGCCTGAGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAGAG
AGAAATGGGACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTG
AATACTTTTGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTC
TTGACTTGGGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCA
CTCTAGGTGCACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGA
GGCAGAAGAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCT
GGTGGGACAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACG
TCAGATCCGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACC
CTTAATACTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGT
AGCCAGACATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACT
CTAGTCTACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTC
TGTTGTAGATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGG
TCCTGGCTGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCT
GATGGGCCTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACT
GCCTTGGTCAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCA
ACAGCAGGTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGA
GGGCAGAAGGTACTGAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCCAC
TCTGAGACTCAGCTGGCCACAGGCCAGGTTCTGAAGTTCCTTTCTTCCAAGCCAG
TGATTCTGGTTCTTGGACAAGGTGTTGAGGAACACTAGAAACAGAGGGGACTGT
GACCTGGGGACTTTTTCTGCAGGAAGAAAACAGCCCAAAGATGAGAGTGATTCG
CGTGGGTACCCGCAAGAGCCAGGTGGGTGCAGGAGCCGGGGTGGAGGAGGTTTG
TCAGAACAGTTATGATGCTCACAGCATCACAAATTGGGGGACTCAGAGGGTTAG

TTCCTAGTATGAAGGAGATGGGGTGGCTGGGCGTTAAGTTCCCCGGGAA**ATGGC**
**AGATTACATTCTATGGCA**

**Tile 8.3 (3726 bp)**

**AGGTCCACTGTCGCAATGTT**CCAATTCCCGCCCAGAGGGAGGGACCTCCCCTTC
GAGGGAGGGCGCCGGAAGTGACGCGAGGCTCTGCGGAGACCAGGAGTCAGACT
GTAGGACGACCTCGGGTCCCACGTGTCCCCGGTACTCGCCGGCCGGAGCCCCCG
GCTTCCCGGGGCCGGGGGACCTTAGCGGCACCCACACACAGCCTACTTTCCAAG
CGGAGCCATGTCTGGTAACGGCAATGCGGCTGCAACGGCGGTGAGTGCTGAGCC
GGTGACCAGCACACTTTGGGCTTCTGGACGAGCCGTGCAGCGATTGGCCCCAGG
TTGCCATCCTCAGTCGTCTATTGGTCAGAACGGCTATCTTTTTTTTTTTTTTTTT
TTTTTTTTTGGTCCGAGTAGCTTTTAAAGGGCCAGTAGCTCGGTTGCCCTCCGGA
AGGAATGGGGAAATCAGAGAGCGGTGATACTGGGTTAAGAGTGGAAGGATTGTT
TGGAACGGAACTCCGGTCCCTGCGGGCATCTGGGTGGGATTCCCATCAGGCCTG
GGATGCACGGCTCTAGATTTAGTGACCCAGACCAAGAACGTTCGTCTACACAGA
CGGGGTCCTTTCATTCGAGGCTGGGCTGAGGCGGATGCAGATACGGCCCCTTTGG
GAAGACACGTTCCACTTTTGATTCATAGGAGAGAGTATCAGCCAAGCCTCCGAA
CTGCACACAAACGTCTTAGAAGTGCGCCTTCTTTTTGTGTTATAGTGGTCTCCCAG
CCACAGCCAACGCTCCAAGTCCCCAGCTGTGACACACCTACTGAATTACTACCGT
GGGTGGGAGGCCGCCGTGGGCCTTTCCATTACGAGCCTGCTTGCCGAGCCCTGGG
CTTGTGCACAGACAAACTGCAGAGCTGGTGGAGGCCACTGCCAGGCCGAGATAA
GAAAGAGATGGGGAGCTGCTAATCTCCCCCTGTCCAGCCTGTTGGTGAGGGCTG
GGATCTTTGCTCTTGCAGTCATTCCAGAGCCCTGGACTAGGAGTAGGAAGATCTG
AATTGTGGCCCCAACTCTCTTTCGGTTATTAGCTCTGTGACCCTAGGCAAGTCAC
CTCATCCCTTGATGCCACCCGTTGCTTCTGTAACATGGTCCCAAAGGTGCCTGTCT
TGTCCACCTGATAGGATTTTTGAGACGACAACAATATGCAAAAGCAATAGCTTCA
ACATAGAAGTGCTCAGTGTTTTATTTTTTAATGAAACGGTTTGACTTGGATATGCT
GTGCACATTCAATGAACTTAAGGAATTGTTTGAACCTAGTAGTTCTGGGACCTTA
GAGTCCTTTCTGTGGGCTCCCTGTGGCCCAGAATTTTGGTGGCCACGTTTAATATC
AAGCCTAGCCTAATTTGCAAAGGGTCTCCCAGGGTTAATTTATTGGAGTGATCAC
ATGGAGTAGACCAGAGTCTGAGGGCAGAAAGCTGTCACCTGCTTCGGCAATAGA
GGCCCCAGATGTCTGGGTGCAAAAGAACTCCATAGCACCCCGACCAACATGGTG
AAACCCCGTCTCTACTAAAAATATAAAAATTAGGCCGGGCACAGTGGCTCATGC
CTGTAATCCTAGCACTTTGGGAGGCCGAGGCAGGTGGATTGCCTGAGCTCAGGA

GTTCGAGACCAGCCTAGGGAACACAGTGAAACCCCGTTTCTACTAAAAATACAA
AAAATTAGCCGACGTGGTGGCATGCGCCTGTAGTCCCAGCTACTTGGGAGGCTA
AGACAGGAGAATCGCTTGAACCTGGGAGGTGGAGGTTGCACTGAGCCGAGACCG
CGCCATTGCACTCCAGCCTGGGTGACAGAGCGCGACTCCCCCTCAAAAAAAGAA
AAAAAAAAAAAATATATATATATATATACACACACACACATATTTTAGCTGGGC
ATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTTGGGAGGCTGAGTCAGGAGA
ATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAGCTGAGAACATGCCACTGCA
CTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCAAAAAAAAAAAAAAAAGGA
ACTACATAGGATGAACATCCCAGATCAGGGAATGTTGACTGTCGACAGTATCAG
TATCTACAGTGGCTACTGTCTGATGTAGAAAGAAATGGGATCAGGCTAGGCGTG
GTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCTGGGGCAGGAGGATCACA
AGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCCGTCTCTACTAAAAATGTG
AAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTCCAGCTTGAACCCAGGGG
TGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACTCCAGCCTGAGCAAAACA
GTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAGAGAGAAATGGGACCTCCGT
CTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTGAATACTTTTGTCCAAGG
TACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTCTTGACTTGGGCTCTGGC
CTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCACTCTAGGTGCACGAACT
TTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGAGGCAGAAGAAACCTGA
GTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCTGGTGGGACAGTGTACC
CAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACGTCAGATCCGTTTACCT
CAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACCCTTAATACTGTCAGCT
TTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGTAGCCAGACATCTGGGC
CTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACTCTAGTCTACTCCATGT
GGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTCTGTTGTAGATAATGA
ATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGGTCCTGGCTGCCCTGC
CCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCTGATGGGCCTTATCTC
TTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACTGCCTTGGTCAAGGTG
GGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCAACAGCAGGTCCTACT
ATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGAGGGCAGAAGGTACT
GAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCCACTCTGAGACTCAGCT
GGCCACAGGCCAGGTTCTGAAGTTCCTTTCTTCCAAGCCAGTGATTCTGGTTCTT
GGACAAGGTGTTGAGGAACACTAGAAACAGAGGGGACTGTGACCTGGGGACTTT
TTCTGCAGGAAGAAAACAGCCCAAAGATGAGAGTGATTCGCGTGGGTACCCGCA

AGAGCCAGGTGGGTGCAGGAGCCGGGGTGGAGGAGGTTTGTCAGAACAGTTATG
ATGCTCACAGCATCACAAATTGGGGGACTCAGAGGGTTAGTTCCTAGTATGAAG
GAGATGGGGTGGCTGGGCGTTAAGTTCCCCGGGAAATGGCAGATTACATTCTAT
GGCAAGATCATCCCTAGGCTGGGAAAATTGTTGGAGTGCAGAGGGCT**CCCAAGC**
**CCCTTCTCATG**

**Supplementary figure 10.5:**

Genomic sequence of the mutant luciferase constructs, including tile 8a_protective type and the five mutants. The sequences were retrieved from SnapGene viewer web tool. Marked in bold is the forward and reverse primers respectively, and the red colored bold letter is the mutant (from protective sequence to risk sequence).

**Tile 8a_WT_protective**

TGTCACCTGCTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCAT
AGCACCCCGACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAG
GCCG**G**GCACAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAG
GTGGATTGCCTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACC
CCGTTTCTACTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTG**T**AGT
CCCAGCTACTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGAG
GTTGCACTGAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGC**G**
ACTCCCCCTCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACACA
CACACATATTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTT
GGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAG
CTGAGAACATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCA
AAAAAAAAAAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATGT
TGACTGTCGACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAAT
GGGATCAGGCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
TGGGGCAGGAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCC
GTCTCTACTAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTC
CAGCTTGAACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACT
CCAGCCTGAGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAA**AG**AG
AGAAATGGGACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTG
AATACTTTTGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTC
TTGACTTGGGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCA

CTCTAGGTGCACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGA
GGCAGAAGAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCT
GGTGGGACAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACG
TCAGATCCGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACC
CTTAATACTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGT
AGCCAGACATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACT
CTAGTCTACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTC
TGTTGTAGATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGG
TCCTGGCTGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCT
GATGGGCCTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACT
GCCTTGGTCAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCA
ACAGCAGGTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGA
GGGCAGAAGGTACTGAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCC

**Tile 8a_rs1799991_G>A**

TGTCACCTGCTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCAT
AGCACCCCGACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAG
GCCGAGCACAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAG
GTGGATTGCCTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACC
CCGTTTCTACTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGTAGT
CCCAGCTACTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGAG
GTTGCACTGAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGCG
ACTCCCCCTCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACACA
CACACATATTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTT
GGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAG
CTGAGAACATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCA
AAAAAAAAAAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATGT
TGACTGTCGACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAAT
GGGATCAGGCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
TGGGGCAGGAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCC
GTCTCTACTAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTC
CAGCTTGAACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACT
CCAGCCTGAGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAGAG
AGAAATGGGACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTG

AATACTTTTGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTC
TTGACTTGGGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCA
CTCTAGGTGCACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGA
GGCAGAAGAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCT
GGTGGGACAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACG
TCAGATCCGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACC
CTTAATACTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGT
AGCCAGACATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACT
CTAGTCTACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTC
TGTTGTAGATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGG
TCCTGGCTGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCT
GATGGGCCTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACT
GCCTTGGTCAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCA
ACAGCAGGTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGA
GGGCAGAAGGTACTGAGGAAGGTTAAA

**Tile 8a_rs1799992_T > C**

TGTCACCTGCTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCAT
AGCACCCCGACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAG
GCCGGGCACAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAG
GTGGATTGCCTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACC
CCGTTTCTACTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGCAG
TCCCAGCTACTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGA
GGTTGCACTGAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGC
GACTCCCCCTCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACAC
ACACACATATTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACT
TGGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTA
GCTGAGAACATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTC
AAAAAAAAAAAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATG
TTGACTGTCGACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAA
TGGGATCAGGCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGG
CTGGGGCAGGAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCC
CGTCTCTACTAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTT
CCAGCTTGAACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCAC

TCCAGCCTGAGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAGA
GAGAAATGGGACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGT
GAATACTTTTGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTT
CTTGACTTGGGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGC
ACTCTAGGTGCACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTG
AGGCAGAAGAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATC
CTGGTGGGACAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTA
CGTCAGATCCGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGA
CCCTTAATACTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTT
GTAGCCAGACATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGA
CTCTAGTCTACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTG
TCTGTTGTAGATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGT
GGTCCTGGCTGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGG
CTGATGGGCCTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAA
CTGCCTTGGTCAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAG
CAACAGCAGGTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAG
GAGGGCAGAAGGTACTGAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCC

**Tile 8a_rs2846282_G>A**

TGTCACCTGCTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCAT
AGCACCCCGACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAG
GCCGGGCACAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAG
GTGGATTGCCTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACC
CCGTTTCTACTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGTAGT
CCCAGCTACTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGAG
GTTGCACTGAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGCA
ACTCCCCCTCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACACA
CACACATATTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTT
GGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAG
CTGAGAACATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCA
AAAAAAAAAAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATGT
TGACTGTCGACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAAT
GGGATCAGGCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
TGGGGCAGGAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCC

GTCTCTACTAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTC
CAGCTTGAACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACT
CCAGCCTGAGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAGAG
AGAAATGGGACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTG
AATACTTTTGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTC
TTGACTTGGGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCA
CTCTAGGTGCACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGA
GGCAGAAGAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCT
GGTGGGACAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACG
TCAGATCCGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACC
CTTAATACTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGT
AGCCAGACATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACT
CTAGTCTACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTC
TGTTGTAGATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGG
TCCTGGCTGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCT
GATGGGCCTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACT
GCCTTGGTCAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCA
ACAGCAGGTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGA
GGGCAGAAGGTACTGAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCC

**Tile 8a_rs11297272_G>-**

TGTCACCTGCTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCAT
AGCACCCCGACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAG
GCCGGGCACAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAG
GTGGATTGCCTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACC
CCGTTTCTACTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGTAGT
CCCAGCTACTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGAG
GTTGCACTGAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGCG
ACTCCCCCTCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACACA
CACACATATTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTT
GGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAG
CTGAGAACATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCA
AAAAAAAAAAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATGT
TGACTGTCGACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAAT

GGGATCAGGCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
TGGGGCAGGAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCC
GTCTCTACTAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTC
CAGCTTGAACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACT
CCAGCCTGAGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAAGA
GAAATGGGACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTGA
ATACTTTTGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTCT
TGACTTGGGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCAC
TCTAGGTGCACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGAG
GCAGAAGAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCTG
GTGGGACAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACGT
CAGATCCGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACCC
TTAATACTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGTA
GCCAGACATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACTC
TAGTCTACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTCT
GTTGTAGATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGGT
CCTGGCTGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCTG
ATGGGCCTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACTG
CCTTGGTCAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCAA
CAGCAGGTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGAG
GGCAGAAGGTACTGAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCC

**Tile 8a_rs1799993 C>A**

TGTCACCTGCTTCGGCAATAGAGGCCCCAGATGTCTGGGTGCAAAAGAACTCCAT
AGCACCCCGACCAACATGGTGAAACCCCGTCTCTACTAAAAATATAAAAATTAG
GCCGGGCACAGTGGCTCATGCCTGTAATCCTAGCACTTTGGGAGGCCGAGGCAG
GTGGATTGCCTGAGCTCAGGAGTTCGAGACCAGCCTAGGGAACACAGTGAAACC
CCGTTTCTACTAAAAATACAAAAAATTAGCCGACGTGGTGGCATGCGCCTGTAGT
CCCAGCTACTTGGGAGGCTAAGACAGGAGAATCGCTTGAACCTGGGAGGTGGAG
GTTGCACTGAGCCGAGACCGCGCCATTGCACTCCAGCCTGGGTGACAGAGCGCG
ACTCCCCCTCAAAAAAAGAAAAAAAAAAAAAATATATATATATATATACACACA
CACACATATTTTAGCTGGGCATGGTGGTGTGCGTCTGTAGTAGTCCCAGCTACTT
GGGAGGCTGAGTCAGGAGAATCGCTTGAACCTGGAAGGCAGTGGTTGTAGTTAG

CTGAGAACATGCCACTGCACTCCAGCCTGGGCAACAGAGGGAGACTCTGTCTCA
AAAAAAAAAAAAAAAGGAACTACATAGGATGAACATCCCAGATCAGGGAATGT
TGACTGTCGACAGTATCAGTATCTACAGTGGCTACTGTCTGATGTAGAAAGAAAT
GGGATCAGGCTAGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
TGGGGCAGGAGGATCACAAGTTCGAGACCAGCCTGGCCAACACAGTGAAACCCC
GTCTCTACTAAAAATGTGAAAATTAGCTGGGCATGGTGGAACATGCCTGTAGTTC
CAGCTTGAACCCAGGGGTGGAGGTTGTAGTGAGCCTAGATCACGCCACTGCACT
CCAGCCTGAGCAAAACAGTGAGACTCTGTCTAAAAAAAAAAAAAAAAAAAGAG
AGAAATGGGACCTCCGTCTTAGACTGAAGAATTCAGTTCTACGTGCTTAGCAGTG
AATACTTTTGTCCAAGGTACTCTGGCAGGAGGAAGAGGCGTGTCCTCTTGAGTTC
TTGACTTGGGCTCTGGCCTGTTAATATTTCCATGTTGGTGAAACCAGAGGCAGCA
CTCTAGGTGAACGAACTTTAGGCAGCGCAGCCTCCTAGTCTTATGGAACATCTGA
GGCAGAAGAAACCTGAGTCCAACCTTTTCATTTTATAGATGAACAAACAGATCCT
GGTGGGACAGTGTACCCAAGGTCACCCAGCCAAGAGGCTGAGCAGGACTGTACG
TCAGATCCGTTTACCTCAGTCCTTAATGCATGCAGTCCAGCCAGATTAAGGGACC
CTTAATACTGTCAGCTTTCCCCACTGTGGGATCTTCATCCTCTTGACTTCTTTTGT
AGCCAGACATCTGGGCCTCTTGCTGGAGAAGGTGGCAGCTTGCTGCTCTTAGACT
CTAGTCTACTCCATGTGGCATCTGGATGGCACTGAAATTTTCTCAAGTGCCTTGTC
TGTTGTAGATAATGAATCTATCCTCCAGTGACTCAGCACAGGTTCCCCAGTGTGG
TCCTGGCTGCCCTGCCCCTGCCAGCTGCAGGCCCCACCCTTCCTGTGGCCAGGCT
GATGGGCCTTATCTCTTTACCCACCTGGCTGTGCACAGCACTCCCACTGACAACT
GCCTTGGTCAAGGTGGGCTTCAGGGCTCAGTGTCCTGGTTACTGCAGCGGCAGCA
ACAGCAGGTCCTACTATCGCCTCCCTCTAGTCTCTGCTTCTCTGGATCCCTGAGGA
GGGCAGAAGGTACTGAGGAAGGTTAAAGGGACCAGCCTTGGAGTATTTCCCC