THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOBEHAVIORAL HEALTH

Characterization of the Promoter of *SLC6A3,* A Gene Relevant to Substance Use Disorder

KATHLEEN CARDONE
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biobehavioral Health
with honors in Biobehavioral Health

Reviewed and approved* by the following:

David Vandenbergh
Professor of Biobehavioral Health
Thesis Supervisor

Helen Kamens
Associate Professor of Biobehavioral Health
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Substance use disorder is a serious health issue. In 2020, 40.3 million Americans had a substance use disorder (Substance Abuse and Mental Health Services Administration, 2021) and 92,000 individuals died of overdose (Ahmad et al., 2021). Genetics play a role in substance use disorder (Goldman et al., 2005) and one gene, the human dopamine transporter gene or Solute Carrier, Family 3, Subfamily A, member 3 (*SLC6A3*), has been associated with substance use disorders (Reith et al., 2021). A few Variable Number of Tandem Repeats (VNTRs) loci have been identified in and around this gene that are suggested to differentially affect gene expression by acting through the gene's promoter (Reith et al., 2021); however, little attention has been paid to shorter repeats (short tandem repeats or STRs). Given the role a gene's promoter plays in gene expression (Tirosh et al., 2009), the first goal of this study was to identify whether an STR within the *SLC6A3* promoter is variable. This STR has previously been identified in the Repeat Masker Track of the Genome Browser (accessed from http://genome.ucsc.edu). Although the *SLC6A3* promoter has been studied, studies assessing variability of this STR or association studies with disease phenotypes were not described in the literature, so two methods were used to test for genetic variation in the STR. The first utilized polymerase chain reaction (PCR) and gel electrophoresis, which showed that the number of repeats does not vary between individuals. In the second method, 62 publicly-available-long-read genome sequences were aligned. Analysis confirmed that there was no variation in the number of repeats between alleles, but one single nucleotide polymorphism (SNP) and two novel insertions/deletions (in/dels) were identified. Recognition that this region was GC rich led to a second hypothesis that G-Quadruplexes (G4s) form in this STR (Kolesnikova and Curtis, 2019). Thus, the second goal of the project was to identify whether G-Quadruplexes formed, which was tested with an online prediction software

and by circular dichroism (CD) spectroscopy. The CD Spectra showed that G4s form in this STR. These findings may have implications for gene expression because transcription factors bind to G4s (Stevens et al., 2017). Additionally, some variants identified are within transcription factor binding sites, further supporting a role for genetic variation in the promoter as relevant to regulation of expression of *SLC6A3*. In summary, this work increased understanding of genetic variation and secondary structure formation within the *SLC6A3* promoter, but further research is needed to determine any effects on gene expression or substance use disorder development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. David Vandenbergh for giving me the opportunity to conduct research in his laboratory. I would also like to thank him for assisting me with countless research projects and critiquing my thesis. I will always be grateful for his support over the last three years. I would also like to thank Dr. Helen Kamens for providing me with guidance and critiques on my thesis. Additionally, I would like to thank fellow research assistants Carly Brogan and Wesley Stone, who helped trim large chromosomal sequence files for analysis, and Andrew Burich for writing the code that streamlined the process of preparing chromosomal sequences for trimming. This thesis could not have been completed without their help. This research was funded by internal funds from Penn State University.

Introduction

Substance Use Disorder

Substance use disorder, defined as a chronic and relapsing disorder in which the affected individuals compulsively seek out the substance and continually use it despite destructive consequences (American Psychiatric Association, 2017), is a serious health issue. In a yearly survey, the National Survey on Drug Use and Health, conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA), 40.3 million Americans had a substance use disorder in 2020 (Substance Abuse and Mental Health Services Administration, 2021). Of this population, 28.3 million had alcohol use disorder, 1.3 million had cocaine use disorder and 1.5 million had methamphetamine use disorder (Substance Abuse and Mental Health Services Administration, 2021). Additionally, 51.7 million people used tobacco (Substance Abuse and Mental Health Services Administration, 2021). As described below, these drugs are of particular relevance to the gene that is the target of this research study.

As defined by the Diagnostic and Statistical Manuel of Mental Disorders Version 5 (DSM-5), the key feature of substance use disorders is continued use despite causing significant problems, such as relationships issues, problems fulfilling obligations at home, school or work, or health problems (American Psychiatric Association, 2017). According to the National Vital Statistics System, which aggregates health statistics collected by states, about 92,000 people died from drug overdose in 2020 (Ahmad et al., 2021). Since 2012, deaths from psychostimulants, especially cocaine and methamphetamine, have increased every year (Ahmad et al., 2021). These

facts illustrate the severity of substance use disorder and thus the importance of finding not only treatments, but the biological underpinnings of the disease.

SLC6A3: the Dopamine Transporter Gene

Genetics accounts for 40-60% of an individual's risk for substance use disorder (Goldman et al., 2005). One gene of interest is the human dopamine transporter gene, or Solute Carrier, Family 3, Subfamily A, member A (*SLC6A3*), which is located on chromosome 5. This gene encodes a protein termed the dopamine transporter (DAT1), which functions to transport dopamine from the synapse back into the presynaptic cell (Mortensen & Amara, 2003). This transport functions to terminate the dopamine signal and thus regulates signaling (Miller et al., 2021). Dopamine is involved in reward, and thus proteins that bind to dopamine, such as transporters, receptors, or enzymes, are targets of many abused drugs (Mortensen & Amara, 2003). Psychostimulants such as cocaine and methamphetamine specifically affect DAT1 (Mortensen & Amara, 2003). Cocaine is an antagonist and binds to the transporter to stop it from functioning (Mortensen & Amara, 2003). This activity extends the dopamine signal and leads to a sense of euphoria, which is a powerful factor in addiction (Mortensen & Amara, 2003). Methamphetamine and other amphetamines weakly affect DAT1 activity in comparison to cocaine (Mortensen & Amara, 2003). Methamphetamine and amphetamine have similar structures to dopamine, so they are transported into the cell instead of dopamine, which also extends dopamine signaling, but through a more complicated mechanism than for cocaine (Mortensen & Amara, 2003). Methamphetamine and amphetamine's predominant effects occur

by interacting with a second transporter, the vesicular monoamine transporter, which causes dopamine release from vesicles, which promotes reverse transport through DAT1 (Mortensen & Amara, 2003). Other substances indirectly affect DAT1 activity. Alcohol causes increased dopamine release (Zhu and Reith, 2008), which leads to upregulation of transporters in the synapse to compensate (Mortensen & Amara, 2003). Nicotine does not directly act on DAT1, but it leads to increased clearance of dopamine in the synapse (Zhu and Reith, 2008), which may occur through nicotinic acetylcholine receptors (nAChRs) because they are on dopamine neurons (Xiao et al., 2020) (Zhu and Reith, 2008). In summary, drugs of abuse can directly and indirectly modulate DAT1 activity; thus, *SLC6A3* is a gene of interest in substance use disorder. Additionally, DAT1 is expressed in almost every dopaminergic neuron, a neuron that releases dopamine as its primary neurotransmitter, further highlighting the protein's importance (Miller et al., 2021).

Known Genetic Variation Within *SLC6A3*

Because of *SLC6A3*'s relevance to substance use disorder, sequence variability within the gene may be a biological underpinning of the disease. There are multiple types of tandem repeats in the human genome, which are DNA sequences that are repeated head-to-tail. Those with a repeating unit longer than six base pairs, but shorter than 100 are termed Variable Number Tandem Repeats (VNTRs) (Bakhtiari et al., 2021). The number of repeats can vary from person to person (Brookes, 2013) (Reith et al., 2021). Three polymorphic VNTRs have been identified in and around *SLC6A3* that may affect gene expression (Reith et al., 2021). The first is in the 3'-

untranslated region (3'-UTR) (Vandenbergh et al., 1992). It has two common alleles termed 9r and 10r (Vandenbergh et al., 1992). Findings are conflicting regarding the association between the alleles and promoter activity. Particularly, *in vitro* studies show conflicting results. One study showed that having the 10r allele has been associated with greater promoter activity, and experiments with a Luciferase-based reporter assay demonstrated a functional effect on expression of the gene, suggesting that this allele leads to more transporters (Fuke et al., 2001). Another study utilized immunoblotting and radioligand binding techniques in cell lines and showed that the 10r allele led to greater DAT1 density than the 9r allele (VanNess et al., 2005). However, other studies used a Luciferase assay and found that there were no differences in expression between the two alleles (Mill et al., 2005) (Hill et al., 2010). Another study supported this finding, showing that the 3'-UTR VNTR was not correlated with mRNA levels in postmortem human midbrains (Zhou et al., 2014). Conversely, *in vivo* studies utilizing PET and SPECT scans from multiple studies showed that the 9r allele was associated with increased DAT1 density in humans (Faraone et al., 2013) (Shumay et al., 2011).

Another VNTR is in Intron 8 (Vandenbergh et al., 2000). This VNTR also has two common alleles coined 5r and 6r (Vandenbergh et al., 2000). In an *in vitro* study, a Luciferase assay was used to show that the 5r allele in Intron 8 VNTR had a functional effect on expression of *SLC6A3*. (Hill et al., 2010). Another *in vitro* study demonstrated a functional effect of the 5r allele using a *Renilla* vector in mouse derived Substania Nigra cells (Guindalini et al., 2006). *In vivo* studies demonstrate the same result, as PET scans showed that the 5r allele was associated with higher DAT1 levels in humans (Shumay et al., 2011). Another study utilizing a Luciferase assay demonstrated that the haplotype 9r allele in the 3'-UTR and the 5r allele in Intron 8 had the largest functional effect on *SLC6A3* expression of all combinations of those alleles (Zhao et al.,

2013). The third VNTR is upstream of the gene in the 5' flank (Zhou et al., 2014). mRNA levels in postmortem human midbrains demonstrated that 15-18 copies of this repeat was correlated with greater expression than 14 copies in healthy controls (Zhou et al., 2014). It is hypothesized that this VNTR is an enhancer (Zhou et al., 2014). In summary, some studies show an association of the VNTRs with differential gene expression, and in some cases, the VNTRs have been demonstrated to have a functional effect, although findings are conflicting regarding the 3'-UTR.

Certain VNTR alleles are associated with substance use disorder for several drugs, including alcohol, nicotine, cocaine, and methamphetamine, although findings on some VNTRs are conflicting (Reith et al., 2021). The 9r allele in the 3' UTR was associated with alcohol dependence (Ma et al., 2016) and unsuccessful smoking cessation (Ma et al., 2016); however, other studies found conflicting results. One study found that the 9r allele was associated with a lower risk of smoking (Tiili et al., 2020). Other studies found that the 3'-UTR's VNTR was not associated with polysubstance abuse (Persico et al., 1993) or methamphetamine dependence (Hong et al., 2003). To contrast, the 10r allele in the 3'-UTR was associated with ADHD (Grünblatt et al., 2019). Additionally, the 6r allele in intron 8 was associated with cocaine abuse (Guindalini et al., 2006), and the 5r allele was associated with a lower risk of smoking (Tiili et al., 2020). It is suggested that an increased *SLC6A3* expression, which leads to an increased number of transporters, which might be due to the genetic variation described above, leads to chronic dopamine depletion, and thus a depletion in reward (Reith et al., 2021). Those who experience fewer rewarding feelings in general are more likely to become dependent on a substance to make them feel good (Reith et al., 2021). Thus, these individuals are more at risk

for becoming dependent on the rewarding effects of a substance (Reith et al., 2021). In fact, chronic dopamine depletion is a risk factor for all substance use disorders (Reith et al., 2021).

Evidence suggests that all these VNTRs affect gene expression by interacting with the promoter in an unknown manner, highlighting the significance of the promoter of *SLC6A3* in substance use disorder. Promoters contain numerous binding sites for transcription factors, and mutations at the binding sites can have implications for gene expression (Tirosh et al., 2009). In addition, tandem repeats are common in promoters and can influence transcription (Sawaya et al., 2013).

An additional type of tandem repeat is a Short Tandem Repeat (STR), which has a repeating unit of six base pairs or less, although some authors use seven base pairs as the cutoff (Bakhtiari et al., 2021). There is additional evidence that these types of tandem repeats are also influencing gene expression (Sawaya et al., 2013). STRs within *SLC6A3* have not been studied, but one is the focus of the research described here.

G-Quadruplexes

One secondary structure in DNA, G-Quadruplexes (G4s), is commonly found in promoters, so they are potentially relevant to this study (Kolesnikova and Curtis, 2019). G4s are secondary structures that form in nucleic acids by stacking guanine tetrads (Kolesnikova and Curtis, 2019). These structures are extremely stable (Kolesnikova and Curtis, 2019). The minimum requirement for a sequence to form a G4 is 4 runs of at least two guanine bases separated by other bases, ranging in number from 1-7 nucleotides (Kolesnikova and Curtis,

2019). There are two types of G4s, parallel and antiparallel (Kypr et al., 2009). Parallel G4s have all four guanine chains in the same orientation, 5’ to 3’ (Kejnovská et al., 2019). Antiparallel G4s have two guanine chains oriented from 5’ to 3’, and two guanine chains oriented from 3’ to 5’ (Kejnovská et al., 2019). G4s are most likely to form in guanine rich sequences.

G4s are common in promoters, as 40% of human genes are predicted to have at least one in their promoter (Kolesnikova and Curtis, 2019). Some transcription factors have been demonstrated to bind to these sequences, offering a potential role in gene expression (Kolesnikova and Curtis, 2019). For example, a transcription factor *MAZ* binds to a G4 in the promoter of *cMYB* and decreases gene expression (Kolesnikova and Curtis, 2019). Thus, G4s may be functionally significant.

G4s cause problems when attempting to detect genetic variation. One study showed that G4s and methylation block amplification by Polymerase Chain Reaction (PCR) (Stevens et al., 2017). This was demonstrated in the promoter of seven genes, including *MEST, AIM1, PLAGL1, GBR10, BLCAP, KCNQ1,* and *DNMT1.* In all these genes, the researchers had difficulty amplifying the promoter despite trying to optimize the reaction with different reagents and conditions (Stevens et al., 2017). Additionally, G4s were identified in all these genes, so the authors suggested that G4s were inhibiting the reaction (Stevens et al., 2017). G4s are highly stable, making DNA harder to denature, which might explain their ability to block amplification by PCR (Kolesnikova and Curtis, 2019).

Hypotheses

This study had two hypotheses. The first was to identify whether an STR within the *SLC6A3* promotor is variable. This STR has previously been identified in the Repeat Masker Track of the Genome Browser (accessed from http://genome.ucsc.edu), but variability within it has not been studied, nor have any association studies with disease phenotypes been conducted. The second was to identify whether G4s form at this STR because the region is GC rich. Results from tests of these hypotheses will provide evidence of the potential role of the STR in modifying expression of *SLC6A3*.

Methods

Identification of Repeat

A hexamer repeat in the 150 nucleotides upstream of *SLC6A3* was identified using the hg38 version of the human genome in the Genome Browser hosted by the University of California, Santa Cruz (UCSC) Genome Browser. The UCSC Genome Browser (http://genome.ucsc.edu) is a tool used to display the genome and functional elements such as promoters, repetitive regions, single nucleotide polymorphisms (SNPs), tandem repeats, and transcription factor binding sites (Kent et al., 2002).

Testing for Genetic Variability: Wet Lab Technique

The repeat was tested for variability using two methods. The first was a wet lab technique utilizing PCR and gel electrophoresis. Previous research assistants recruited human research study participants and collected and purified DNA samples. Participants were recruited by word of mouth under IRB STUDY00008832 (PI Vandenbergh). A saliva sample was taken from each participant using a buccal swab without personal identifiers. DNA was extracted from the samples using an organic solvent, a mixture of phenol and chloroform, and then ethanol precipitated before dissolving in water.

PCR is a method in which selective sequences of DNA are amplified so the DNA is in a large enough quantity that it can be detected in gel electrophoresis (Green & Sambrook, 2019). Primers that flank the STR were selected using Primer-BLAST, a software tool that selects

optimal primer sequences for a PCR reaction (Ye et al., 2012). The primers selected were 1445290R and 1445617F, labeled based on their chromosomal coordinates of their 5’ end. Their sequences were as follows.

1445290R: (5')-CTTCGGGGTCTGCGGAAG-(3’)

1445617F: (3’)-CAGGACAGTGGAGCCGTG-(5’)

The PCR reaction produced inconsistent and unclear results at first, so different reaction conditions were tried to optimize the reaction, such as different primers, enzymes, additives, and annealing temperatures. The most successful results were obtained with Q5 enzyme and GC enhancer from New England Biolabs following the manufacturer’s instructions. Eight reactions were done, but extra reaction mix was made, specifically enough for nine reactions, to allow for pipetting error.

Reagents	1 reaction	9 reactions
Water	0.25 µL	2.25 µL
5X GC Enhancer	5.0 µL	45.0 µL
5X Buffer	2.5 µL	22.5 µL
2 mM dNTP	2.5 µL	22.5 µL
2 mM 1445617F	6.25 µL	56.25 µL
2 mM 1445290R	6.25 µL	56.25 µL
Q5 Enzyme	0.25 µL	2.25 µL

Table 1: PCR Reaction Mix

Reagents were added to a microtube in the order they appear in the table. After adding the enzyme, the sample was mixed by gently flicking the tube. Then, the reaction mix was pipetted into eight tubes. Then, 2.0 µL of DNA samples were added to seven of the tubes, while water was added in place of DNA in the eighth tube because seven DNA samples were used plus

a negative control for the eighth tube. The reaction was then carried out using a GeneAmp PCR System 9700 (ThermoFisher). The DNA was denatured at 95.0°C for three minutes followed by thirty-five cycles of 92°C for thirty seconds, 63°C for thirty seconds, and 72°C for thirty seconds. Then, the DNA was held at 72°C for ten minutes before refrigerating at 4°C.

After the PCR reaction, the length of the repeat was identified utilizing polyacrylamide gel electrophoresis. Gel electrophoresis separates DNA fragments by size using an electrical field (Lee et al., 2012). Fragment size is determined by comparison to a DNA ladder, which consists of fragments of known lengths (Lee et al., 2012). A negative result, indicating no variability, would exhibit bands that are all the same length. A positive result, indicating variability in the number of repeats, would exhibit bands that are different lengths and/or individuals with multiple bands, which indicates heterozygosity. 10.0 μL of each sample was mixed with 1.0 μL of tracking dye with glycerol. These mixtures were loaded into the wells in the gel, along with a DNA ladder of DNA fragments that differed by 100 base pairs. The gel was run for one hour at 200 V current and 100 mA resistance. After this, the gel was soaked in ethidium bromide for five minutes. It was then photographed under UV illumination.

Testing for Variability: Bioinformatic Method

The second method used to test for variability was a bioinformatic technique, by aligning the sequences of 62 unrelated individuals. This method was intended to serve as a confirmation for the results found in the wet lab experiment and identify other variation, such as SNPs, that could not be identified by PCR and gel electrophoresis. Sixty-four long read sequences of

chromosome five were obtained from the Human Genome Structural Variant Consortium (Ebert et al., 2021). A python code written by Andrew Burich, another research assistant associated with Dr. Vandenbergh's research group, trimmed the sequences to only include the repeat of interest. Two barcodes, or DNA sequences, were selected to be complementary to sequences flanking the repeat region and were used to trim the sequences to the left and right of the repeat. The barcodes' sequences were as follows:

5' barcode: (5')-CGCGAAGCCTCCCCTCCCGCTCCGCAGCGCT-(3')

3' barcode: (5')-TCCGAGACGTCGCGGATCTAGACGGGCGCCT-(3')

All the trimmed sequences were concatenated into one file using another python code written by Andrew Burich. Then, a multiple alignment of all the trimmed sequences was done using SnapGene® Software (from Insightful Science, available at snapgene.com). Research assistants Carly Brogan and Wesley Stone assisted in running the codes and doing the multiple alignment.

Detection of G-Quadruplexes

The G4 formation in this repeat was first hypothesized with an online analysis tool called QGRS Mapper (Kikin et al., 2006). QGRS Mapper is a software that predicts G4 formation based on the sequence (Kikin er al., 2006). The sequence of the repeat was obtained from the Repeat Masker Track of the UCSC Genome Browser and loaded into QGRS Mapper for analysis.

The presence of G4s was confirmed by circular dichroism (CD) spectroscopy. CD spectroscopy uses circular polarized light to detect the difference in absorbed light between left- and right-handed molecules (Bishop & Chaires, 2003). B-DNA, the standard double helix, is right-handed whereas Z-DNA (G4s) is left-handed (Kypr et al., 2009). Thus, more absorbance by left-handed molecules indicates the presence of G4s (Kypr et al., 2009).

A protocol for the spectroscopy was designed based on the protocol provided by (Kejnovská et al., 2019). The full protocol is available in Appendix A. Both single stranded and double stranded DNA were tested, even though it is suggested that CD cannot identify G4s in double-stranded DNA, only in single-stranded DNA (Kreig et al., 2015). It has been suggested that double stranded DNA does not produce a spectrum consistent with those expected for G4s (Kreig et al., 2015). However, this experiment was of value to try because it more closely represents formation in vivo.

An oligonucleotide encompassing a sequence identified by QGRS Mapper and its reverse complement were ordered from Integrated DNA Technologies (idtdna.com). This specific sequence was selected because it was the longest that could be ordered as a primer. The reverse complement was obtained from the Sequence Manipulation Site, a website that contains programs useful to formatting and analyzing DNA sequences (Stothard, 2000). The primers selected were labeled 1445541R and 1445569F. 1445541R is the strand hypothesized to form G4s. Their sequences were as follows:

1445541R: (5’)-GGGAGGCGGGGGCGGGGGCGCGGCCCGGG-(3’)

1445569F: (3’)-CCCGGGCCGCGCCCCCGCCCCCGCCTCCC -(5’)

The experiment consisted of the following steps. First, 100 mM oligonucleotide was mixed with 0.2 M sodium phosphate buffer (pH: 7.0) to a final concentration of 10 mM. The

DNA was then denatured at 90°C for 5 minutes using a thermocycler. Then, the sample was allowed to cool to room temperature on the lab bench overnight. After this, a CD spectrum was taken on a Jasco 1500 spectrophotometer at the Huck X-Ray Crystallography Core Facility. This spectrum was intended to test DNA in baseline conditions that does not allow G4 formation as a comparison (Kypr et al., 2009) (Kejnovská et al., 2019). Potassium induces G4 formation, so after the first recording of a spectrum, 0.2 M potassium phosphate buffer (pH: 7.0) was added to induce G4 formation and another spectrum was taken. Last, 3 M KCl was added to further induce G4 formation, followed by a final spectrum.

Three experiments were conducted. The first tested 1445541R alone. The amounts of sodium phosphate buffer, potassium phosphate buffer and potassium chloride were based on the proportions in the protocol (Kejnovská et al., 2019). The second experiment tested 1445569F, an identical reaction mix to that above, but with 1445569F instead of 1445541R. The third experiment tested both 1445541R and 1445569F to test for G4 formation in double stranded DNA. The amount of sodium buffer was adjusted to allow for 20 μL of the extra primer.

Reagent	Amount
100 mM 1445541R	20 μL
0.2 M Na_2HPO_4	84.28 μL
0.2 M NaH_2PO_4	84.28 μL
0.2 M K_2HPO_4	2.4 μL
0.2 M KH_2PO_4	2.4 μL
3 M KCl	6.66 μL
	200 μL

Table 2: Reaction Mix for CD spectrum of Single Stranded DNA (1445541R)

Data from the sodium phosphate buffer test was used as a baseline measurement and thus was subtracted from the other spectra. G4 formation was tested at 25°C (room temperature). The samples were scanned between 190 and 350 nm because DNA has a characteristic pattern of absorbance of light between these wavelengths (Kejnovská et al., 2019).

The spectrophotometer produces measurements in millidegrees (mdeg). However, CD spectra in the literature utilize ellipticity because it controls for DNA concentration (Kejnovská et al., 2019). Thus, the units were converted to ellipticity using the following equation:

$$\varepsilon = \frac{(mdeg\ x\ M)}{32980\ x\ C\ x\ L}$$

Where "M" is the molecular weight of the DNA (in g/mol), "C" is the molar concentration of DNA, and "L" is the optical path of light (in cm). The complete calculations can be found in Appendix B.

G4s have a few characteristic features on a CD spectrum. Parallel G4s have a positive peak at 260 nm and a negative peak at 240 nm, whereas antiparallel G4s have a positive peak at 290 nm and negative peak at 260 nm (Kypr et al., 2009). Both forms have a positive peak at 210 nm (Kypr et al., 2009). For a comparison, see Figure 4 in (Kypr et al., 2009). These features will be examined in this experiment to identify whether G4s are present.

Results

Identification of the Repeat

According to the UCSC genome browser, the sequence indicated by a black bar in the figure below is an STR (see Figure 1). The repeat is located on chromosome 5 and its chromosomal nucleotide coordinates are chr5: 1445477-1445570. The STR is a hexamer with a repeated sequence of GGCGGG, which was obtained by clicking on the black box (sequence is not displayed in Figure 1). The estimated length of the repeat is 327 base pairs. This tandem repeat is slightly upstream of *SLC6A3*. The blue box at the upper left of the figure is exon 1 and the blue line is intron 1 (transcription goes to the left) of *SLC6A3*. This repeat occurs within the promoter, and is indicated by lighter blue and red boxes with the label "promoter" on the left side of the figure. The red promoter in the track titled, "ENCODE Candidate Cis-Regulatory Elements (cCREs) combined from all cell types", was identified as sequence with promoter like sequence signatures (Moore et al., 2020), while the blue promoter in the track titled, "Promoters from EPDNew human version 006", was experimentally validated (Dreos et al., 2013).

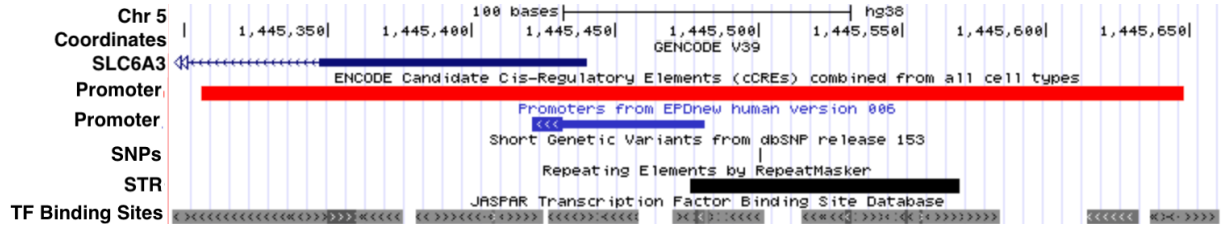


Figure 1: The STR and Nearby Functional Elements, Accessed From http://genome.ucsc.edu.

There is a SNP within the repeat, which is displayed by the black tick mark in the track titled "SNPs". The SNP's accession number is rs2975226 and its chromosomal coordinate is chr5: 1445501. The reference allele is adenine (A), and the alternative allele is thymine (T). The SNP was identified in the SNP database (dbSNP, Build 153) (Sherry et al., 1999), and among

different studies presented in dbSNP, the frequency of the A allele ranges from 37.0%-62.9%. These studies encompassed multiple populations including European, Hispanic/Latino, African, and Asian ancestries. Populations of European ancestry were most commonly studied.

In Figure 2 below, the gray boxes are predicted transcription binding sites from the JASPAR core collection, a database for experimentally identified transcription factor (TF) binding sites (Castro-Mondragon et al., 2022). A minimal view is shown in the previous figure, while the complete collection of transcription factors is shown in the figure below.

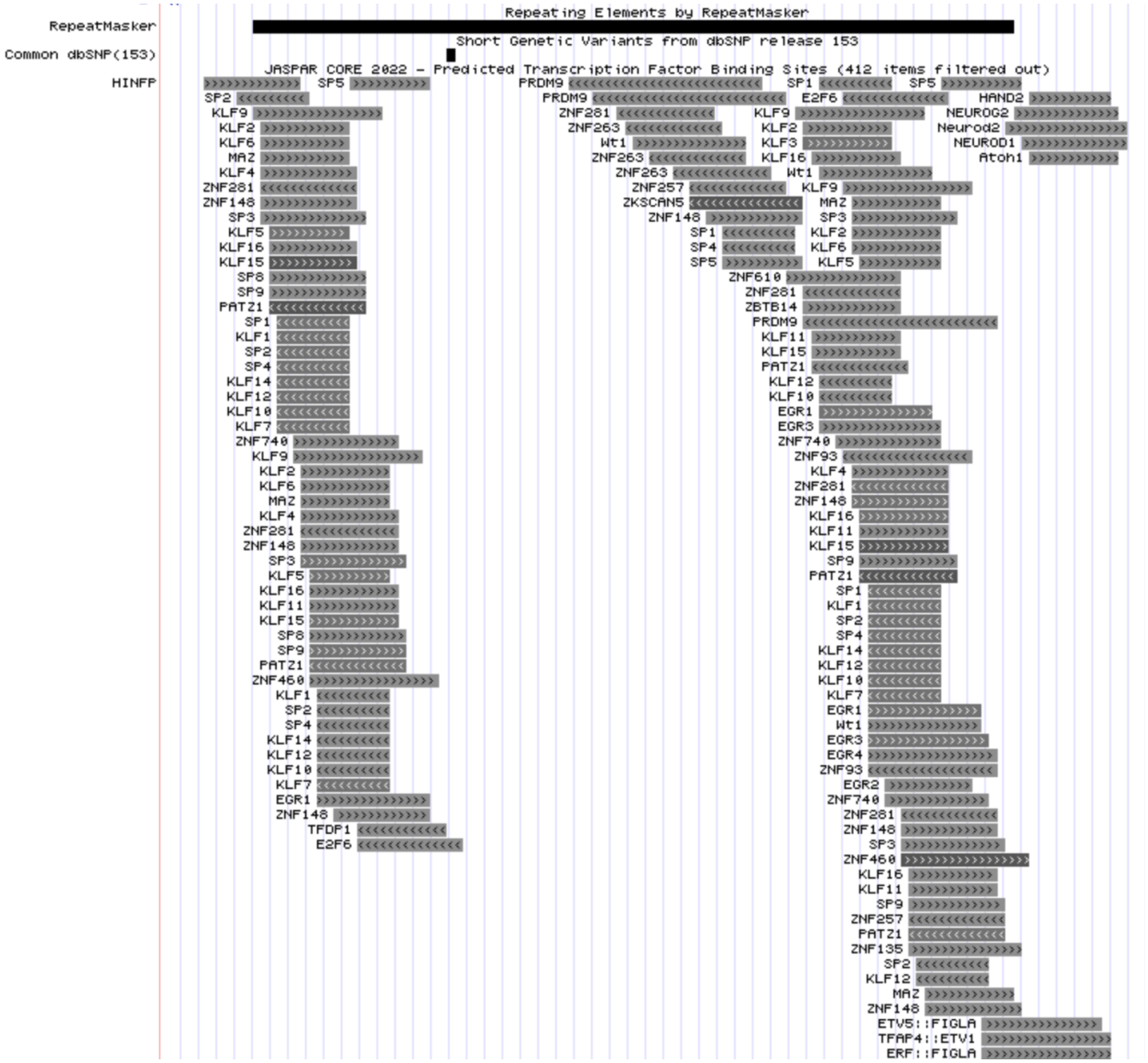


Figure 2: Transcription Factor Binding Sites Within the STR from the Genome Browser, Accessed From http://genome.ucsc.edu.

There are numerous transcription factor bindings sites within the STR that is shown by the black bar above the binding sites. Additionally, rs2975226 is within the E2F6 binding site.

The figure below shows expression of transcription factor E2F6 in different tissues throughout the body. These data were obtained from the Genotype-Tissue Expression Project, a database with information relating genetic variants and gene expression (GTEx Consortium, 2013). A box above a specific tissue indicates that E2F6 is being expressed in that tissue. The red

arrow is pointing to the substantia nigra, where dopaminergic neurons are present and express the dopamine transporter (Sonne et al., 2021). Although this finding is consistent with the expression of E2F6 in dopaminergic neurons, it remains to be tested experimentally.

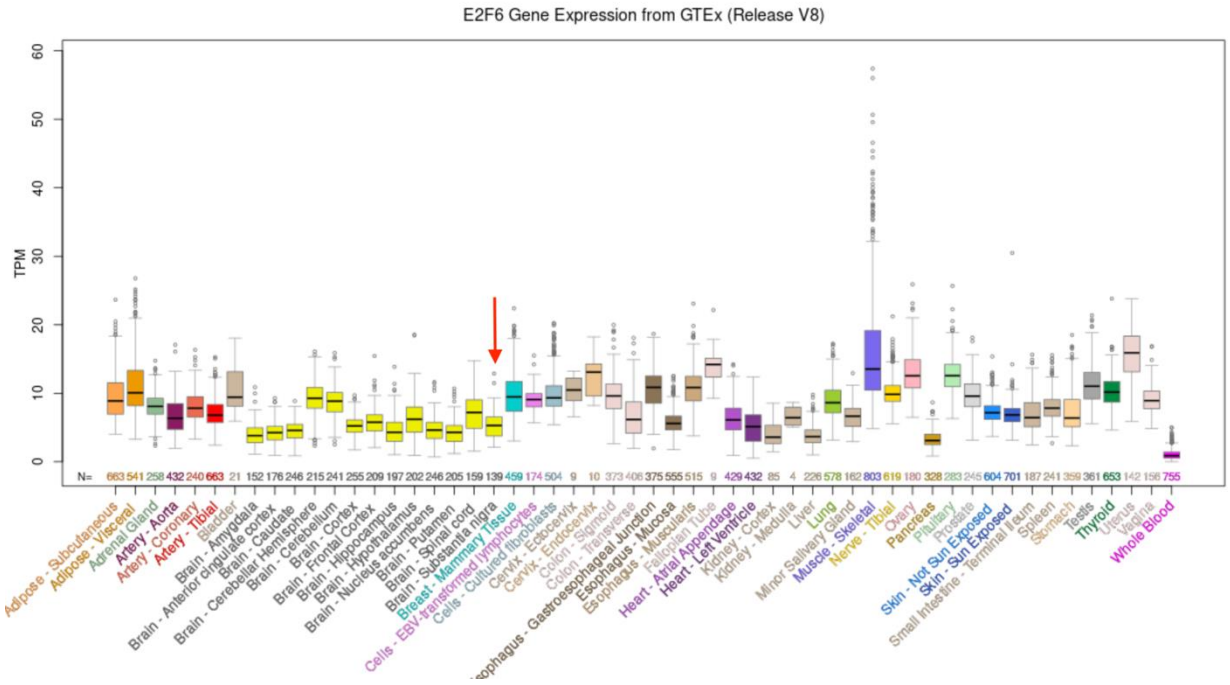


Figure 3: Expression of E2F6 in Different Tissues Throughout the Body

Genetic Variation Between Individuals: Wet Lab Technique

The gel images in Figure 4 (below) show the PCR reaction intended to amplify the STR. Each number represents a different human DNA sample. The arrow is pointing to the expected product, 327 base pairs, which was predicted by the Genome Browser. Figure 4A shows the best amplification produced. There is some background amplification present, highlighting the difficulty amplifying this region. There is no variation in the bands. They appear to be the same length, between 300 and 400 base pairs, which matches the expected length. Because the bands

are the same length, there is no variability in the number of repeats between individuals, refuting the hypothesis. One limitation of this gel is that it did not include a negative control. Figure 3B compensates for this omission. This image shows a gel of the same reaction, but with a negative control (lane 0). Because of the difficulty amplifying this region, the reaction could not be replicated to include a negative control and clean amplification. Multiple experiments varying conditions of the PCR reaction were unable to resolve the background problem. Thus, this gel contains a lot of background amplification, but everyone had the 327 base pair band. There is no product in the control lane, indicating that the only DNA present in the reaction mixes was the sample DNA.

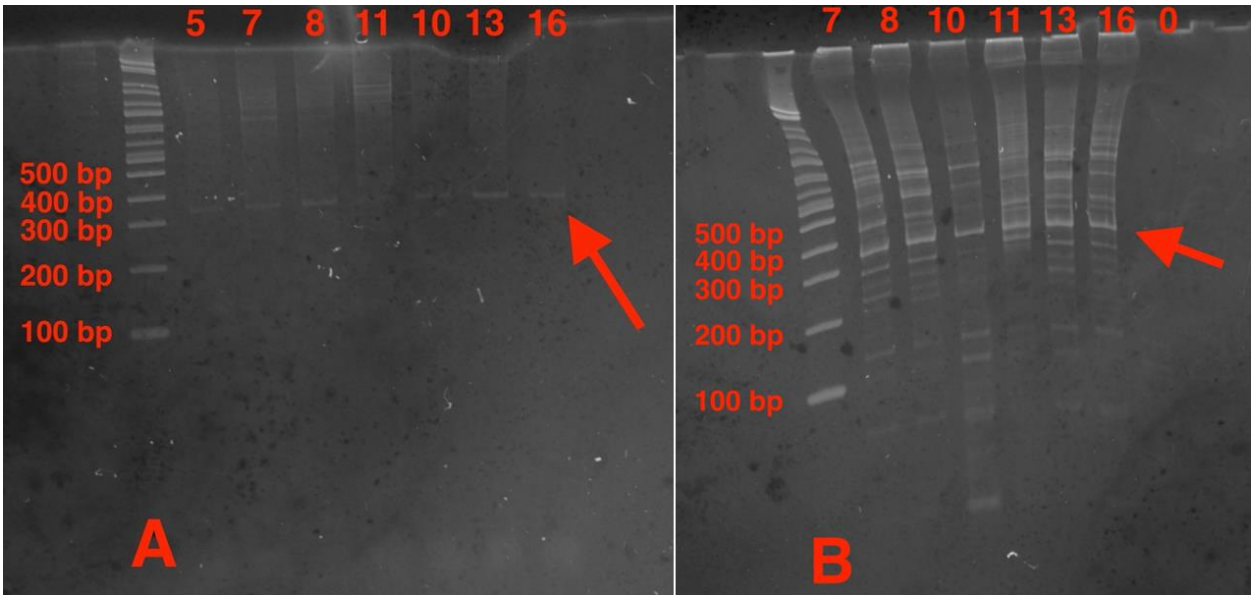


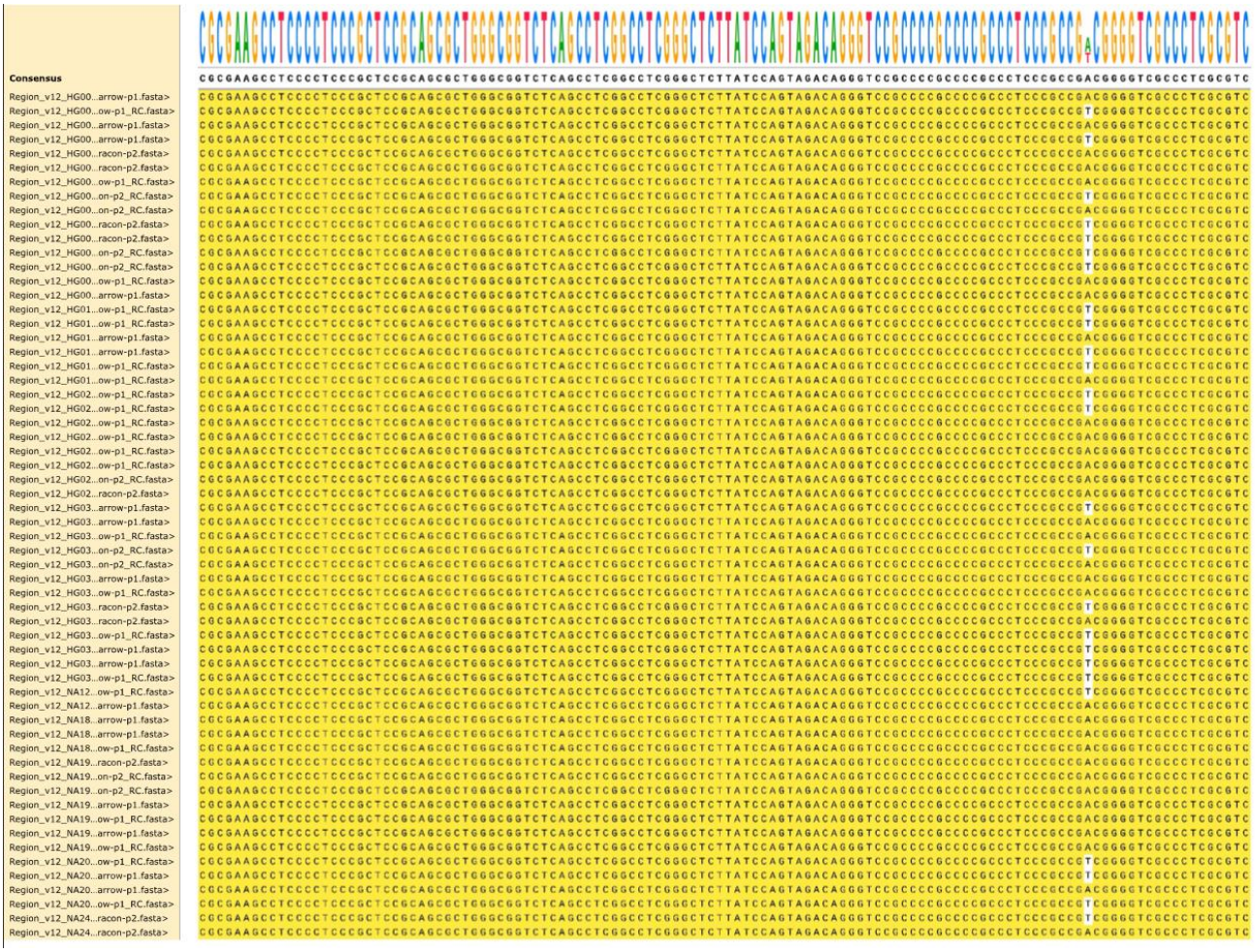
Figure 4: Gel Images of Amplified STR

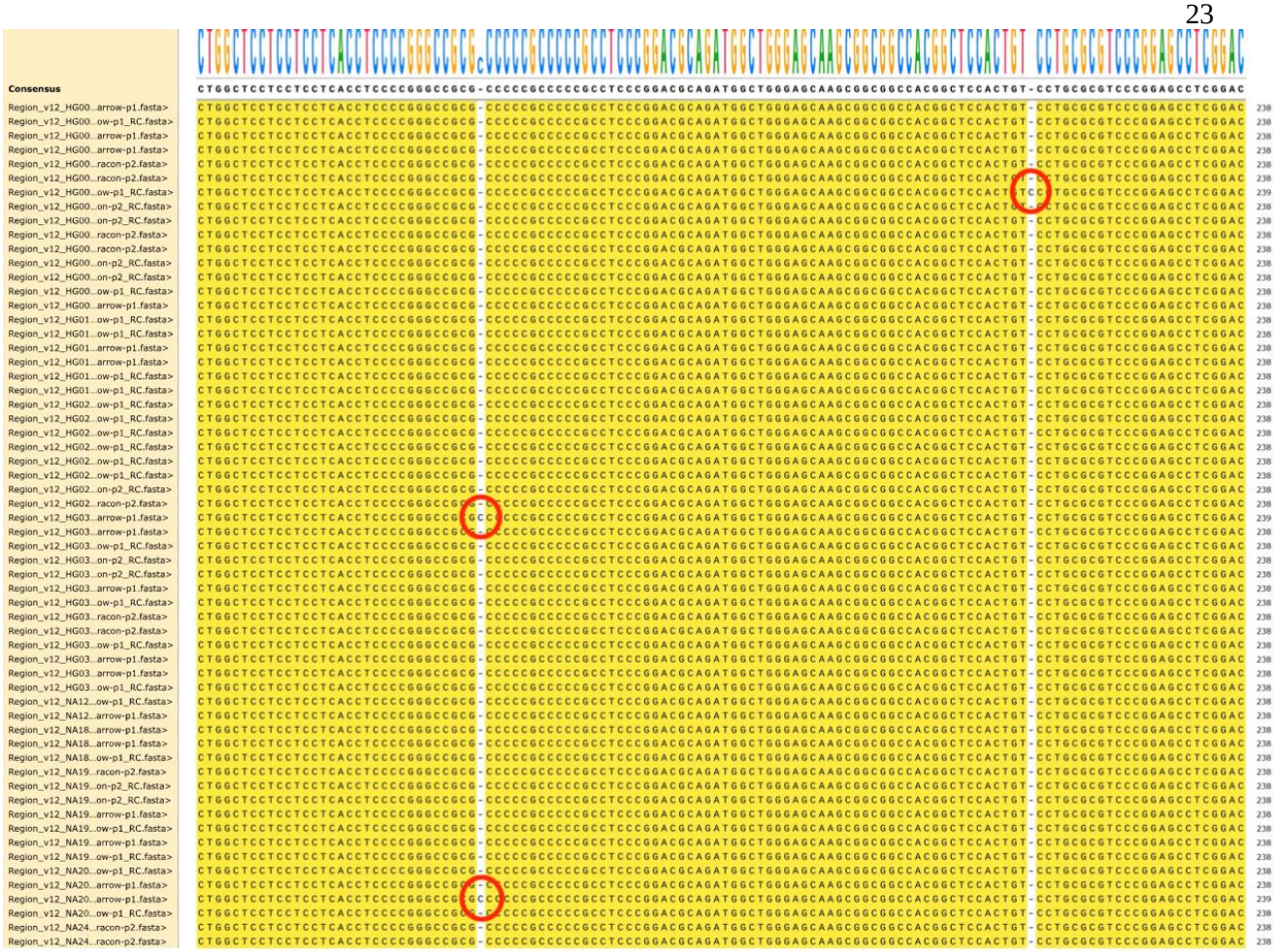
Genetic Variation Between Individuals: Bioinformatic Technique

In Figure 5 (below), each line of sequence represents a different individual's trimmed sequence flanked by barcodes described in the methods. The sequence was split into three sections due to length. The numbers on the right-hand side are the base pair numbers within the trimmed sequence. The letters at the top of the sequence are the bases that are present at that coordinate and the size of the letters refers to the proportion of sequences that have that base. If only one letter is present and large, then all the sequences have that base. However, if a letter is smaller, then only some of the sequences have that base. Two letters on top of each other indicates there is a SNP, specifically a single base substitution. There is no variation in the number of repeats in each row, indicating that the number of repeats does not vary between chromosomes from different individuals. However, there is still variation between the sequences. The A/T SNP identified by the UCSC Genome Browser (accession number: rs2975226, Figure 1) is present as expected. 59.7% of alleles have the A allele and 40.3% of alleles have the T allele, which is within the expected range reported in the ALFA project, which provides aggregate allele frequency from dbGaP from multiple global populations (https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/). Known associations between this SNP and disease phenotypes are reviewed in the discussion section. Additionally, there are two novel single base insertions in three different sequences. One insertion is present in two sequences. Each insertion is a cytidine (C). They are identified with a red circle in the figure.

Chromosomal coordinates for the insertions were estimated utilizing the sequencing data and known coordinates from the genome browser. The first insertion's estimated coordinate is chr5: 1445451. It is not within the repeat but is still within the promoter. It is in four predicted transcription factor binding sites, including ZFX, TFAP2A, TFAP2C, and TFAP2B. The second

insertion's estimated coordinate is chr5: 1445388. Like the other insertion, it is within the promoter but not the repeat. It is within two predicted transcription factor binding sites, ZNF417 and ZNF574.




Consensus
CTGGCTCCTCCTCCTCACCTCCCCGGGCCGCG-CCCCCGCCCCCGCCTCCCGGACGCAGATGGCTGGGAGCAAGCGGCGGCCACGGCTCCACTGT-CCTGCGCGTCCCGGAGCCTCGGAC
Region_v12_HG00...arrow-p1.fasta>
Region_v12_HG00...ow-p1_RC.fasta>
Region_v12_HG00...arrow-p1.fasta>
Region_v12_HG00...arrow-p1.fasta>
Region_v12_HG00...racon-p2.fasta>
Region_v12_HG00...racon-p2.fasta>
Region_v12_HG00...ow-p1_RC.fasta>
Region_v12_HG00...on-p2_RC.fasta>
Region_v12_HG00...on-p2_RC.fasta>
Region_v12_HG00...racon-p2.fasta>
Region_v12_HG00...racon-p2.fasta>
Region_v12_HG00...on-p2_RC.fasta>
Region_v12_HG00...on-p2_RC.fasta>
Region_v12_HG00...ow-p1_RC.fasta>
Region_v12_HG00...arrow-p1.fasta>
Region_v12_HG01...ow-p1_RC.fasta>
Region_v12_HG01...ow-p1_RC.fasta>
Region_v12_HG01...arrow-p1.fasta>
Region_v12_HG01...arrow-p1.fasta>
Region_v12_HG01...ow-p1_RC.fasta>
Region_v12_HG01...ow-p1_RC.fasta>
Region_v12_HG02...ow-p1_RC.fasta>
Region_v12_HG02...ow-p1_RC.fasta>
Region_v12_HG02...ow-p1_RC.fasta>
Region_v12_HG02...ow-p1_RC.fasta>
Region_v12_HG02...ow-p1_RC.fasta>
Region_v12_HG02...ow-p1_RC.fasta>
Region_v12_HG02...on-p2_RC.fasta>
Region_v12_HG02...racon-p2.fasta>
Region_v12_HG03...arrow-p1.fasta>
Region_v12_HG03...arrow-p1.fasta>
Region_v12_HG03...ow-p1_RC.fasta>
Region_v12_HG03...on-p2_RC.fasta>
Region_v12_HG03...on-p2_RC.fasta>
Region_v12_HG03...arrow-p1.fasta>
Region_v12_HG03...ow-p1_RC.fasta>
Region_v12_HG03...racon-p2.fasta>
Region_v12_HG03...racon-p2.fasta>
Region_v12_HG03...ow-p1_RC.fasta>
Region_v12_HG03...arrow-p1.fasta>
Region_v12_HG03...arrow-p1.fasta>
Region_v12_HG03...ow-p1_RC.fasta>
Region_v12_NA12...ow-p1_RC.fasta>
Region_v12_NA12...arrow-p1.fasta>
Region_v12_NA18...arrow-p1.fasta>
Region_v12_NA18...arrow-p1.fasta>
Region_v12_NA18...ow-p1_RC.fasta>
Region_v12_NA19...racon-p2.fasta>
Region_v12_NA19...on-p2_RC.fasta>
Region_v12_NA19...on-p2_RC.fasta>
Region_v12_NA19...arrow-p1.fasta>
Region_v12_NA19...ow-p1_RC.fasta>
Region_v12_NA19...arrow-p1.fasta>
Region_v12_NA19...ow-p1_RC.fasta>
Region_v12_NA20...ow-p1_RC.fasta>
Region_v12_NA20...arrow-p1.fasta>
Region_v12_NA20...arrow-p1.fasta>
Region_v12_NA20...ow-p1_RC.fasta>
Region_v12_NA24...racon-p2.fasta>
Region_v12_NA24...racon-p2.fasta>
238
239

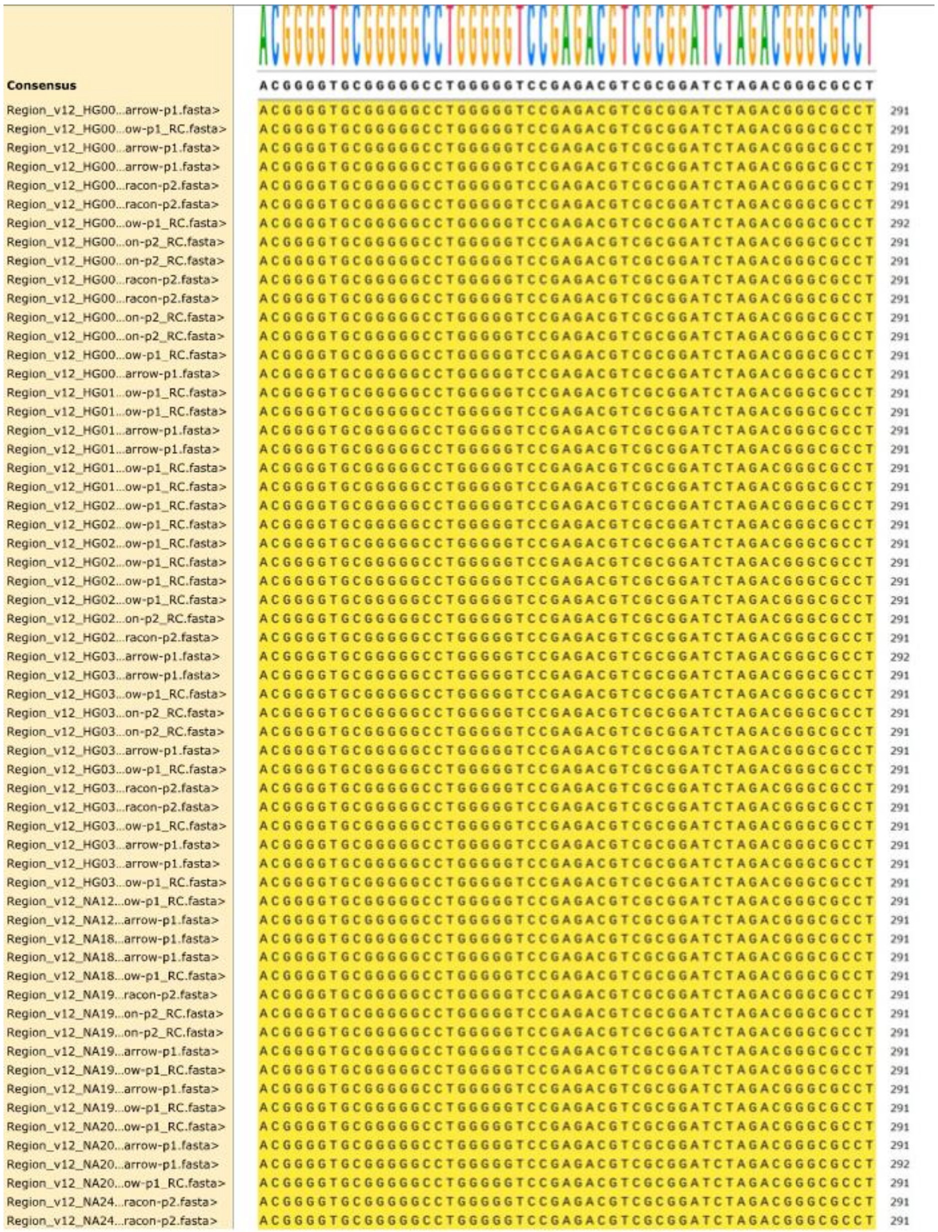


Figure 5: Multiple Alignment of 62 Sequences of the STR

G-Quadruplex Formation

The QGRS Mapper output identified three potential sequences in which G4s might be present (Figure 6). The G-Score refers to the likelihood that a sequence will form a G4 (Kikin et

al., 2006). A higher G-Score means a higher likelihood (Kikin et al., 2006). The first sequence was chosen for analysis by circular dichroism because it was the longest and had the second highest G-score.

Position	Length	QGRS	G-Score
1	29	GGGAGGCGGGGGCGGGGCGCGGCCCGGG	38
32	13	GGTGAGGAGGAGG	19
74	17	GGGAGGGCGGGGCGGGG	41

Figure 6: QGRS Mapper Output

The CD spectrum below in Figure 7 utilized the oligonucleotide 1445541R which was hypothesized to form G4s by QGRS Mapper. The x-axis represents wavelength, and the y-axis represents ellipticity. The blue line is DNA in sodium phosphate buffer alone, whereas the orange line is DNA in sodium phosphate buffer plus potassium phosphate buffer. The gray line represents DNA, sodium phosphate buffer, potassium phosphate buffer, and potassium chloride. This graph exhibits that this strand of DNA forms parallel G4s. All reaction mixes have a peak at 210 nm, a trough at 240 nm and another peak at 260 nm, which are all characteristic of parallel G4 spectra. The DNA and sodium phosphate buffer mix even has these characteristics, which is not expected, but shows that G4s might form in this region even without the presence of potassium. Additionally, the magnitude of the peaks increases as additional potassium is added, which is expected as potassium induces G4 formation (Kypr et al., 2009). This increase in signal strength with increasing concentrations of potassium further supports the hypothesis that G4s are forming.

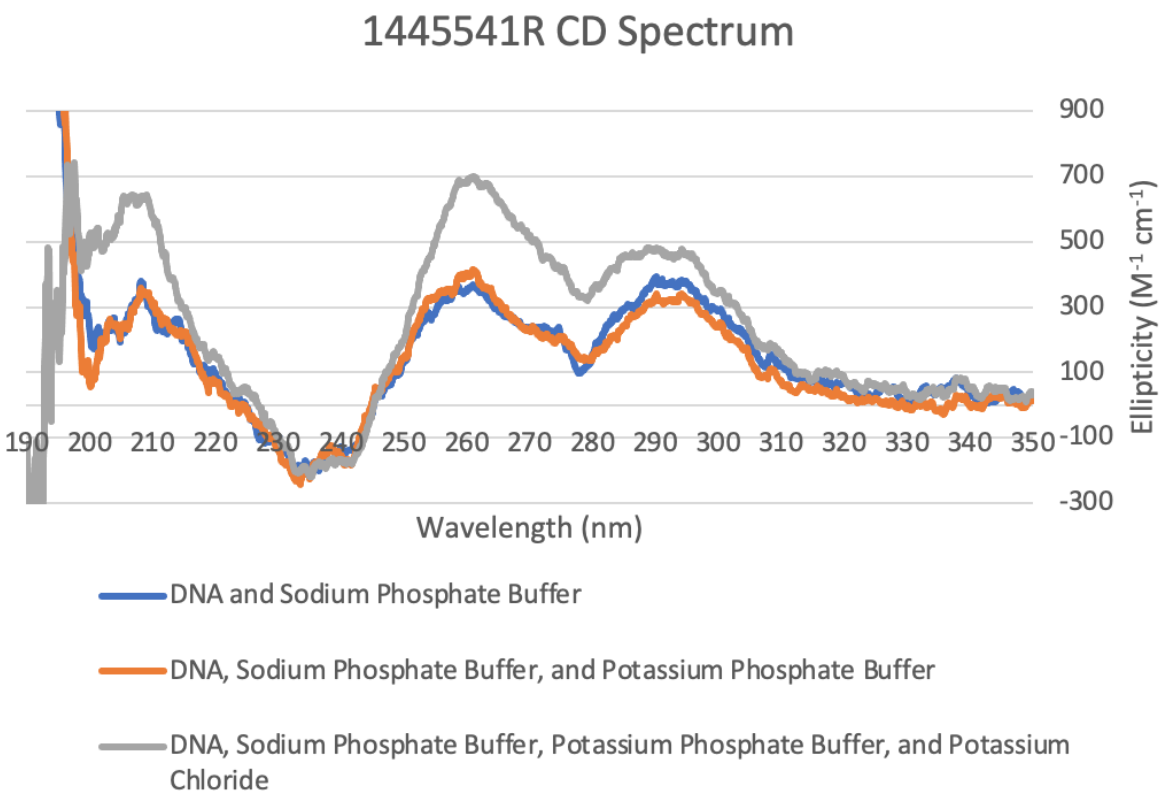


Figure 7: CD Spectrum of STR Utilizing Single Stranded DNA (1445541R)

The next CD spectrum (Figure 8) utilized oligonucleotide 1445569F, the reverse complement of the oligonucleotide hypothesized to form G4s. This graph is a control and shows single stranded DNA that does not form G4s. It does not have the characteristic features of G4s, including a peak at 210 nm, a trough at 240 nm, and a peak at 260 nm. All the reaction mixes in this experiment have amplitudes of similar magnitude, indicating that the addition of potassium did not have any effect. This exhibits a CD spectrum of DNA that does not form G4s.

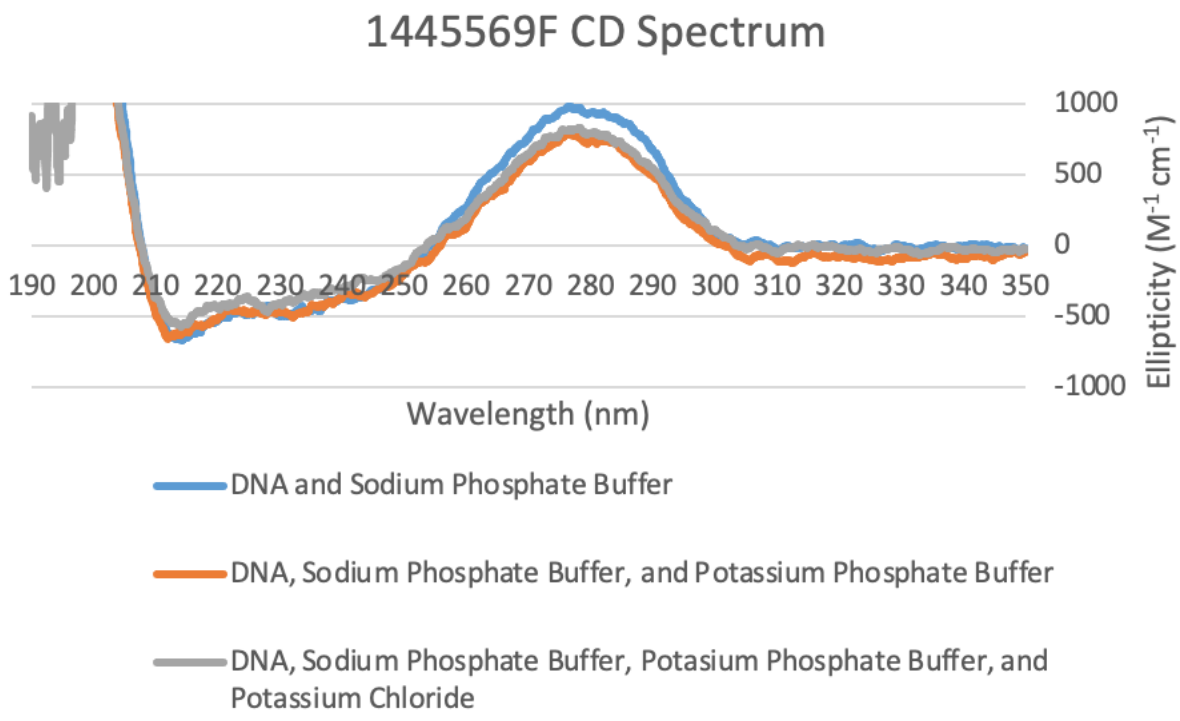


Figure 8: Control CD Spectrum Utilizing 1445569F

The final CD spectrum (Figure 9) utilized oligonucleotides 1445541R and 1445569F to test double stranded DNA. 1445541R forms G4s and 1445569F is its reverse complement. This graph exhibits that, when double stranded, these two oligonucleotides do not form G4s. Although the characteristic trough at 240 nm is present, the peak at 210 nm is absent and the peak at 260 nm is shifted to the right to about 265 nm. All three curves from the three reaction mixes have similar magnitudes, indicating that the addition of potassium did not have any effect. This analysis indicates that G4s did not form; however, CD spectroscopy does not accurately assess G4 formation in double stranded DNA, so this result is expected (Kreig et al., 2015).

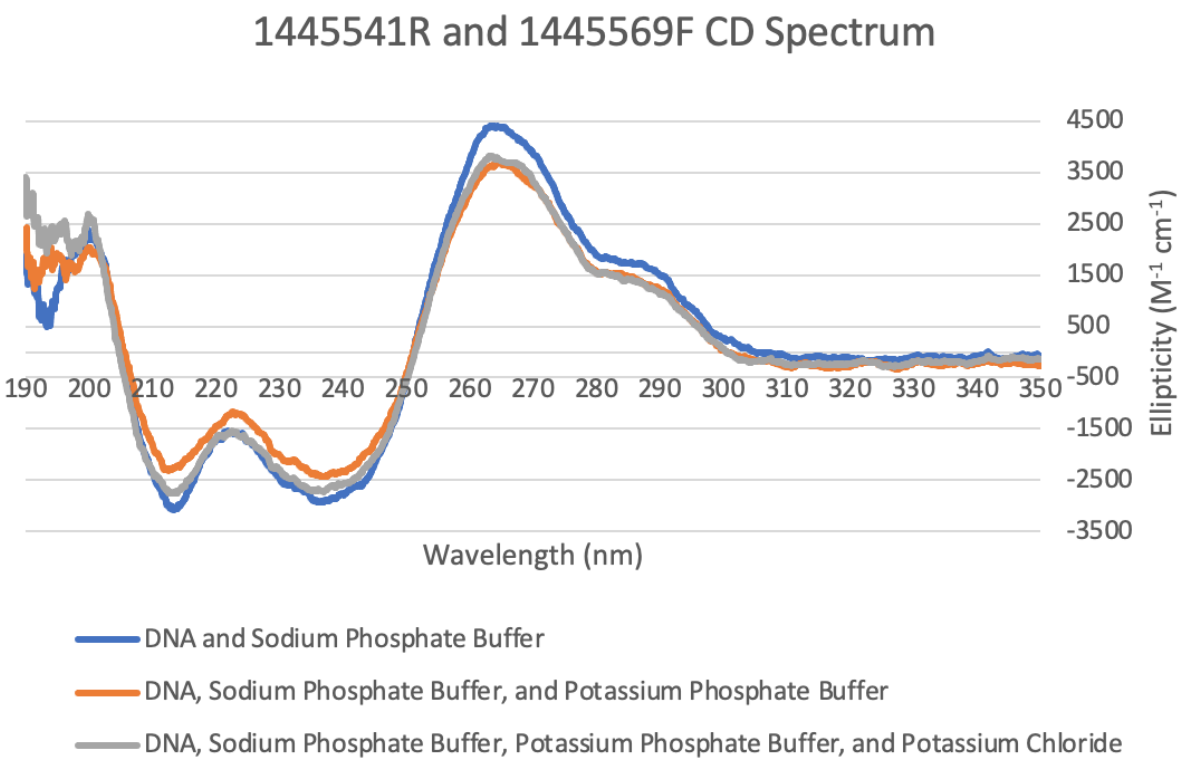


Figure 9: CD Spectrum of STR Utilizing Double Stranded DNA (1445541R and 1445569F)

Discussion

Genetic Variation

These results from this research refute the hypothesis that this repeat is a variable STR. Both the wet lab and the bioinformatics methods show that the number of repeats does not vary from individual to individual. However, there is still other variation present in this STR. As expected, rs2975226 was present in the sequencing data, and the frequency was consistent with results from population studies reported in dbSNP. Additionally, three novel single-base insertions, in three different sites, were identified. Although the data refuted the original hypothesis, novel variation was still found.

rs2975226 is associated with a few diseases. Two studies showed that rs2975226 is associated with schizophrenia and another study showed that it is associated with ADHD (Xu et al., 2020; Huang et al., 2010; Xu et al., 2009). Another study found that the SNP is weakly associated with bipolar disorder (Huang et al., 2015). However, no studies have been done on the association between rs2975226 and substance use disorder. Thus, future studies should be done to understand whether these are linked, given that other variants in the gene are associated with substance use disorder.

This variability detected in the *SLC6A3* promoter could have implications for gene expression. rs2975226 is within the predicted binding site for transcription factor, E2F6, which is a member of a family of transcription factors that regulate cell proliferation (Helin, 1998). Specifically, they are required for initiation of DNA replication (Helin, 1998). This transcription factor is potentially expressed in dopaminergic neurons, so it is possible that this SNP affects transcription factor binding, and future studies, perhaps with gel shift assays, should be done to

test variability in binding of E2F6 to the two alleles at rs2975226. If this SNP does affect transcription factor binding, this could have implications for cell proliferation, but it is unclear if this will have any effect on substance use disorder risk.

The first insertion is within binding sites for transcription factors ZFX, TFAP2A, TFAP2C, and TFAP2B. Three of these factors, TFAP2A, TFAP2C, and TFAP2B, are not expressed in dopaminergic neurons (GTEx Consortium, 2013). ZFX likely activates transcription and may play a role in cancer development (Rhie et al., 2018). However, additional research must be conducted in order to conclusively identify its function. The second insertion is within ZNF417 and ZNF574 binding sites. The functions of ZNF417 and ZNF574 were not identified in the literature.

G-Quadruplexes

The results presented support the hypothesis that G4s form in this repeat. The CD spectrum of 1445541R, which was predicted to form G4s by QGRS Mapper, clearly showed a spectrum consistent with G4 formation. All the characteristic features of a G4 in a CD spectrum, a peak at 210 nm, a trough at 240 nm, and a peak at 260 nm, were present. The magnitude of the spectrum increased with the addition of potassium, further confirming the presence of G4s. This spectrum had one feature that was not shown in previous studies, a peak at 290 nm, but it is unknown why this peak is present. The sequence utilized in this study differed from the one used in the protocol published by (Kejnovská et al., 2019), so differences in base composition might account for this difference. Even though this one feature is different, this spectrum still shows

that G4s are present because it has the required features. Another interesting difference was the ellipticity magnitude. The magnitude in this experiment was 35 times greater than spectra in the literature (Kejnovská et al., 2019). However, this may be due to the increased amount of DNA utilized. The protocol suggested using 1.35 μL of 10 mM DNA, but the spectrophotometer did not produce a signal with this low of concentration. The recommended amount of DNA may have been inadequate for this experiment because they utilized a different spectrophotometer (Kejnovská et al., 2019). Thus, the amount was increased to 20 μL of 100 mM DNA.

On the other hand, the CD spectrum of double stranded DNA (with 1445541R and 1445569F), did not have the characteristic features of G4s. However, this is expected, as previous studies found that G4s cannot be detected in double stranded DNA by CD spectroscopy (Kreig et al., 2015). The spectrum produced, however, did not match ones in the literature (Kreig et al., 2015). Previous studies exhibited a trough a 250 nm and a peak at 270 nm (Kreig et al., 2015). For more information about the shape of the line, see Figure S4 in (Kreig et al., 2015). In the spectrum produced by this experiment, the peak was closer to 265 nm. Additionally, there were two troughs, and they were at 220 and 240 nm. Like the previous spectrum, the ellipticity magnitude was much greater than in previous studies (Kejnovská et al., 2019). The magnitude in this experiment was 225 times greater than spectra in the literature (Kejnovská et al., 2019). Again, this may be due to the increased amount of DNA utilized. It is plausible that the magnitude was larger than the last experiment because double the amount of DNA was used.

The CD spectrum of 1445569F exhibits DNA that does not form G4s. The spectrum did not have the characteristics features of G4s, including a peak at 210 nm, a trough at 240 nm, and a peak at 260 nm. Additionally, the addition of potassium did not change the magnitude of the spectrum, further confirming this finding. This shows that 1445541R forms G4s because a

cytosine-rich strand did not produce the same spectrum under the same reaction conditions. Similar to the other spectra, the ellipticity magnitude was much 50 times than spectra in the literature (Kejnovská et al., 2019). It is unclear why the magnitude difference was different than the 1445541R spectrum because the same amount of DNA was used. However, two different sequences were used, so this could have played a role.

G4s are common in promoters, so this finding is plausible (Kolesnikova and Curtis, 2019). This finding could have implications for gene expression, as transcription factors that bind to G4s in promoters have been identified (Kolesnikova and Curtis, 2019). There are numerous transcription factor binding sites within the sequence studied. Thus, it is likely that transcription factors bind to this G4. It is unknown how G4 formation affects transcription factor binding for *SLC6A3*, so additional studies should be done to understand this relationship.

G4s may play a role in disease development, as mutations that disrupt G4 formation might contribute to disease development (Lee et al., 2020). However, it seems that mutations within the G4 motif are most important, and the mutations identified in this study do not appear to be within these motifs (Lee et al., 2020). G4s are known to be involved in cancer and neurological diseases (Nakanishi, & Seimiya, 2020; Simone et al., 2015). Cancer cells have greater G4 formation than healthy cells (Nakanishi, & Seimiya, 2020). G4s in DNA increase genetic instability, so this may accelerate tumor development (Nakanishi, & Seimiya, 2020). Additionally, expansion of a G4-forming hexamer repeat in *C9orf72* is the underlying disease mechanism for frontotemporal dementia and amyotrophic atherosclerosis (Simone et al., 2015). G4s may also play a role in Fragile X Syndrome, as the protein that is not expressed in the disease, FMRP, binds to G4s (Simone et al., 2015). However, this role is unclear as Fragile X Syndrome is caused by loss of *FMRP* (Simone et al., 2015). Given that G4s play a role in other

diseases, it is plausible that their formation plays a role in substance use disorder. However, studies must be done to understand the association between G4s in the *SLC6A3* promoter and substance use disorder.

Conclusion

This research increased the understanding of the promoter of *SLC6A3*. The number of repeats in the STR does not vary between individuals but does contain one SNP and three novel insertions. The SNP and two insertions are within a transcription factor binding sites likely to be expressed in dopaminergic neurons, but it is unclear how these would contribute to substance use disorder. Additionally, G4s form in this region. It is likely that transcription factors bind to this sequence, but the effect G4 formation has on gene expression is unclear. Further research should be done to not only determine the relationship between these findings and gene expression, but also the development of substance use disorder.

References

Ahmad FB, Rossen LM, Sutton P. Provisional drug overdose death counts. National Center for Health Statistics. 2021. Designed by LM Rossen, A Lipphardt, FB Ahmad, JM Keralis, and Y Chong: National Center for Health Statistics.

American Psychiatric Association. (2017). Diagnostic and statistical manual of mental disorders: Dsm-5. *DSM Library*. American Psychiatric Association. Retrieved November 15, 2021, from https://dsm.psychiatryonline.org/doi/full/10.1176/appi.books.9780890425596.dsm16.

Bakhtiari, M., Park, J., Ding, Y. C., Shleizer-Burko, S., Neuhausen, S. L., Halldórsson, B. V., Stefánsson, K., Gymrek, M., & Bafna, V. (2021). Variable number tandem repeats mediate the expression of proximal genes. *Nature communications*, *12*(1), 2075. https://doi.org/10.1038/s41467-021-22206-z

Bishop, G. R., & Chaires, J. B. (2003). Characterization of DNA structures by circular dichroism. *Current protocols in nucleic acid chemistry*, *Chapter 7*, . https://doi.org/10.1002/0471142700.nc0711s11

Brookes K. J. (2013). The VNTR in complex disorders: the forgotten polymorphisms? A functional way forward?. *Genomics*, *101*(5), 273–281. https://doi.org/10.1016/j.ygeno.2013.03.003

Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., Fornes, O., Leung, T. Y., Aguirre, A., Hammal, F., Schmelter, D., Baranasic, D., Ballester, B.,

Sandelin, A., Lenhard, B., Vandepoele, K., … Mathelier, A. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic acids research*, *50*(D1), D165–D173. https://doi.org/10.1093/nar/gkab1113

Dreos, R., Ambrosini, G., Cavin Périer, R., & Bucher, P. (2013). EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic acids research*, *41*(Database issue), D157–D164. https://doi.org/10.1093/nar/gks1233

Ebert, P., Audano, P. A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M. J., Sulovari, A., Ebler, J., Zhou, W., Serra Mari, R., Yilmaz, F., Zhao, X., Hsieh, P., Lee, J., Kumar, S., Lin, J., Rausch, T., Chen, Y., Ren, J., Santamarina, M., … Eichler, E. E. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science (New York, N.Y.)*, *372*(6537), eabf7117. https://doi.org/10.1126/science.abf7117

ENCODE Project Consortium, Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E. L., Freese, P., Gorkin, D. U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., Williams, B. A., … Weng, Z. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, *583*(7818), 699–710. https://doi.org/10.1038/s41586-020-2493-4

Faraone, S. V., Spencer, T. J., Madras, B. K., Zhang-James, Y., & Biederman, J. (2014). Functional effects of dopamine transporter gene genotypes on in vivo dopamine transporter functioning: a meta-analysis. *Molecular psychiatry*, *19*(8), 880–889. https://doi.org/10.1038/mp.2013.126

Fuke, S., Suo, S., Takahashi, N., Koike, H., Sasagawa, N., & Ishiura, S. (2001). The VNTR polymorphism of the human dopamine transporter (DAT1) gene affects gene expression. *The pharmacogenomics journal*, *1*(2), 152–156. https://doi.org/10.1038/sj.tpj.6500026

Goldman, D., Oroszi, G., & Ducci, F. (2005). The genetics of addictions: uncovering the genes. *Nature reviews. Genetics*, *6*(7), 521–532. https://doi.org/10.1038/nrg1635

Green, M. R., & Sambrook, J. (2019). Polymerase Chain Reaction. *Cold Spring Harbor protocols*, *2019*(6), 10.1101/pdb.top095109. https://doi.org/10.1101/pdb.top095109

Grey, C., Baudat, F., & de Massy, B. (2018). PRDM9, a driver of the genetic map. *PLoS genetics*, *14*(8), e1007479. https://doi.org/10.1371/journal.pgen.1007479

Guindalini, C., Howard, M., Haddley, K., Laranjeira, R., Collier, D., Ammar, N., Craig, I., O'Gara, C., Bubb, V. J., Greenwood, T., Kelsoe, J., Asherson, P., Murray, R. M., Castelo, A., Quinn, J. P., Vallada, H., & Breen, G. (2006). A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(12), 4552–4557. https://doi.org/10.1073/pnas.0504789103

Grünblatt, E., Werling, A. M., Roth, A., Romanos, M., & Walitza, S. (2019). Association study and a systematic meta-analysis of the VNTR polymorphism in the 3'-UTR of dopamine transporter gene and attention-deficit hyperactivity disorder. *Journal of neural transmission (Vienna, Austria : 1996)*, *126*(4), 517–529. https://doi.org/10.1007/s00702-019-01998-x

GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. *Nature genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Helin K. (1998). Regulation of cell proliferation by the E2F transcription factors. *Current opinion in genetics & development*, *8*(1), 28–35. https://doi.org/10.1016/s0959-437x(98)80058-0

Hill, M., Anney, R. J., Gill, M., & Hawi, Z. (2010). Functional analysis of intron 8 and 3' UTR variable number of tandem repeats of SLC6A3: differential activity of intron 8 variants. *The pharmacogenomics journal*, *10*(5), 442–447. https://doi.org/10.1038/tpj.2009.66

Hong, C. J., Cheng, C. Y., Shu, L. R., Yang, C. Y., & Tsai, S. J. (2003). Association study of the dopamine and serotonin transporter genetic polymorphisms and methamphetamine abuse in Chinese males. *Journal of neural transmission (Vienna, Austria : 1996)*, *110*(4), 345–351. https://doi.org/10.1007/s00702-002-0790-8

Huang, C. C., Lu, R. B., Yen, C. H., Yeh, Y. W., Chou, H. W., Kuo, S. C., Chen, C. Y., Chang, C. C., Chang, H. A., Ho, P. S., Liang, C. S., Cheng, S., Shih, M. C., & Huang, S. Y. (2015). Dopamine transporter gene may be associated with bipolar disorder and its personality traits. *European archives of psychiatry and clinical neuroscience*, *265*(4), 281–290. https://doi.org/10.1007/s00406-014-0570-0

Huang, S. Y., Chen, H. K., Ma, K. H., Shy, M. J., Chen, J. H., Lin, W. C., & Lu, R. B. (2010). Association of promoter variants of human dopamine transporter gene with schizophrenia in Han Chinese. *Schizophrenia research*, *116*(1), 68–74. https://doi.org/10.1016/j.schres.2009.10.004

Kejnovská, I., Renčiuk, D., Palacký, J., & Vorlíčková, M. (2019). CD Study of the G-Quadruplex Conformation. *Methods in molecular biology (Clifton, N.J.)*, *2035*, 25–44. https://doi.org/10.1007/978-1-4939-9666-7_2

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome research*, *12*(6), 996–1006. https://doi.org/10.1101/gr.229102

Kikin, O., D'Antonio, L., & Bagga, P. S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic acids research*, *34*(Web Server issue), W676–W682. https://doi.org/10.1093/nar/gkl253

Kolesnikova, S., & Curtis, E. A. (2019). Structure and Function of Multimeric G-Quadruplexes. *Molecules (Basel, Switzerland)*, *24*(17), 3074. https://doi.org/10.3390/molecules24173074

Kreig, A., Calvert, J., Sanoica, J., Cullum, E., Tipanna, R., & Myong, S. (2015). G-quadruplex formation in double strand DNA probed by NMM and CV fluorescence. *Nucleic acids research*, *43*(16), 7961–7970. https://doi.org/10.1093/nar/gkv749

Kypr, J., Kejnovská, I., Renciuk, D., & Vorlícková, M. (2009). Circular dichroism and conformational polymorphism of DNA. *Nucleic acids research*, *37*(6), 1713–1725. https://doi.org/10.1093/nar/gkp026

Lee, D., Ghanem, L. R., & Barash, Y. (2020). Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nature communications*, *11*(1), 527. https://doi.org/10.1038/s41467-020-14404-y

Lee, P. Y., Costumbrado, J., Hsu, C. Y., & Kim, Y. H. (2012). Agarose gel electrophoresis for the separation of DNA fragments. *Journal of visualized experiments : JoVE*, (62), 3923. https://doi.org/10.3791/3923

Ma, Y., Fan, R., & Li, M. D. (2016). Meta-Analysis Reveals Significant Association of the 3'-UTR VNTR in SLC6A3 with Alcohol Dependence. *Alcoholism, clinical and experimental research*, *40*(7), 1443–1453. https://doi.org/10.1111/acer.13104

Ma, Y., Yuan, W., Cui, W., & Li, M. D. (2016). Meta-analysis reveals significant association of 3'-UTR VNTR in SLC6A3 with smoking cessation in Caucasian populations. *The pharmacogenomics journal*, *16*(1), 10–17. https://doi.org/10.1038/tpj.2015.44

Mill, J., Asherson, P., Craig, I., & D'Souza, U. M. (2005). Transient expression analysis of allelic variants of a VNTR in the dopamine transporter gene (DAT1). *BMC genetics*, *6*, 3. https://doi.org/10.1186/1471-2156-6-3

Miller, D. R., Guenther, D. T., Maurer, A. P., Hansen, C. A., Zalesky, A., & Khoshbouei, H. (2021). Dopamine Transporter Is a Master Regulator of Dopaminergic Neural Network Connectivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *41*(25), 5453–5470. https://doi.org/10.1523/JNEUROSCI.0223-21.2021

Mortensen, O. V., & Amara, S. G. (2003). Dynamic regulation of the dopamine transporter. *European journal of pharmacology*, *479*(1-3), 159–170. https://doi.org/10.1016/j.ejphar.2003.08.066

Nakanishi, C., & Seimiya, H. (2020). G-quadruplex in cancer biology and drug discovery. *Biochemical and biophysical research communications*, *531*(1), 45–50. https://doi.org/10.1016/j.bbrc.2020.03.178

Persico, A. M., Vandenbergh, D. J., Smith, S. S., & Uhl, G. R. (1993). Dopamine transporter gene polymorphisms are not associated with polysubstance abuse. *Biological psychiatry*, *34*(4), 265–267. https://doi.org/10.1016/0006-3223(93)90081-n

Reith, M., Kortagere, S., Wiers, C. E., Sun, H., Kurian, M. A., Galli, A., Volkow, N. D., & Lin, Z. (2021). The dopamine transporter gene SLC6A3: multidisease risks. *Molecular psychiatry*, 10.1038/s41380-021-01341-5. Advance online publication. https://doi.org/10.1038/s41380-021-01341-5

Rhie, S. K., Yao, L., Luo, Z., Witt, H., Schreiner, S., Guo, Y., Perez, A. A., & Farnham, P. J. (2018). ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream of transcription start sites at the majority of CpG island promoters. *Genome research*, *28*(3), 310–320. Advance online publication. https://doi.org/10.1101/gr.228809.117

Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M. A., & Gemmell, N. (2013). Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PloS one*, *8*(2), e54710. https://doi.org/10.1371/journal.pone.0054710

Sherry, S.T., Ward, M. and Sirotkin, K. (1999) dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res.*, **9**, 677–679.

Shumay, E., Chen, J., Fowler, J. S., & Volkow, N. D. (2011). Genotype and ancestry modulate brain's DAT availability in healthy humans. *PloS one*, *6*(8), e22754. https://doi.org/10.1371/journal.pone.0022754

Simone, R., Fratta, P., Neidle, S., Parkinson, G. N., & Isaacs, A. M. (2015). G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS letters*, *589*(14), 1653–1668. https://doi.org/10.1016/j.febslet.2015.05.003

Sonne, J., Reddy, V., & Beato, M. R. (2021). Neuroanatomy, Substantia Nigra. In *StatPearls*. StatPearls Publishing.

Stevens, A. J., Taylor, M. G., Pearce, F. G., & Kennedy, M. A. (2017). Allelic Dropout During Polymerase Chain Reaction due to G-Quadruplex Structures and DNA Methylation Is Widespread at Imprinted Human Loci. *G3 (Bethesda, Md.)*, *7*(3), 1019–1025. https://doi.org/10.1534/g3.116.038687

Stothard, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques 28: 1102-1104.

Substance Abuse and Mental Health Services Administration. (2021). Key substance use and mental health indicators in the United States: Results from the 2020 National Survey on Drug Use and Health (HHS Publication No. PEP21-07-01-003, NSDUH Series H-56). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Tiili, E. M., Mitiushkina, N. V., Sukhovskaya, O. A., Imyanitov, E. N., & Hirvonen, A. P. (2020). The effect of SLC6A3 variable number of tandem repeats and methylation levels on individual susceptibility to start tobacco smoking and on the ability of smokers to quit smoking. *Pharmacogenetics and genomics*, *30*(6), 117–123. https://doi.org/10.1097/FPC.0000000000000403

Tirosh, I., Barkai, N., & Verstrepen, K. J. (2009). Promoter architecture and the evolvability of gene expression. *Journal of biology*, *8*(11), 95. https://doi.org/10.1186/jbiol204

Vandenbergh, D. J., Persico, A. M., Hawkins, A. L., Griffin, C. A., Li, X., Jabs, E. W., & Uhl, G. R. (1992). Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. *Genomics*, *14*(4), 1104–1106. https://doi.org/10.1016/s0888-7543(05)80138-7

Vandenbergh, D. J., Thompson, M. D., Cook, E. H., Bendahhou, E., Nguyen, T., Krasowski, M. D., Zarrabian, D., Comings, D., Sellers, E. M., Tyndale, R. F., George, S. R., O'Dowd, B. F., & Uhl, G. R. (2000). Human dopamine transporter gene: coding region conservation among normal, Tourette's disorder, alcohol dependence and attention-deficit hyperactivity disorder populations. *Molecular psychiatry*, *5*(3), 283–292. https://doi.org/10.1038/sj.mp.4000701

VanNess, S. H., Owens, M. J., & Kilts, C. D. (2005). The variable number of tandem repeats element in DAT1 regulates in vitro dopamine transporter density. *BMC genetics*, *6*, 55. https://doi.org/10.1186/1471-2156-6-55

Xiao, C., Zhou, C. Y., Jiang, J. H., & Yin, C. (2020). Neural circuits and nicotinic acetylcholine receptors mediate the cholinergic regulation of midbrain dopaminergic neurons and nicotine dependence. *Acta pharmacologica Sinica*, *41*(1), 1–9. https://doi.org/10.1038/s41401-019-0299-4

Xu, F. L., Ding, M., Wu, X., Liu, Y. P., Xia, X., Yao, J., & Wang, B. J. (2020). A Meta-analysis of the Association Between SLC6A3 Gene Polymorphisms and Schizophrenia. *Journal of molecular neuroscience : MN*, *70*(2), 155–166. https://doi.org/10.1007/s12031-019-01399-5

Xu, X., Mill, J., Sun, B., Chen, C. K., Huang, Y. S., Wu, Y. Y., & Asherson, P. (2009). Association study of promoter polymorphisms at the dopamine transporter gene in Attention Deficit Hyperactivity Disorder. *BMC psychiatry*, *9*, 3. https://doi.org/10.1186/1471-244X-9-3

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Zhao, Y., Xiong, N., Liu, Y., Zhou, Y., Li, N., Qing, H., & Lin, Z. (2013). Human dopamine transporter gene: differential regulation of 18-kb haplotypes. *Pharmacogenomics*, *14*(12), 1481–1494. https://doi.org/10.2217/pgs.13.141

Zhou, Y., Michelhaugh, S. K., Schmidt, C. J., Liu, J. S., Bannon, M. J., & Lin, Z. (2014). Ventral midbrain correlation between genetic variation and expression of the dopamine transporter gene in cocaine-abusing versus non-abusing subjects. *Addiction biology*, *19*(1), 122–131. https://doi.org/10.1111/j.1369-1600.2011.00391.x

Zhu, J., & Reith, M. E. (2008). Role of the dopamine transporter in the action of psychostimulants, nicotine, and other drugs of abuse. *CNS & neurological disorders drug targets*, *7*(5), 393–409. https://doi.org/10.2174/187152708786927877

Appendix A

Circular Dichroism Protocol

Materials

- 100 mM oligonucleotides (1445541R and 1445569F)
- 0.2 M Na_2HPO_4 Buffer
- 0.2 M NaH_2PO_4 Buffer
- 0.2 M K_2HPO_4 Buffer
- 0.2 M KH_2PO_4 Buffer
- 3M KCl

Denaturation of DNA and Preparation of Solutions

1. Turn on the PCR machine
2. Prepare DNA solution

Reagents	Amount
100 mM oligonucleotide	20 µL
0.2 M Na_2HPO_4	84.28 µL
0.2 M NaH_2PO_4	84.28 µL
	188.55 µL

Table 3: Single Stranded DNA Reaction Mix for Denaturation

Reagents	Amount
100 mM oligonucleotide	20 µL
100 mM oligonucleotide	20 µL
0.2 M Na_2HPO_4	74.28 µL
0.2 M NaH_2PO_4	74.28 µL
	188.55 µL

Table 4: Double Stranded DNA Reaction Mix for Denaturation

a. Vortex the solution
b. Centrifuge it to bring down all the liquid

c. Pipette this entire volume into a PCR tube

3. Denature the DNA
   a. Set thermocycler to 1 cycle for 90 degrees for 5 minutes
   b. Take sample out of the machine when the sample is complete and let it sit on the counter overnight. This will allow the DNA to come back to room temperature on its own.
4. Prepare potassium solutions
   a. Potassium Buffer Solution

Reagents	Amount
0.2 M K_2HPO_4	2.4 μL
0.2 M KH_2PO_4	2.4 μL
	4.8 μL

Table 5: Potassium Buffer Solution Reaction Mix for CD Spectra

   b. 6.66 μL of 3 M KCl
5. Prepare sodium buffer solutions

Reagents	Amount
0.2 M Na_2HPO_4	100 μL
0.2 M NaH_2PO_4	100 μL
	200 μL

Table 6: Sodium Buffer Solution Reaction Mix for CD Spectrum Baseline

Circular Dichroism

1. Preparation of Jasco J-1500 Spectrophotometer
   a. Open the nitrogen tank
   b. Turn on the water heater
   c. Turn on the spectrophotometer
   d. Log in to the computer
   e. Open CD program: SpectraMaster
   f. Allow the nitrogen tank to be open for 5 minutes before running an experiment
   g. Open the shutter
   h. Ensure the 1 mm cuvette is clean by holding it up to the light and inspecting it for any debris
2. Run a spectrum with sodium buffer solution for a baseline
   a. Pipette 200 μL of the sodium buffer solution into the 1 mm cuvette
   b. Place the lid on the cuvette and put it into the machine
   c. Ensure the hole in the shim is facing down to allow light to pass through

d. Open the parameters
e. Ensure that the cell unit temperature is set to 25°C
f. Measure the sample

3. Run a spectrum with the DNA and sodium buffer solution
   a. Dispose of the sodium buffer.
   b. Pipette 188.55 μL of the DNA and sodium buffer solution into the 1 mm cuvette
   c. Repeat steps b-f under the sodium buffer measurement instructions
   d. Process the spectra
      i. Subtract the sodium buffer spectrum from the DNA and sodium buffer spectrum
      ii. Collect the data and make a graph with it in Excel
4. Run a spectrum with DNA, sodium buffer, and potassium buffer solution
   a. Pipette 4.8 μL of potassium buffer into the cuvette with the DNA and sodium buffer solution
   b. Repeat steps c-d.ii under the DNA and sodium buffer measurement instructions
5. Run a spectrum with DNA, sodium buffer, potassium buffer, and potassium chloride solutions
   a. Pipette 6.66 μL of 3 M potassium chloride into the cuvette with the DNA, sodium buffer, and potassium buffer.
   b. Repeat steps c-d.ii under the DNA and sodium buffer measurement instructions

Convert Millidegrees to Ellipticity

1. Calculate the molar weight of the DNA sample
   a. Complete calculations are in Appendix B
2. Calculate the concentration DNA in the solution
   a. Complete calculations are in Appendix B

$$M_1V_1 = M_2V_2$$

3. Calculate ellipticity
   a. This equation was inserted into excel and the calculations were done with this program

$$\varepsilon = \frac{(mdeg\ x\ M)}{32980\ x\ C\ x\ L}$$

Appendix B

Unit Calculations

Molar Weight of 1445541R

Sequence contains 21 guanines, 6 cytosines, and 1 adenosine.

$$Deoxyguanosine\ monophosphate_{molecular\ weight} = 347.22\ \frac{g}{mol}$$

$$Deoxyadenosine\ monophosphate_{molecular\ weight} = 331.222\ \frac{g}{mol}$$

$$Deoxycytidine\ monophosphate_{molecular\ weight} = 307.1971\ \frac{g}{mol}$$

$$1445541R_{molecular\ weight} = \left(21\ x\ 347.22\ \frac{g}{mol}\right) + \left(1\ x\ 331.222\ \frac{g}{mol}\right) + (6\ x\ 307.1971\ \frac{g}{mol})$$

$$1445541R_{molecular\ weight} = 9466.0246\ \frac{g}{mol}$$

Molar Weight of 1445569F

Sequence contains 21 cytosines, 6 guanines, and 1 thymine.

$$Deoxythymidine\ monophosphate_{molecular\ weight} = 320.1926\ \frac{g}{mol}$$

$$1445569F_{molecular\ weight} = \left(21\ x\ 307.1971\ \frac{g}{mol}\right) + \left(1\ x\ 320.1926\ \frac{g}{mol}\right) + (6\ x\ 347.22\ \frac{g}{mol}$$

$$1445569F_{molecular\ weight} = 8854.6517\ \frac{g}{mol}$$

Concentration of DNA in Solution

$$M_1V_1 = M_2V_2$$

$$(100\ mM\ x\ 20\ \mu L) = (M_2\ x\ 200\ \mu L)$$

$$M_2 = 10\ mM$$

ACADEMIC VITA

Kathleen M. Cardone

EDUCATION

Bachelor of Science in Biobehavioral Health **May 2022**
The Pennsylvania State University Schreyer Honors College, University Park, PA
Minors: Neuroscience, Psychology, Biology
Dean's List: 7/7
Awards: Alumni Recognition for Student Excellence (April 2022)

RESEARCH EXPERIENCE

Undergraduate Research Assistant **November 2019-Present**
The Pennsylvania State University, Department of Biobehavioral Health, University Park, PA
Principle Investigator: David Vandenbergh, Ph.D.

- Studying genetic factors associated with substance use disorder.
- Analyzing the dopamine transporter gene using PCR and gel electrophoresis to identify variable tandem repeats.
- Utilizing online sequencing programs to analyze DNA sequences and primers.
- Utilizing software prediction tools circular dichroism to test for G-Quadruplexes.
- Analyzing publicly available sequencing data utilizing a python code and bioinformatic tools.
- Writing an honors thesis on findings.

INTERNSHIPS

Research Strategic Initiative Department Undergraduate Intern **June 2021-Present**
Geisinger Medical Center, Danville, PA
Mentors: Natasha Strande, Ph.D., Marina DiStefano, Ph.D, Melissa Kelly, Ph.D

- Analyzing variant curations to collect case data, segregation data, and functional data, and draw conclusions on these findings.
- Evaluated cancer genes on version 3 of American College of Medical Genetics & Genomics (ACMG) Secondary Findings List for Geisinger MyCode Community Health Initiative.
- Utilized online databases and scientific literature to collect data on genetic factors including gene-disease validity, clinical actionability, population-based prevalence and penetrance, mosaicism, and pseudogenes.
- Created an abstract and a poster on my research project.
- Presented findings at Susquehanna Valley Undergraduate Research Symposium.

Research and Development Summer Intern **May 2019-August 2020**
CRC Industries Inc., Warminster, PA
Supervisors: David Nicholson, Valerie Stephens

- Improved current products and developed new products at an aerosol-based chemical industrial company.
- Designed a research project measuring the differences between the company's distilled water and tap water and coauthored an internal scientific report on the findings.

- Conducted research on competitor products and presented findings to the marketing directors.
- Collaborated with multiple departments including quality control, regulatory, and upper management to solve a quality incident.

PRESENTATIONS

Cardone K, Mayk C, Kelly M, Strande N, DiStefano M. Evaluation of Cancer Genes on ACMG Secondary Findings List for Geisinger MyCode Project. Poster session presented at: Susquehanna Valley Undergraduate Research Symposium; 2021 July 28; Virtual.

TECHINCAL SKILLS

- Micropipetting
- Polymerase chain reaction
- Gel Electrophoresis
- Circular Dichroism
- Density
- pH
- Conductivity
- Moisture
- Gas Chromatography
- IR Spectroscopy
- Water Hardness
- Viscosity
- Plastic Safety Testing
- PrimerBLAST (software)
- UCSC Genome Browser (online database)
- NCBI Blast (software)
- Tandem Repeat Finder (software)
- Zotero (software)
- QGRS Mapper (software prediction tool)
- Mega11 (software)
- SnapGene (software)
- Microsoft Office 365
- Clinical Genome Resource (ClinGen) (online database)
- Gene-Disease Curation Coalition (GenCC) (online database)
- ClinVar (online database)
- GeneReviews (online database)
- Ensembl (software)
- IBM SPSS (software)

CAMPUS ENGAGEMENT

Pennsylvania State University Marching Blue Band **August 2018-January 2022**
The Pennsylvania State University, University Park, PA
Trumpet Section Leader

- Co-directed a group of 60 peers in a supervisory role while building a culture of efficacy, clear communication, and discipline
- Well versed in fast-paced environments and executing tasks quickly and skillfully by performing pregame and halftime shows at football games each week.
- Honed time management and organization by balancing 30 hours of practices and performances with an honors level course load.

Lead the Pride Leadership Workshop **November 2019**
The Pennsylvania State University, University Park, PA

- Learned how to network with prospective employers.

- Attended seminars on leadership, diversity, and networking.

College of Health and Human Development Ambassadors **September 2019-May 2020**
The Pennsylvania State University, University Park, PA

- Spoke to prospective students about my college experience during accepted student programs.
- Strengthened communication skills by speaking in front of large crowds of people.
- Developed professionalism by networking with Penn State alumni.
- Attended multiple workshops about leadership and speaking to prospective students.

COMMUNITY SERVICE

State College Community Theatre **January 2021-April 2021**

- Utilized graphic design software to make marketing materials for upcoming events
- Researched plays and musicals to collect information on requested shows

WORK EXPERIENCE

Schreyer Honors College Distinguished Honors Faculty Program. **November 2020-Present**
The Pennsylvania State University, University Park, PA
Faculty Assistant

- Coordinating events sponsored by the Penn State Schreyer Honors College in which faculty members share their work with honors students who want to learn about something outside of their major.
- Communicating professionally with faculty to schedule events.
- Composing event descriptions and advertisements for students in a professional manner.

Starbucks **November 2020-December 2021**

- Prepared food and drinks for customers.
- Worked in a fast-paced environment while maintaining quality and customer service.
- Completed monetary transactions.

REFERENCES

David Vandenbergh, Ph.D.
The Pennsylvania State University, University Park, PA
djv4@psu.edu
(814)-863-8430

Natasha Strande, Ph.D.
Geisinger Medical Center, Danville, PA
ntstrande@geisinger.edu
(440)-371-7582

Marina DiStefano, Ph.D.
Geisinger Medical Center, Danville, PA
mtdistefano@geisinger.edu
(724)-991-9360