The serotonin transporter gene (*SLC6A4*) shows differential regulation between children with ASD and typically developing children in a South African population

Michaella Hulley (HLLMIC014)

Thesis presented for the degree of Master of Science in the Department of Molecular and Cell Biology
Faculty of Science
University of Cape Town

Submitted: April 2016
Supervisor: Dr. Colleen O'Ryan
Department of Molecular and Cellular Biology

PLAGIARISM DECLARATION

I, Michaella Hulley, know the meaning of plagiarism and declare that all of the work in this thesis, save for that which is properly acknowledged, is my own.

1. I know that plagiarism is wrong. Plagiarism is to use another‘s work and pretend that it is one‘s own.
2. I have used the Harvard convention for citation and referencing. Each contribution to this thesis from the works of other people has been cited and referenced.
3. This thesis is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature: _____

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Colleen O'Ryan, for the opportunity to complete my Master's work. She has contributed to both the funding of this work and provided guidance throughout the completion of this thesis. I would also like to thank the remainder of the Molecular Autism Research in South Africa lab, 405, of the MCB department of UCT, including Dr. Sofia Stathopolous, Gerrit van Wyk and Marita de Waal, for their continued support throughout the completion of my Masters. Specific thanks to Dr. Colleen O'Ryan, Dr. Sofia Stathopoulos, Dr. Shane Murray and Gerrit van Wyk for their instrumental help and suggestions in the writing process of this thesis and to Dr. Laura Roden, of the Biological Rhythms and Blooms lab at UCT, for her continued guidance throughout the completion of this work. Additional acknowledgement should go to Inqaba Biotech in Pretoria for their instrumental help in the setting up of the EpiTYPER assay and allowing me to complete this in their laboratory. I would also like to thank the parents and participants involved in this study for their contribution and time. Thank you to my parents for providing me with this opportunity and their constant support. Lastly, thank you to the National Research Foundation (NRF) for awarding me the NRF Scarce Skills scholarship, without which I would not have been able to complete this work.

ABSTRACT

Autism spectrum disorders (ASD) are characterised by impairments in social interaction and communication, and repetitive and compulsive behaviours. The heterogeneous nature of ASD is underpinned by complex genetic networks; hence there is no known single genetic cause of ASD. It is therefore imperative to identify both genetic and epigenetic loci associated with ASD or specifically ASD endophenotypes. This may aid in earlier interventions for ASD if molecular biomarkers were identified. The serotonergic system has a longstanding association with ASD, and the differential expression of the serotonin transporter gene (*SLC6A4*) is linked to particular ASD traits. This study focuses on a length polymorphic repeat region (5-HTTLPR) upstream of the *SLC6A4* promoter, and *SLC6A4* promoter DNA methylation. The short variant of 5-HTTLPR is directly linked to lowered *SLC6A4* expression and is at higher frequencies in ASD populations. DNA methylation at the promoter of *SLC6A4* is also known to affect *SLC6A4* expression.

This study hypothesis is that there is differential regulation of *SLC6A4*, through changes in 5-HTTLPR and DNA methylation, between an ASD and control cohort. This differential regulation was also predicted to differ between ASD endophenotypes based on severity levels categorised by ADOS-2 (Autism Diagnostic Observation Schedule-2). ASD children (n=50) were compared to an age-matched control group (n=13), all of whom were characterised phenotypically by an ADOS-2 assessment. The 5-HTTLPR and *SLC6A4* promoter DNA methylation were analysed using restriction fragment length polymorphic analysis and the EpiTYPER assay, respectively.

Significant differences were found in *SLC6A4* regulation between the ASD and control group. A significant increase in frequency of individuals homozygous for the 5-HTTLPR long variant in the control group was observed when compared to the ASD cohort (p=0.049); with the long allele conferring reduced risk of ASD. Overall DNA methylation at the *SLC6A4* promoter region was significantly decreased in the ASD cohort (p=0.011), the moderate ASD endophenotype (p=0.004), language impaired endophenotype (p=0.003), as well as the more severe socially and repetitive and restricted behaviour groups (p=0.006 and 0.045, respectively), when compared to the control cohort. Reduced levels of DNA methylation at CpG 30 in the target region were found to be significantly associated with higher levels of repetitive behaviour (p=0.001).

The data from this study implicates the involvement of the serotonin transporter in overall ASD aetiology, specifically within the language impaired and repetitive and restricted behaviour endophenotypes. The data highlight the importance of maintaining appropriate methylation levels in order to modulate *SLC6A4* expression. The regulatory mechanisms that control DNA methylation at the *SLC6A4* promoter are unknown and need to be identified to completely understand how dysfunction of the serotonergic system is involved in ASD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1. Complex phenotypes and Autism Spectrum Disorders (ASD)

Human genetic studies on complex traits and phenotypes have primarily focussed on DNA sequence variants or polymorphisms, with comprehensive and curated collation of this data on record (Amberger *et al*, 2015). These polymorphisms contribute to conferring susceptibility to disorders, such as Fragile X syndrome and sickle cell anaemia (Hessl *et al*., 2008; Ingram, 1957), and moderate responses to targeted therapy. However, there has been a recent trend of analysing epigenetic regulation of gene expression to assess the potential interplay with environmental factors. This environmentally induced epigenetic regulation of genes has been observed in a number of complex phenotypes, such as schizophrenia (Dong *et al*., 2015) and other central nervous system disorders (Szyf, 2015). These complex phenotypes likely result from interplay among multiple causal genes with both epigenetic and environmental influences. This interplay complicates the identification of specific causes or exact molecular contributions towards the aetiology of certain disorders, resulting from the polygenic nature of these disorders (Marian, 2012). A large number of neuropsychiatric disorders are polygenic, such as schizophrenia and bipolar disorder (Neale and Sklar, 2015). These disorders display large phenotypic heterogeneity; this is especially true for Autism Spectrum Disorders.

Autism Spectrum Disorders (ASD) encompass a range of complex neuro-developmental disorders affecting the central nervous system (Lintas *et al*., 2012). ASD is characterised by persistent impairments in core domains of social communication and social interaction, as well by the presence of restricted and repetitive acts or behaviours. Individuals with ASD fall upon a spectrum; with various symptoms of ASD, such as restricted behaviour and impaired social communication, ranging from mild to severe. ASD is a life-long condition, typically diagnosed in early childhood, between 12-24 months of age (American Psychiatric Association, 2013).

Global estimations place the incidence or diagnosis of ASD ranging from 0.6% to 1% of the population (Brugha *et al*., 2011; Williams *et al*., 2006; Elsabbagh *et al*., 2012; Sun *et al*., 2013; Hirtz *et al*., 2007), with a 4:1 male to female bias (Hirtz D *et al*., 2007). Between 2002 and 2008, the prevalence of ASD increased by 78% (Autism and Developmental Disabilities Monitoring Network, 2008). This most likely results from possible inclusion of below-threshold cases, differences in methodology, increased awareness, as well as a true ASD incidence rise (American Psychiatric Association, 2013). Regardless of the cause for this increased incidence or diagnosis of ASD, effective alleviative treatment of ASD relies on early intervention methods.

Early diagnosis of ASD allows for quick intervention practices leading to an improvement in social, communication and cognitive impairments, as well as a decrease in any maladaptive behaviour (Landa, 2007; Rogers and Vismara, 2008). However, most diagnoses occur long after symptoms have presented (Howlin and Asgharian, 1999; Wiggins *et al*., 2006). Both the diagnostic system and intervention services present in developing countries, such as South Africa, are limited. This places pressure on such services available; with data on incident rates, prevalence and ASD impact almost non-existent (Malcolm-Smith *et al*., 2013). Therefore, studies should focus on identifying methodology to aid in the diagnosis of ASD at earlier stages in children. In particular, identifying molecular biomarkers could aid in anticipating the specific set of impairments that could present in children. This is pertinent for children with ASD worldwide, but particularly in undeveloped countries, where resources are sorely lacking. Earlier intervention and therefore more targeted treatment to alleviate the negative effects of ASD are imperative. In order to accomplish this, one needs to address several challenges faced that come hand in hand with the study of ASD.

Challenges of studying ASD: heterogeneity and heritability

The main challenge of studying ASD is the phenotypic heterogeneous nature of this spectrum disorder. The phenotypic heterogeneity of ASD suggests a genetically complex molecular underpinning. Multiple genes or molecular pathways may be at play simultaneously, as well as in accordance with both epigenetics and environmental factors (Persico and Bourgeron, 2006; Tordjman *et al*., 2014). This

heterogeneity complicates any genetic or biological analysis (Hu *et al*., 2009). Therefore, some studies have attempted to reduce this phenotypic heterogeneity by clustering the complex phenotypes into more well-defined endophenotypes based on ASD and trait severity (Hu *et al*., 2009; Hu and Steinberg, 2009). The focus of ASD studies would then be on ASD-like traits as opposed to ASD as a whole. This allows for specific characteristics of ASD to be associated with certain genetic, environmental or epigenetic influences and could then aid in the identification of genes or causal pathways. Therefore, potential biomarkers may be identified that could be associated with specific traits of ASD. This in turn, may contribute to the effectiveness of the ASD diagnostic system.

The current ASD diagnostic measures rely solely on clinical diagnoses. Therefore, ASD diagnostic criteria rely largely on the identification of a set of observable characteristics (Mandy *et al*., 2014). This is made ambiguous by the phenotypic heterogeneity of ASD, as well as by ASD co-morbidities. These include depression, attention deficit hyperactivity disorder, speech and language development delays and other psychiatric disorders (Morsi *et al.,* 2011). Due to the lack of understanding surrounding the molecular underpinnings of ASD, no current diagnostic biomarkers are applicable (Mandy *et al*., 2014).

Another challenge of studying ASD concerns the complex nature of ASD heritability. Twin studies have attempted to determine the degree of heritability by examining the environmental and genetic factors that play a role in ASD aetiology. Certain monozygotic twin studies estimate that heritability is as high as 70%-90% (Robinson *et al*., 2012; Bailey *et al*., 1995). However, a study by Hallmayer et al. (2011) reported lower heritability for monozygotic twins with a probandwise concordance rate of 0.58. This study also reported a higher than previously reported concordance rate for dizygotic twins (Hallmayer *et al*., 2011). These reported rates suggest a larger role for environmental factors in ASD aetiology than initially proposed. Heritability estimates are further made ambiguous by other influences that may reduce the genetic similarity between monozygotic twins. These include *de novo* mutations, copy number variations (CNVs) and any epigenetic changes occurring in one twin only (Bruder *et al*., 2008). On the contrary, other factors work to enhance the genetic similarity between dizygotic twins. This includes the shared gestational

environment that allows for comparable epigenetic alterations (Titlestad *et al*., 2002). However, genetic polymorphisms are still considered the main contributing factor, with over 100 genomic loci identified with a contributory role in ASD aetiology (Betancur, 2011). Most genetic-based research into the aetiology of ASD, primarily through genome wide association studies, has not been conclusive (Newschaffer *et al*., 2007; Sykes and Lamb, 2007), likely due to the potential interplay with environmental factors through epigenetic mechanisms.

1.2. Genetic/epigenetic interplay

It is estimated that only 10-20% of ASD cases have known genetic causes; these include chromosomal abnormalities, copy number variants and single causal genes (Geschwind, 2011). The apparent increasing incidence of ASD cases highlights the value of investigating a potential interplay between the environment and genetics in ASD aetiology (Newschaffer *et al*., 2007; CDC, 2012). This environmental contribution is hypothesized to act via epigenetic mechanisms. These mechanisms allow for reversible heritable changes that alter or regulate the expression of genes without changing the DNA sequence (Gropman and Batshaw, 2010). Many non-coding DNA regions are imperative in the modulation of gene transcription and are often under the control of epigenetic regulation. This control is primarily observed in developmental cellular processes, including cell differentiation, tissue arrangement and the regulation of certain cell lines (Gropman and Batshaw, 2010). These processes result from changes in the patterns of gene expression in differing cells that ultimately govern a cell‘s developmental fate. Furthermore, these epigenetic mechanisms allow for the interplay between genetics and the environment by responding to various environmental cues and subsequently controlling gene function (Van Vliet *et al*., 2007). For example, an epigenetic model based on imprinting in Angelman and Prader-Willi syndrome, has been proposed for ASD aetiology (Jiang *et al*., 2004). Therefore, investigation of these epigenetic regulatory mechanisms is important to understanding the molecular basis of the development of ASD and ASD-like traits. The main epigenetic mechanism implicated in ASD aetiology is that of DNA methylation, typically observed in the promoter regions of various candidate genes (Schanen, 2006; Samaco *et al*., 2005; Lopez-Rangel and Lewis, 2006).

DNA Methylation

DNA methylation is the covalent chemical modification of a methyl group being added to the 5th position of the cytosine pyridine ring (Figure 1.1). This generally occurs when the cytosine is positioned to the 5‘ end of a guanine; known as a CpG site (Dhingra *et al*., 2014). DNA methylation is central in the regulation of cell differentiation and alters protein transcription that affects downstream pathways (De Leon-Guerrero *et al*., 2011; LaSalle, 2011). This is because methylation can alter the interactions between DNA and proteins, ultimately leading to chromatin restructuring, and either an increase or decrease in transcriptional activity. The addition of the methyl group through DNA methylation can also allow for the binding of specific proteins that can then regulate transcriptional activity (Jones and Takei, 2001). CpG islands comprise of clustered CpG sites that are generally non-methylated (Larson *et al*., 1992), usually located upstream of the genes, or in the promoter regions (De Leon-Guerrero *et al*., 2011; LaSalle, 2011). Methylation of these regions can lead to altered gene regulation and transcription.

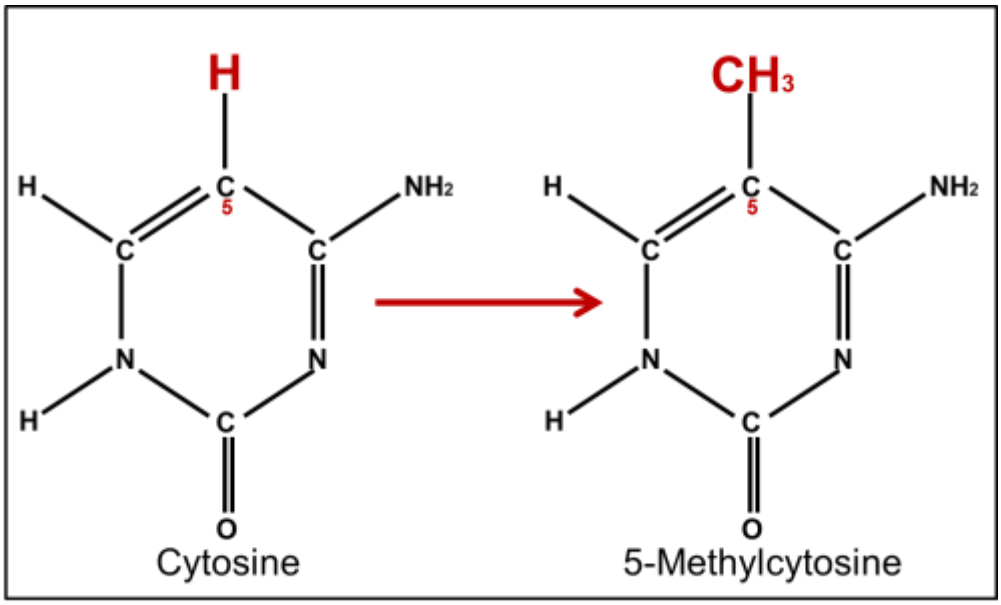

Figure 1.1. Schematic diagram of the chemical conversion of a cytosine to 5-methyl-cytosine during DNA methylation. The unmethylated cytosine is to the left, while the methylated cytosine, now 5-methylcytosine is to the right.

The role of DNA methylation in gene regulation has long been hypothesized as a contributing factor in psychiatric disorders (Tsankova *et al*., 2007; Grayson *et al*., 2013; Mill *et al*., 2008). Several studies have shown that early life stress, such as childhood abuse, modulates regulation of gene expression via differential methylation patterns in the brain which can then alter behaviour (McGowan *et al*., 2009; Reaume and Sokolowski, 2011). Additionally, Wong *et al*. (2014) completed a comprehensive analysis of DNA methylation between monozygotic twins discordant for ASD and ASD-traits, as well as between ASD cases and controls. They observed a greater number of CpG sites with large methylation differences between ASD-discordant twins than within control twins. Wong *et al*. (2014) also observed DNA methylation differences between their ASD cohort and control individuals. These studies have provided platforms to identify various candidate genes and pathways implicated in the development of ASD and other neuropsychiatric disorders.

Candidate genes/pathways

Multiple genes and metabolic pathways have been implicated in ASD aetiology from studies of both DNA methylation regulation, and genetic polymorphisms. For example, a study by Zhu *et al*. (2014) implicated epigenetic dysregulation of *SHANK3* in ASD individuals. SHANK3 is involved in connecting neurotransmitter receptors and other membrane proteins to the G-protein-coupled signalling pathways (Boeckers *et al*., 2002). This study observed significantly increased DNA methylation and therefore altered expression of the *SHANK3* gene in 15% of brain tissue of ASD cases. Other studies highlighted differential DNA methylation patterns between ASD individuals and any non-ASD siblings (Wong *et al*., 2014; Nguyen *et al*., 2010). Epigenetic dysregulation of the retinoic acid-related orphan receptor alpha gene (*RORA*) within ASD was identified (Nguyen *et al*., 2010). RORA acts within the transcriptional regulation of genes involved in circadian rhythms (Ko and Takahashi, 2006). In the first global methylation analysis of ASD using 3 pairs of monozygotic twins and 2 pairs of siblings, Nguyen et al. (2010) showed that increased methylation at the RORA promoter was responsible for decreased expression of the gene in ASD. This study also showed a reduction in RORA protein in post-mortem brain tissues in ASD individuals compared to controls (Nguyen *et al*., 2010). These studies support the role of underlying epigenetic mechanisms in the development of ASD.

More specifically, one of the main pathways identified in neuropsychiatric disorders, including ASD and ASD-like traits, involves the serotonergic system (Dolen *et al*., 2013). Altered serotonergic signalling has been implicated in a variety of conditions, such as depression (Philibert *et al*., 2008), obsessive compulsive disorder (Voyiaziakis *et al*., 2011) and anxiety disorders (Katsuragi *et al*., 1999; Lesch *et al*., 1996; Naslund *et al*., 2015). This supports the involvement of the serotonergic system within neuropsychiatric disorders. In addition, many attributes of the above mentioned disorders are also found in certain ASD cases, highlighting the need for further analysis of the role of the serotonergic system within an ASD population.

1.3. Serotonergic signalling and the serotonin transporter (SLC6A4)

Serotonin (5-HT) is an integral monoamine neurotransmitter involved in serotonergic signalling within the central and peripheral nervous systems. The raphe 5-HT system moderates mood, motor function, cognition and circadian and neuroendocrine rhythms. These rhythms are involved in food intake, sleep and reproductive activity (Heils *et al*., 1996). Furthermore, the serotonergic system plays an initiatory role in signalling during development and maturation of tissues, including regulation of neurite growth, neuronal differentiation and structure synaptogenesis and neurogenesis (Madden and Zup, 2014, and references within). Thus, the functioning of the serotonergic system is crucial for typical neural development. Therefore, any disruptions to this system may lead to susceptibility to the aforementioned neuropsychiatric disorders.

One of the key regulators and therefore most susceptible to disrupting the serotonergic system is the serotonin transporter protein (Lesch *et al*., 1996). This protein, called the 5-hydroxytryptamine transporter (5-HTT) is encoded by the *SLC6A4* (solute carrier family 6, member 4) gene. It directs the scale and duration of post-synaptic responses. In addition, it is involved in carrier-facilitated transport of serotonin in and out of the presynaptic neuron (Heils *et al*., 1996). The specific mode of action of the serotonin transporter and the location of its coding gene on chromosome 17 is observed in an informative diagram from Canli and Lesch (2007) in Figure 1.2. This diagram illustrates the main regulatory role of this transporter within the serotonergic system. Disrupted functioning of the serotonergic system,

potentially through changes in serotonin transporter action, may then confer susceptibility to ASD and other neuropsychiatric disorders.

Link to ASD

The specific link of the serotonergic system to ASD stems from a variety of findings. Firstly, reduced binding of the serotonin transporter was observed in specific brain regions of individuals with ASD (Makkonen *et al*., 2008; Murphy *et al*., 2006). ASD has also had longstanding associations with several polymorphisms in genes working within the serotonergic system, including *SLC6A4* (Huang *et al*., 2008; Kistner-Griffin *et al*., 2011; Guhathakurta *et al*., 2009; Hranilovic *et al*., 2010). In addition, a study by McDougle *et al*. (1996) observed increased ASD-related sensory motor behaviours following the depletion of tryptophan, which is involved in the production of serotonin. This implicates changes in behaviours associated with ASD with altered levels of serotonin.

Other ASD behaviours, such as abnormal reciprocal social interaction and repetitive behaviours were also improved in some ASD cases after exposure to selective serotonin reuptake inhibitors (SSRIs). These act directly on the serotonin transporter (Gordon *et al*., 1993; Kolevzon *et al*., 2006; McDougle *et al*., 1996). SSRIs are similar in structure to serotonin, and therefore compete for binding of the serotonin transporter leading to an increase in the levels of synaptic serotonin. Due to the alleviation of ASD-related symptoms through the use of SSRIs, serotonergic system malfunctioning is again implicated with ASD and other psychiatric ASD comorbid disorders.

Gastrointestinal problems are also a known comorbidity of ASD (Parracho *et al*., 2005). Serotonin acts as a paracrine signalling molecule at the gastrointestinal level via serotonergic interneurons. Up to five types of receptors that play a role in intestinal peristalsis, brain-gut axis signalling and secretion are activated by serotonin. Thus, abnormal levels of serotonin within this system can affect all of these actions. This may result from the overproduction of serotonin, therefore excessive release, as well as inadequate uptake of serotonin and subsequent inactivation. Therefore, any modifications or alterations to the action of the serotonin transporter may impair gastrointestinal motility (Yeo *et al*., 2004).

In addition, the observed male to female bias of ASD may also be partially accounted for by improper regulation of the serotonergic system (Madden and Zup, 2014, and references within). The modulating fundamental mechanisms by serotonin were established very early on and differ between males and females, with a sexually dimorphic profile of serotonin release and metabolism. A relatively stable level of serotonin is observed in the female brain in early postnatal development, while the male brain shows a temporary and significant decrease in serotonin (Giulian *et al*., 1973; Wilson *et al*., 1986). Therefore, neurological sexual differentiation resulting in physiological and behavioural changes relies on appropriate regulation of serotonin signalling. Through the introduction of SSRIs, increases of serotonin signalling are observed which impacts upon cellular and behavioural processes. This increase is more significant in males (Rodriguez-Porcel *et al*., 2011; Simpson *et al*., 2011). Therefore, Madden and Zup (2014) suggest that the developing brain of males, especially during early postnatal development, is more sensitive and therefore susceptible to any alterations to normal serotonin signalling. This in turn may be at least a contributing factor to the observed male bias for ASD.

Lastly, one of the main linking factors of the serotonergic system to ASD involves elevated whole blood serotonin levels being observed in up to 35% of individuals with ASD (Hranilovic *et al*., 2007). The cause of this observed blood serotonin increase is unknown, but may be the result of serotonin production and metabolism differences of individuals during foetal development (Madden and Zup, 2014, and references within). These elevated levels of blood platelet serotonin have also been hypothesized to result from a change in the serotonin transporter and its function (Cook *et al*., 1993).

Repeat polymorphism of SLC6A4

One of the most studied polymorphisms of the serotonin transporter, which alters its function, is a repeat polymorphic region. This repeat polymorphism is called the serotonin-transporter linked polymorphic region (5-HTTLPR), and found in the promoter of *SLC6A4* upstream of the protein coding region (Lesch *et al*., 1996). The 5-HTTLPR is comprised of two variants, a short allele (S) and a 44bp longer allele (L), shown in Figure 1.2. The presence of the L allele, a 16 repeat, or S allele, a 14

repeat, can affect *SLC6A4* mRNA expression (Delbruck *et al*., 1997; Heils *et al*., 1996; Lesch *et al*., 1996). The S allele also appears to have an overriding mode of action compared to the L allele and therefore the genotypes SS and LS are often grouped together in comparison studies (Hu *et al*., 2006). The L allele also contains an *A* to *G* single nucleotide polymorphism (SNP), rs25531. The A variant (L_A) is associated with increased *SLC6A4* mRNA expression, while the G variant (L_G) acts more similarly to the S allele. This substitution creates an allelic polymorphism with three potential variants (Hu *et al*., 2006; Wendland *et al*., 2006), with the frequency distributions of this polymorphism varying across populations (Arieff *et al*., 2010).

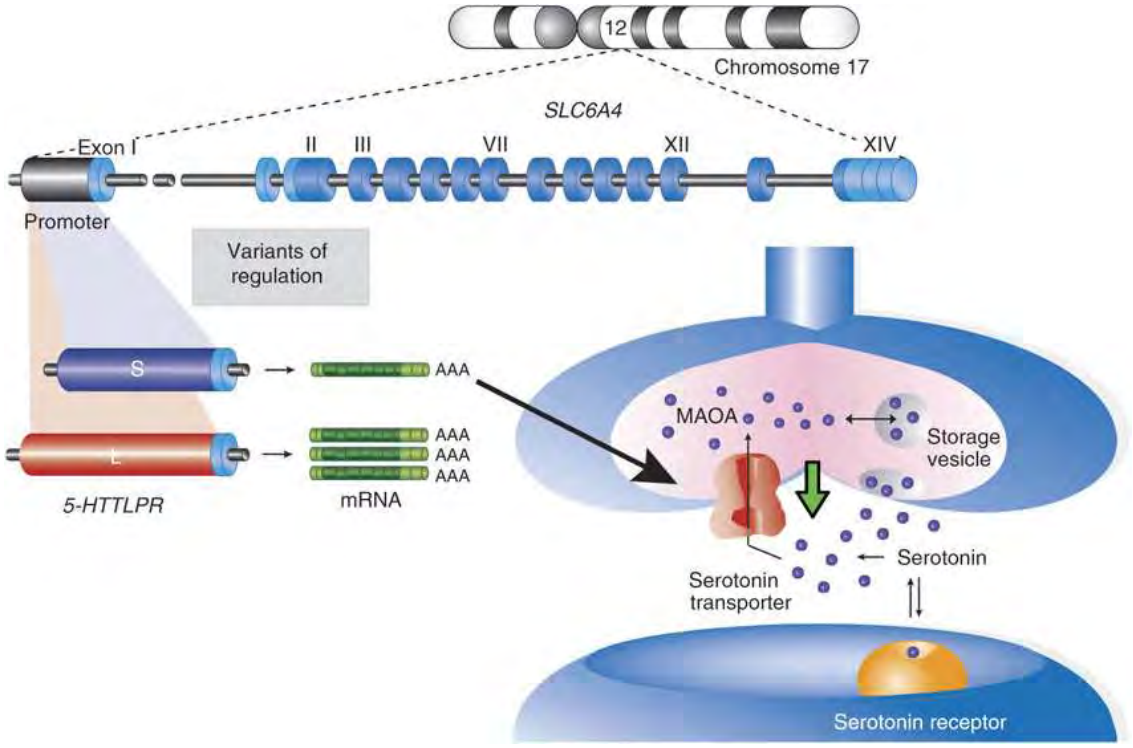

Figure 1.2. Diagram of the SLC6A4 gene and transporter function. Position on chromosome 17, allelic variation of the repeat polymorphic region (5-HTTLPR) and serotonin reuptake mechanism is shown. Image is taken from Canli and Lesch (2007). The S and L alleles are shown in blue and red, respectively. The S allele leads to lowered mRNA transcripts of SLC6A4 compared to the mRNA transcripts from the long allele. The carrier-facilitated transport of serotonin is indicated by the arrow.

South African population based studies report an increased presence of the S allele in an ASD population. A study by Esau *et al*., (2008) reported genotypic frequencies of the 5-HTTLPR in a control population, with the homozygous LL individuals at 61.40%, the LS heterozygous at 33.92% and the homozygous SS at 4.68%. A second study by Arieff *et al*., (2010) proceeded to do the same for an ASD cohort. They observed frequencies of 52.00% for homozygous LL individuals, 18.00% for LS heterozygotes and 30.00% for the homozygous SS individuals. This increased number of homozygous SS individuals in an ASD population suggests a role for the S allele, and reduced *SLC6A4* function, in ASD development.

The S allele has also been shown to be associated with various ASD-like traits, such as anxiety (Canli and Lesch, 2007; Lesch *et al*., 1996) and rigid and compulsive behaviours (Lin, 2007; McDougle *et al*., 1998; Sutcliffe *et al*., 2005). Several studies have attempted to identify a link between the 5-HTTLPR and ASD as a whole, but this has proven inconclusive. Studies have shown an association between the S allele and ASD (Kistner-Griffin *et al*., 2011; Arieff *et al*., 2010), while others implicate the L allele with ASD (Cho *et al*., 2007; Yirmiya *et al*., 2001). However, other studies were unable to replicate or corroborate these findings (Ma *et al*., 2010; Persico *et al*., 2002; Ramoz *et al*., 2006). The difficulty in finding clear, replicable associations with the 5-HTTLPR may be due to the phenotypic heterogeneity of ASD.

Studies have attempted to narrow these phenotypic differences using a more focused approach. For example, Nyffeler *et al*., (2014) focused their study on high functioning ASD cases only, with specific reference to intellectual and language ability, as well as cognitive function. They reported an association; with the S allele as a significant risk factor in ASD aetiology (Nyffeler *et al*., 2014). However, the lack of a conclusive gene variant association implicates other mechanisms, such as epigenetic DNA methylation, in the aetiology of ASD.

Promoter DNA methylation of SLC6A4

Studies have attempted to link the serotonergic system to ASD, and other neuropsychiatric disorders, not only through the 5-HTTLPR, but also by analysing promoter DNA methylation of the *SLC6A4* gene. Interestingly, DNA methylation of *SLC6A4* shows an intrinsic gender difference, with females reportedly having higher

methylation levels at birth compared to males (Philibert *et al*., 2008; Beach *et al*., 2010). The cause of this difference and specific effects remain unclear, however this is in line with the involvement of DNA methylation within ASD aetiology, and the male bias observed.

Focus has also been on how the methylation status of the *SLC6A4* promoter interacts with the 5-HTTLPR. *SLC6A4* expression is significantly moderated through chromatin remodelling, which includes DNA methylation of CpG cytosine residues. Increased CpG methylation at the promoter region found upstream of *SLC6A4* has previously been found to correlate with reduced *SLC6A4* mRNA levels (Philibert *et al*., 2007) as well as brain serotonin synthesis (Wang *et al*., 2012). Therefore, DNA methylation induces a similar effect to that observed by the presence of the S allele of the 5-HTTLPR. A longitudinal study by Wong *et al*., (2010), posited that the variation of DNA methylation observed at the *SLC6A4* locus likely results from environmental factors which may account for interindividual DNA methylation differences. However, other studies have demonstrated that the presence of the S allele can influence DNA methylation levels of the *SLC6A4* promoter (Philibert *et al*., 2007; Vijayendran *et al*., 2012). This supports the notion that *SLC6A4* DNA methylation is regulated to a certain extent by the status of the 5-HTTLPR.

Furthermore, many studies have attempted to elucidate the impact of environmental factors on DNA methylation of *SLC6A4*. One study showed that higher levels of methylation of *SLC6A4* were associated with increased adult reactivity following infantile exposure to early life stress in *rhesus macaques* (Kinnally *et al*., 2011). Beach *et al*. (2010) connected increased DNA methylation of the *SLC6A4* promoter with childhood physical abuse, and this correlation was replicated in a second study that focused on women (Beach *et al*., 2011). Lower DNA methylation levels of *SLC6A4* have also been associated with an increased susceptibility to posttraumatic stress disorder (PTSD). Higher DNA methylation levels create a protective effect against PTSD (Koenen *et al*., 2011). However, higher DNA methylation levels in the *SLC6A4* promoter were also associated with increased trauma in 5-HTTLPR homozygous LL individuals, while fewer traumas are observed in homozygous SS individuals (van Ijzendoorn *et al*., 2010). DNA methylation of *SLC6A4* has been implicated in the moderation of responses to stress, as well as in depression

(McGowan *et al*., 2009; Caspi *et al*., 2003). These studies are indicative of a more complex interaction between the 5-HTTLPR and environmental factors which influence the DNA methylation status of the *SLC6A4* promoter.

One of the most promising regions for potential differential DNA methylation of *SLC6A4* that may interact with the 5-HTTLPR is shown in Figure 1.3. The relative position of CpG sites to the transcription start site (TSS) and 5-HTTLPR is indicated. This image is taken from a study by Wang *et al*., (2012) who analysed DNA methylation at the promoter of *SLC6A4* and found an association with childhood physical aggression. This study observed hypomethylation within this CpG Island. A mean DNA methylation percentage ranging from 4.32% to 5.78%, depending on tissue type and an overall range from 1.17% to 15.89% DNA methylation was reported. This study also suggested a functional role of DNA methylation in *SLC6A4* promoter regulation, with *in vitro* methylation inducing transcriptional activity reduction (Wang *et al*., 2012).

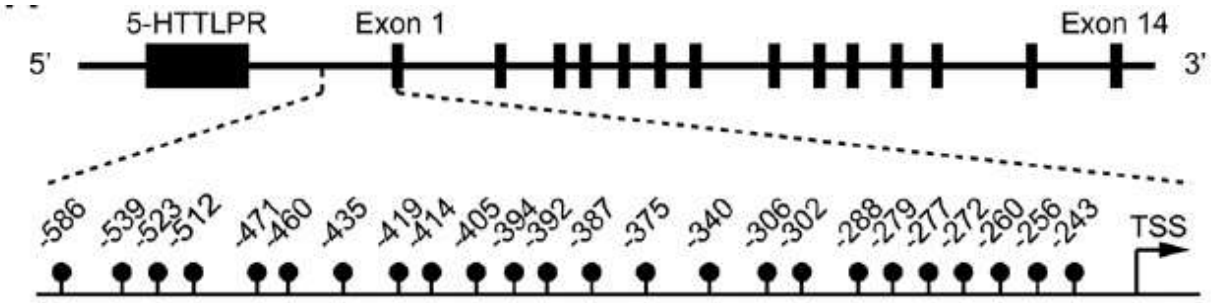

Figure 1.3. SLC6A4 promoter region investigated by Wang *et al*. (2012), showing positions of the CpG sites relative to the TSS and the 5-HTTLPR. DNA methylation at this region is associated with changes in *SLC6A4* mRNA expression.

The *SLC6A4* promoter region is rich in CpG sites, with several CpG islands across a 799bp stretch of DNA, comprising of 81 CpG sites (Philibert *et al*., 2007). A specific CpG-rich region contains 7 CpG sites situated between 479bp and 350bp relative to the transcription start site (accession number: BankIt1577778 *SLC6A4* KC106430) that have been associated with changes in mRNA expression of *SLC6A4* (Philibert *et al*., 2007).

1.4. Research Question and Study Aims

The overarching research question of this study is whether there is differential molecular regulation of *SLC6A4* between a) an ASD and control cohort, and b) different endophenotypes of ASD. The first objective of this study was to determine the endophenotypes of the ASD cohort using ADOS-2 (Autism Diagnostic Observation Schedule-2), which provides the severity level of ASD, as well as individual scores for social affect (SA) and repetitive and restricted behaviour (RRB). Well-defined, smaller subgroupings of the ASD cohort based on these scores allow for easier identification of any associations between the molecular underpinnings and ASD trait severity. This alleviates ASD phenotypic heterogeneity difficulties, and will hopefully allow for the identification of causal mechanisms for ASD traits. This may then inform preventative or alleviative treatments. The second objective was to determine 5-HTTLPR and *SLC6A4* promoter DNA methylation status of each participant and calculate the allelic and genotypic frequencies and mean methylation levels for ASD and control cohorts, and within the ASD-trait endophenotypes.

Several hypotheses are tested in this study. The main hypothesis states that there will be differential regulation of the *SLC6A4* gene between the ASD and control cohorts and across the ASD endophenotypes. Secondary hypotheses are as follows: The genotypic and allelic frequencies of the 5-HTTLPR will differ between ASD and control cohorts and endophenotypes. Then, the S allele will be enriched in the ASD cohort and more severely affected ASD-trait endophenotypes, as studies show S allelic-associations to certain ASD traits (Canli and Lesch, 2007; Lin, 2007; McDougle *et al*., 1998; Sutcliffe *et al*., 2005; Lesch *et al*., 1996). Additionally, differential mean DNA methylation will be observed at individual CpG sites and across the investigated region in the ASD cohort compared to the control cohort, as well as between the ASD-trait endophenotypes. Finally, an interaction between the genotypic and allelic frequencies and promoter DNA methylation of *SLC6A4* will be observed.

These hypotheses will be tested through comparisons of the allelic and genotypic frequencies and mean DNA methylation levels across the two main cohorts (ASD and control) and the ASD endophenotypes using statistical tests. The identification of differential regulation within ASD and the ASD endophenotypes could then

implicate *SLC6A4* regulation within the aetiology of ASD and with ASD-trait development. This could allow for the determination of biomarkers within the ASD diagnostic system, which may eventually aid in earlier diagnosis or intervention of ASD.

CHAPTER 2
Methods and Materials

2.1. Sample cohort & phenotypic information

A cohort of both children with ASD (n=50) and age-and gender-matched controls (n=13) was built for this study. At the time of this study, only 13 control participants were recruited based on volunteer parents, while the 50 children with ASD were recruited from schools and sport clubs in Cape Town, Western Cape (33.9253° S, 18.4239° E). This included a Western Cape Government autism specialist school and a private autism centre. All participants were male, with an age range of 6 to 14 years old. These parameters were chosen because of the 4 to 1 male to female incidence of ASD, and to avoid the confounding factor of sex-dependent changes in DNA methylation patterns present from birth. The study was granted both Western Cape Government approval (Ref: 2014002-37506) and UCT Ethics Approval (FSREC 076-2014); informed parental consent was obtained from all study participants (See Appendix for consent form).

Fifty of the children were recruited from the specialist autism schools, and they all had a working diagnosis of ASD from the local specialist children‘s hospital, a team of educational psychologists and occupational therapists at the respective schools or a neuro-paediatrician. The entire cohort, both the children with an ASD diagnosis, and age-matched controls, underwent an ADOS-2 assessment (Autism Diagnostic Observation Schedule-2). This was conducted by a team of certified assessors, who were also research reliable to ensure that the ADOS scores were comparable both among assessors and over time. This allows for the ADOS-2 classification of autism, autism-spectrum and non-spectrum (control) for each child. An ADOS-2 severity comparison score is also assigned, classifying the participant‘s ASD as ASD-high, ASD-moderate or ASD-low when compared to other children with ASD of a similar developmental age. The ADOS-2 assessment also provides two other useful scores: a) SA, which is comprised of the child‘s communication and reciprocal social interaction, and b) RRB. These core traits are associated with ASD and used in the ADOS-2 algorithm that determines ADOS-2 classification.

Children were excluded from the study cohort if they displayed any co-morbidity, such as Rett Syndrome or known chromosomal disorders (e.g. Fragile X Syndrome or Neurofibromatosis), or if they had a sibling already in the study. Control candidates were excluded if they scored above threshold on ADOS-2.

Buccal cell collection and genomic DNA extraction

DNA was obtained from buccal cells, a non-invasive method of sample collection. Buccal epithelial cells have the advantage of being derived from the embryonic ectodermal layer, the same embryonic layer from which neuronal cells are derived. Hence, it has been suggested that buccal cells are the ideal target tissue in neuro-developmental studies (Olsson *et al*., 2010). DNA methylation patterns of epithelial buccal cells were also found to be more similar to those observed in brain tissue than observed in blood (Lowe *et al*., 2013).

Buccal cells were collected using Catch All Sample Collection Swabs (Epicentre Biotechnologies, Madison, USA) from each participant over at least three separate occasions. Each cheek of the participants was swabbed thoroughly for approximately 30 seconds, with the swab placed in storage buffer comprising of 540uL Lysis buffer (0.4M NaCl, 10mM Tris-HCl, pH 8.0, and 2mM EDTA, pH 8.0) and 60uL of 10 % (w/v) sodium dodecyl sulphate (1% final concentration SDS) immediately. The buccal cells could then be stored at 4°C for a maximum period of two weeks before DNA extraction.

DNA was extracted from these buccal cells using the extraction protocol described in Aljanabi and Martinez (1997) with a few modifications. The first being the addition of 10μL of 20 mg/mL proteinase K to the lysis buffer containing the swab, followed by incubation at 56°C for two hours. After incubation, 225μL of 6M saturated NaCl was added, followed by two rounds of centrifugation at 13,200rpm for 10 minutes on an Eppendorf centrifuge 5415R. The supernatant was placed into a fresh eppendorf after each spin. Total genomic DNA (gDNA) was precipitated by the addition of 800uL isopropanol and incubated at -20°C overnight, followed by a 70% (v/v) ethanol wash step. The precipitated gDNA was resuspended in 40μL of sterile, molecular grade water, and stored at 4°C until use to avoid degradation by the freeze/thaw cycle, with long-term storage at -20°C. The gDNA quantity and quality was assessed

using the Nanodrop ND-1000 spectrophotometer (Nanodrop ® Technologies, USA). Quality was then confirmed by visualisation of gDNA on 0.8% (w/v) agarose gel in 1X Tris Borate EDTA (TBA; Tris, EDTA and boric acid) buffer and stained with ethidium bromide.

2.2. Genetic Methodology: SLC6A4 polymorphisms

The *SLC6A4* region investigated is visualised in Figure 2.1. All amplicons analysed are also illustrated, along with their corresponding coverage of the 5-HTTLPR and CpG sites.

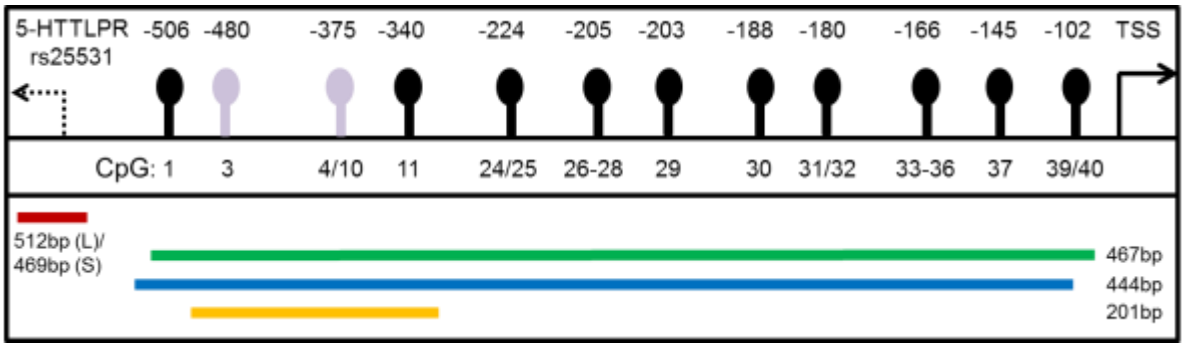

Figure 2.1. Representative image of the target region of the *SLC6A4* promoter region for both the genetic and epigenetic analysis. The TSS is represented by a solid arrow, while the 5-HTTLPR is a broken arrow and upstream of these CpG sites. Each CpG site analysed is represented by a solid circle. CpG sites associated with *SLC6A4* mRNA expression are coloured purple. The fragments of DNA amplified for the 5-HTTLPR analysis (red), bisulphite sequencing (yellow), native sequencing (blue) and the EpiTYPER assay (green) are shown along with relative coverage of the CpG sites and position to TSS.

Repeat polymorphism & SNP rs25531

The DNA fragment of the promoter of *SLC6A4* containing the 5-HTTLPR and the SNP rs25531 is shown in red in Figure 2.1. This fragment was amplified under the following conditions: In a total volume of 25μL, 100ng of DNA were amplified in the presence of 2.5μL of 10× KAPA Taq buffer, 1μL of 10uM dNTP mix, 0.75μL of each primer and 0.2μL of 5U/μL KAPA Taq DNA polymerase (Kapa Biosystems, Boston, USA), with sterile, distilled water making up the reaction to 25μL. The reaction cycling conditions were as follows: initial denaturation at 95 °C for 3 minutes, with 35 cycles of 95°C for 30 seconds, 65°C for 45 seconds, and 72°C for 45 seconds; followed by 10 minute final elongation at 72°C, using the primers described in Dukal *et al*. (2015; See Appendix Table A). The 5-HTTLPR genotype was then identified

using electrophoresis of the PCR product on a 2.5% (w/v) agarose gel in a 1 X TBE buffer stained with ethidium bromide. A 512bp amplicon was expected for the L allele and a 469bp amplicon for the S allele.

The SNP, rs25531, on the L allele of 5-HTTLPR, was then determined through restriction fragment length polymorphism (RFLP) analysis by digesting 20μL of the PCR product with *HpaII* (New England Biolabs) per manufacturer's instructions. This allows for the detection of the rare G allele of rs25531, as the DNA sequence, 5' – *CCAG – 3'*, would be altered to the restriction recognition site, 5' – *CCGG – 3'*. The status of rs25531 was resolved on a 2.5% (w/v) agarose gel in a 1 X TBE buffer stained with ethidium bromide. The L_A and S alleles were expected to remain uncut, while the L_G allele was expected to show amplicons of 402bp and 110bp in length.

2.3. Epigenetic Methodology: DNA Methylation

The DNA methylation analysis was completed in two ways, first qualitatively and the second, quantitatively. Both methods involve sodium bisulphite treatment of the DNA. When treated with sodium bisulphite, any unmethylated cytosines are converted to uracil residues, while methylated sites would be protected and remain as cytosines. The qualitative method involves bisulphite sequencing, developed by Frommer *et al*. (1992) and allows for the detection of methylation-dependent sequence variations. Primers are altered to match the converted DNA. A native fragment (unconverted DNA) of the target region is also sequenced in order to identify potential SNPs which may account for any variations that could create or destroy a CpG site. The second method involves the use of the EpiTYPER assay, conducted on the Sequenom MassARRAY system (Agena Bioscience, San Diego, CA). EpiTYPER is a mass spectrometry bisulphite sequencing method. It allows for quantitative measurement of DNA methylation levels at a single-nucleotide resolution (Suchiman *et al*., 2015). This was also completed, because the qualitative methodology was not sufficient to detect low levels of DNA methylation.

2.3.1. Qualitative analysis

Bisulphite converted DNA sequencing

In order to analyse CpG methylation status, gDNA was treated with sodium bisulphite using the EZ DNA Methylation Direct Kit per manufacturer's instructions

(Zymo Research, Orange, CA, USA). The converted DNA quantity was then assessed using the Nanodrop ND-1000 spectrophotometer (Nanodrop ® Technologies, USA), using the single stranded DNA setting due to the converted nature of the DNA.

A 201bp region, covering the seven CpG sites associated with mRNA expression of *SLC6A4*, and shown in yellow in Figure 2.1, was amplified. In a total volume of 25μL, 100ng of DNA were amplified in the presence of 5μL of 5× KAPA Taq HotStart buffer, 1μL of 10uM dNTP mix, 1μL of each primer and 0.2μL of 5U/μL KAPA Taq HotStart DNA polymerase, with sterile, distilled water making up the reaction to 25μL. The reaction cycling conditions were as follows: initial denaturation at 95 °C for 15 minutes; with 40 cycles of 95°C for 15 seconds, 57°C for 30 seconds, and 72°C for 15 seconds; followed by 5 minute final elongation at 72 °C. Primers described by Kim *et al.* (2013) were used in the amplification, altered according to the converted DNA sequence (See Appendix Table A). Amplification confirmation was completed using gel electrophoresis by running 5uL of the product on a 1.5 % (w/v) agarose gel in a 1 X TBE buffer stained with ethidium bromide. The converted PCR product was then purified from the agarose gel using the Zymoclean^{TM} Gel DNA Recovery Kit (Zymo Research, Orange, CA, USA), and sent to the Centre for Analytical Services (CAF) at the University of Stellenbosch (Stellenbosch, Western Cape, South Africa) for BigDye sequencing on the BigDye Terminator V3.1 sequencing kit (Applied Biosystems) with only slight modifications to the manufacturers‘ protocol. CAF utilises the ABI 3730xl to run the amplicons, as well as the AB Foundation DATA Collection v3.0 and Sequence Analysis v5.3 to complete the data collection and perform sequence analysis. The chromatograms were visualised and edited using Chromas software (downloaded from http://technelysium.com.au/wp/chromas on 11 September 2015, (Technelysium Pty Ltd, Tewantin QLD, Australia).

Native DNA Sequencing

A 444bp target region, containing the CpG sites to be investigated, was amplified using a high-fidelity DNA polymerase in order to sequence the native DNA. This allows for sequence comparison to bisulphite treated DNA. The intended target region is shown in blue in Figure 2.1. In a total volume of 20μL, 20ng of DNA was amplified in the presence of 4μL of 5× Phusion GC buffer, 0.4μL of 10uM dNTP mix,

1μL of each primer and 0.2μL of Phusion DNA polymerase (Thermo Scientific, Vilnius, Lithuania), with sterile, distilled water making up the reaction to 20μL. The two-step reaction cycling conditions were as follows: initial denaturation at 98 °C for 30 sec, with 35 cycles of 98 °C for 10 sec and 72°C for 35 sec; followed by 10 min final elongation at 72 °C. The forward primer 5' – GCCGGTCAGTCAGATAAACG – 3' and reverse primer 5' – CGTCACTTTGAGGCGAATAAA – 3' used were designed in Primer3Plus Design (Untergasser *et al*., 2012; See Appendix Table A). Confirmation of amplification was completed using gel electrophoresis by resolving 5μL of the product on a 1.5 % (w/v) agarose gel in a 1 X TBE buffer stained with ethidium bromide.

The remaining PCR product underwent electrophoresis on a 0.8% (w/v) agarose gel in a 1 X TBE buffer stained with ethidium bromide, in order for the product at the 444bp region to be cut out. The DNA was then extracted from these gel slices using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Orange, CA, USA). The gel extracted amplicons were sent for BigDye sequencing, as described above for the converted DNA. The sequences were aligned using Molecular Evolutionary Genetics Analysis (MEGA) software (downloaded from http://www.megasoftware.net on 15 September 2015; Stecher *et al*., 2013), in order to identify any SNPs that may alter the sequence. Potential transcription factor binding sites that may overlap with CpG sites were predicted using Alibaba 2.1, a program developed by Niels Grabe (http://www.gene-regulation.com/pub/programs/alibaba2/index.html; TRANSFAC® Public).

2.3.2. Quantitative analysis

EpiTYPER MassARRAY Methylation Assay

The EpiTYPER assay involves the examination of PCR-amplified converted DNA with a mass spectrometer. The basic steps involve treating genomic DNA with sodium bisulphite, amplifying the converted DNA with primers tagged with a T7 promoter and treating this with shrimp alkaline phosphatase (SAP). RNase A cleavage follows along with transcription of the amplicon into single RNA strands with the T7 promoter. A final cleaning step involves resin and the samples are then loaded onto a SpectroCHIP R II Array, preparing the samples for separation with the mass spectrometer. The mass spectrometer is a MALDI-TOF device and makes use of laser energy to ionize the RNA fragments. The time of flight is then recorded and

increases with the 16 Da higher mass of a methylated CpG site compared to an unmethylated one. The size of the two peaks can be used in the calculation of the methylation ratio of each target CpG site. Detection of certain CpG sites is limited when fragments of the target region are too similar in mass to another fragment. The final result is given as a DNA methylation percentage of each CpG site (Suchiman *et al*., 2015).

The EpiTYPER assay was conducted on the Sequenom MassARRAY system (Agena Bioscience, San Diego, CA) at Inqaba Biotec (Pretoria, South Africa). Genomic DNA was treated with sodium bisulphite using the EZ DNA Methylation Kit (Zymo Research, Orange, CA, USA) as per manufacturer's instructions. The gDNA quantity and quality was assessed using the Nanodrop ND-1000 spectrophotometer (Nanodrop ® Technologies, USA) prior to conversion, while the post-treatment analysis was completed at Inqaba using Picogreen RNA settings (PicoGreen®).

A 467bp region of the 5-HTT promoter was amplified using bisulphite-conversion based PCR primers, designed in EpiDesigner Beta (www.epidesigner.com): 5'–TTGTTAGGTTTTAGGAAGAAAGAGAGA–3' and 5'– CCCTCACATAATCTAATCTCTAAATAACC – 3' (See Appendix Table A). This region is shown in green in Figure 2.1, and covers 21 of the 40 CpG sites for the massARRAY purposes. This is because certain fragments are too similar in mass. In a total volume of 9μL, 1μL of 10ng/μL (10ng) of DNA was amplified in the presence of 1μL of 10× PCR buffer with 20mM $MgCl_2$, 0.08μL of 25mM dNTP mix, 2μL of each primer, 0.08μL of PCR Enzyme (5U/μL) and 2.84μL sterile, distilled water. The specific PCR enzyme and buffer makeup is copyrighted to the assay and service provider. Amplification was performed at the following cycling conditions: initial denaturation at 94°C for 4 minutes, 45 cycles of 94°C for 20 seconds, 56°C for 30 seconds and 72°C for 1 minute; followed by the final extension for 3 minutes at 72°C. This was also completed for a negative control sample (without bisulphite-treated DNA), as well as controlled methylated DNA at levels of 0%, 25%, 75% and 100%. A T7 promoter-tagged reverse primer was used to allow for *in vitro* transcription at a later stage.

The manufacturer's instructions were followed for the remainder of the EpiTYPER assay. This includes the SAP treatment in order to dephosphorylate unincorporated

dNTPs in the reaction, which was incubated at 37°C for 20 minutes and 85°C for 5 minutes. Following this, the *in vitro* RNA transcription with subsequent T (U)-specific cleavage using RNase A was completed to yield fragmented RNA molecules, through incubation of 37°C carried out for 3 hours. Both methylated and non-methylated regions are cleaved on the reverse strand only, allowing equal length fragments which differ only in the nucleotide composition.

The products were then conditioned for the MALDI-TOF mass spectrometer using Clean Resin, optimising mass spectrometry analysis and transferred to a SpectroCHIP array using the MassARRAY nanodispenser, whereby the different signals generated by the mass difference between the non-methylated and methylation fragments were recorded. In this assay, the mass difference detected by the MassARRAY system would result from guanine/adenine changes rather than cytosine/thymine changes due to the use of the reverse strand. The Analyser program of EpiTYPER 1.2 (Agena Bioscience, San Diego, CA) was used to display the spectra and produce the quantitative methylation results, which included percentage methylation per CpG site of the whole tissue type.

2.4. Statistical Analysis

Hardy-Weinberg exact tests were used to determine whether the 5-HTTLPR L and S allelic frequencies were in Hardy Weinberg equilibrium (HWE) (Simple Hardy-Weinberg Calculator - Court Lab). All remaining statistical analyses were completed using IBM SPSS Statistics 22 (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp), with statistical significance accepted at $p<0.05$. Genotypic and allelic frequency distributions for comparisons across the phenotypic grouping were completed using the two-tailed Fisher‘s exact test, due to the small sample sizes of the ASD endophenotypes. Odds ratios and relative risk associations were completed for the ASD and control cohort comparison to determine the 5-HTTLPR allele risk. The Mann-Whitney U test was used to compare the quantitative DNA methylation results, both median and spread of data, across the comparison groups, due to the non-normal distribution of DNA methylation data and the sample size constraints (Hart, 2001). The quantitative data were reported as the mean methylation percentage per site or region ± standard deviation and presented using box-plots.

CHAPTER 3
Results

3.1. Cohort and ASD endophenotypes

ADOS-2 assessments provided detailed, standardised phenotypic information for sixty three individuals. Using the exclusion criteria and ADOS-2 assessments, 25 of the children diagnosed with ASD and 7 of the control participants did not meet the requirements, highlighting the value of completing an ADOS-2 assessment on both cohorts. The resultant, clearly phenotyped group included participants with ASD (n=25) and age-matched controls (n=6). A summary of the ages and ancestry of the final cohorts are shown in Appendix Table B. The participants were from varied ancestral backgrounds and the mixed ancestry (an admixed population group including African, European and Asian ancestry) and African ancestry groups were combined for statistical analysis purposes due to the small sample size (n=3) of the African group.

The ADOS-2 assessment was used to reduce the phenotypic heterogeneity of the ASD cohort. The ASD cohort was sub-grouped using the ADOS-2 severity comparison scores, the SA scores and the RRB scores. The SA score is comprised of social communication and reciprocal social scores. The choice of which ADOS-2 module to administer is based on the level of expressive language of the participant. Module 1 was used when the participant was pre-verbal or used single words only, while module 2 was used for participants who used phrase speech. Module 3 was used for participants with fluent language, as well as for the entire control cohort. The ADOS-2 module was also used as one of the ways to create ASD language endophenotypes. The resultant ASD endophenotypes used in this study are summarised in Table 3.1. These endophenotypes represent smaller subgroups of the ASD cohort according to specific ASD-traits, thereby reducing the phenotypic heterogeneity observed in ASD as a whole.

Table 3.1. Endophenotypes of the ASD cohort and control cohort using criteria based on ADOS-2

Phenotypic groups and ASD-trait Endophenotypes	ADOS-2 criteria	Sample Size (n)
ASD	ASD on ADOS-2	25
Control	No score on ADOS-2	6
ASD-High	Comparison score ≥ 8	5
ASD-Moderate	Comparison score 5 ≥ n < 8	20
Pre-verbal/Single words	Module 1 completed	14
Phrase Speech	Module 2 completed	8
Fluent Language	Module 3 completed	3
High SA	ADOS-2 SA score ≥ 8	23
Low SA	AODS-2 SA score < 8	2
High RRB	ADOS-2 RRB score ≥ 6	6
Low RRB	ADOS-2 RRB score < 6	19

ASD-Autism spectrum disorder; SA-Social affect; RRB-Repetitive and restricted behaviour; with high and low referring to the level of severity of ASD and ASD-traits.

The endophenotypes listed in Table 3.1 were used in all subsequent genotype-phenotype association analyses. The genotypic and allelic frequencies of the 5-HTTLPR and *SLC6A4* DNA methylation levels were compared across all the ASD endophenotypes using the two tailed Fisher‘s exact test and Mann-U Whitney test, respectively. This was completed to identify any associations between the S or L alleles of the 5-HTTLPR, and levels of *SLC6A4* promoter DNA methylation, with ASD or specific ASD-traits. Pairwise comparisons were done between the control participants and entire ASD cohort, as well as between the control cohort and various ASD endophenotypes. These ASD endophenotypes are ASD-high, ASD-moderate, high SA, high RRB, pre-verbal and phrase speech. The high RRB endophenotype was also compared to the low-RRB endophenotype. Only two ASD participants had low SA scores and three had fluent language, therefore these two endophenotypes were not used in the pairwise comparisons. The detailed results of these comparisons are listed below.

3.2. Repeat polymorphism and SNP rs25331

The 5-HTTLPR and rs25531 SNP in the promoter region of *SLC6A4* were genotyped using RFLP for twenty four of the ASD participants, while all six participants of the control cohort were genotyped. Only twenty four of the twenty five ASD participants were used for RFLP due to shortage of DNA. The genotypic and allelic frequency distributions were calculated. Tests for deviation from HWE were performed for the genotype frequencies at the 5-HTTLPR locus and there was no significant deviation from HWE. This test was completed for the entire cohort, as well for ASD and control cohorts separately (p=0.18, 0.09 and 0.62, respectively).

Repeat polymorphism and SNP rs25531in the ASD and control cohorts

In the ASD cohort, five of the participants were LL homozygous (20.8%); sixteen were LS heterozygous (66.7%), and three were SS homozygous (12.5%). In the control cohort, two of the participants were LS heterozygous (33.3%), while the remaining four control participants were LL homozygous (66.7%). The SS homozygous participants were combined with the LS heterozygous participants to calculate a frequency distribution representing genotypes associated with lowered *SLC6A4* expression, since the S allele is thought to act in a dominant manner (Hu *et al*., 2006). These are the "low expression" individuals, while the LL participants are considered the "high expression" individuals. When the rs25531 SNP is taken into account, any LL participants with at least one G variant of the SNP at the *SLC6A4* promoter are added to the "low expression" group. None of the LL participants in the ASD cohort contained the G variant, while one of the control LL participants had the rs25531 SNP and was thus added to the "low expression" *SLC6A4* group in the analysis. The L/S allelic frequencies and genotypic frequency distributions are shown in Figure 3.1.

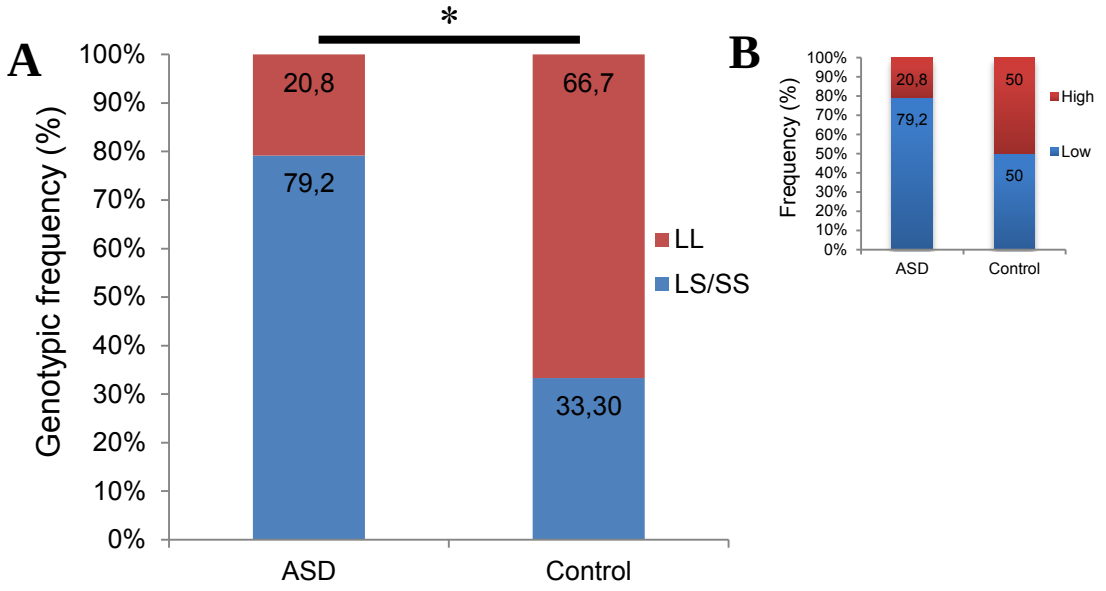

Figure 3.1. Frequency distributions for the 5-HTTLPR genotypes (A) and rs25531genotype modified-expression (B) of participants in the ASD (n=24) and control cohorts (n=6). LL, LS and SS represent the three 5-HTTLPR genotypes, while High and Low represent high expression and low expression of *SLC6A4,* taking the effect of rs25531 into account. These were compared between the ASD cohort (n=24) and control cohort (n=6) using the two-tailed Fisher's exact test. Significance is denoted by * ($p<0.05$).

The genotypic frequency distribution based on the 5-HTTLPR genotypes, where any S allele containing genotypes were combined, was found to be significantly different between the ASD cohort and the control cohort ($p=0.049$, refer to Figure 3.1). An enrichment of the S allele-containing genotypes in the ASD cohort is observed, with a 2.38-fold increase in the frequency of these genotypes (79.2%) compared to the control cohort (33.30%). However, the presence of the rs25531 G variant for one of the LL control participants increased the frequency within the "low expression" group. The adjusted frequency distribution based on the incorporation of the rs25531 SNP into the calculations was not found to be significantly different between the ASD and control cohorts ($p=0.3$), with only a 1.58-fold increase in "low expresser" frequency observed (79.2% for the ASD cohort and 50% for the control cohort).

The significant difference in the genotypic frequency distribution is supported by the odds ratio/relative risk data. The relative risk assessment revealed that the presence of the L allele significantly decreases the risk for ASD to a relative risk of 0.78. (95% confidence interval: 0.622-0.997). This decreased relative risk is observed with a

3.20-fold increase of the LL genotype (Figure 3.1) in the control population compared to the ASD cohort. The relative risk is based on the allelic distribution, however it is must be noted that the allelic frequency distribution was not found to be significantly different between the two cohorts using the two-tailed Fisher's exact test (p=0.1), as seen in Figure 3.2. Although, a 1.5-fold increase in the L allele in the control cohort does correspond to a 2.7-fold decrease of the S allele compared to the ASD cohort, showing an observed enrichment of the S allele in the ASD population compared to the control cohort.

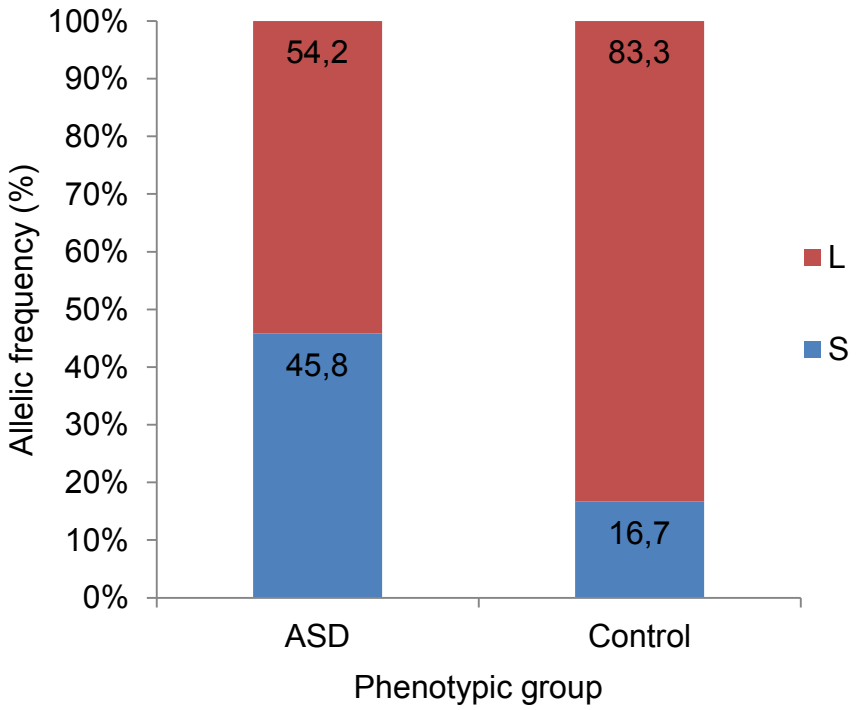

Figure 3.2. Frequency distribution for the 5-HTTLPR L and S alleles in the ASD (n=24) and control cohorts (n=6). The allelic frequencies were compared between the two cohorts using the two-tailed Fisher's exact test (Significance set at $p<0.05$).

Repeat polymorphism and SNP rs25531in the ASD endophenotypes

Both genotypic frequency distributions and the allelic frequencies were also calculated for the all endophenotypes listed in Table 3.1 and then used in the aforementioned comparisons using the two-tailed Fisher's exact test (See Appendix Table C). However, none of these comparisons yielded any significant results, with only the more severely affected ASD endophenotype, ASD-high, approaching the closest significant difference of genotypic frequency distribution (LL vs. LS/SS,

p=0.061) when compared to the control cohort. The comparison of the frequency distributions between participants of different ancestry (African/mixed ancestry and European ancestry) also did not yield any significant differences (p=0.226).

In summary, the presence of the L allele appears to significantly reduce the risk of developing ASD. This is supported by both the genotypic comparison of the LL vs. LS/SS in an ASD cohort to a control cohort and the relative risk result. A significant enrichment of S-allele containing genotypes as well as a 2.7-fold increase of the S allele is also observed in the ASD cohort.

3.3. SLC6A4 Promoter DNA Methylation

The native (or not converted) sequences for both the ASD and control cohorts were aligned using the multiple sequence alignment tools in MEGA (See Appendix Figure A). This allowed for the identification of SNPs in both cohorts‘ native DNA sequences. No SNPs were found to overlap with any of the CpG sites. Only two of the participants contained a single sequence variant and were heterozygous for a *T - C* SNP, rs25533. This SNP was identified in one individual with ASD and one control participant.

Using the native *SLC6A4* promoter sequence as a reference, Alibaba 2.1 predicted potential DNA regions that five transcription factors could potentially bind. There are 15 potential binding sites for these transcription factors in this region covering all but two of the CpG sites. These transcription factors are the: specificity protein 1 (SP1), activating enhancer binding Protein 2 alpha (AP2-alpha), protein c-ets-1 (C-Ets-1), nuclear factor 1 (NF-1) and Represso. The recognition site for Sp1 is 5'-GGGCGG-3' (Giallongo *et al*., 1990); AP2-alpha is 5‘-GCCNNNGGC-3‘ (Williams and Tjian, 1991); NF-1 5'-TTGGCNNNNNGCCAA-3' (De vries *et al*., 1987), while C-Ets-1 has a 5‘-GGAA/T-3‘ core element (Wotton *et al*., 1994). Figure 3.3 maps these sites along the *SLC6A4* promoter region, showing the overlap between these binding sites and CpG sites.

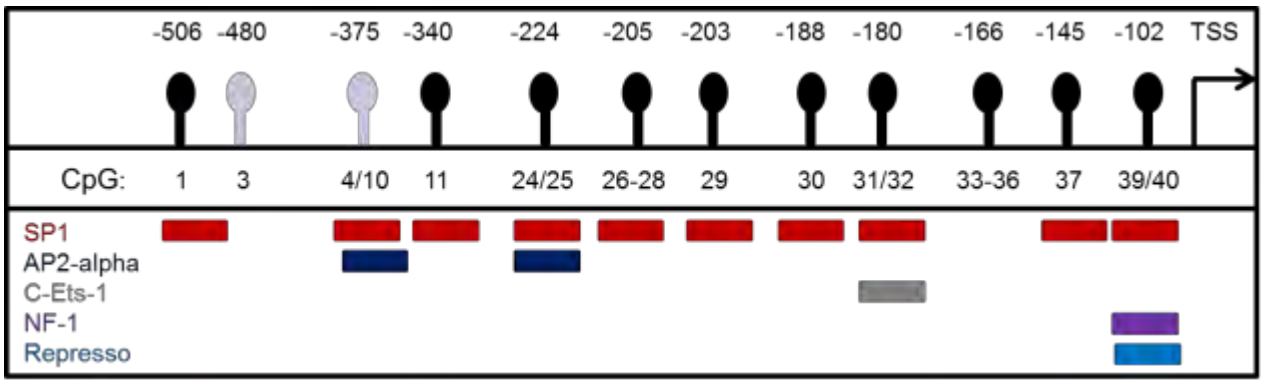

Figure 3.3.Representative image of the target region with the position of each CpG site analysed using the EpiTYPER assay. The Transcription Start Site (TSS) of the *SLC6A4* gene is included, along with any transcription factor binding sites which may overlap the CpG sites. The five transcription factors that bind to these putative sites are Sp1, AP2-alpha, C-Ets-1, NF-1 and Represso. Only CpG 3 and CpG 37 do not overlap with a putative binding site.

DNA sequences were produced for the bisulphite converted DNA of all 31 participants. The chromatograms of the converted sequences were analysed using Chromas 2.4. Visual inspection of the chromatograms allows for the analysis of individual peaks at each nucleotide position of the region investigated. Any remaining methylated cytosine residues would be observed as cytosine peaks, while the unmethylated cytosines would now be thymine peaks. DNA methylation was not evident from the sequence results; there were no clear cytosine peaks observed. On closer inspection, only one site (CpG 4) showed a small peak representative of a conserved cytosine residue. However, as seen in the chromatogram (Appendix Figure B), any background noise renders an unequivocal identification of DNA methylation presence unlikely. The results from bisulphite sequencing suggest very low levels of DNA methylation for the SLC6A4 promoter region; therefore a more sensitive assay of methylation, the EpiTYPER, was used.

Using the EpiTYPER assay, DNA methylation levels of the *SLC6A4* promoter were obtained for all 25 of the ASD participants and six of the controls. Twenty one of the potential 40 CpG sites in the region investigated were covered by the EpiTYPER assay. The final results include data from 12 CpG regions covering these 21 CpG sites, illustrated in Figure 3.4. Certain CpG sites that are closely situated were grouped together, depending on the RNase A cleavage. CpG sites 4 and 10 were

manually combined for data analysis, because identical levels of DNA methylation for these sites were observed for each participant.

SLC6A4 Promoter DNA Methylation in the ASD and control cohorts

The overall DNA methylation level for the entire region investigated was calculated using the individual CpG site DNA methylation levels. Both the average individual CpG site DNA methylation and the average DNA methylation across all CpG sites are shown in Figure 3.4. The technical controls of 0%, 25%, 75% and 100% methylated DNA are shown in Appendix Figure C.

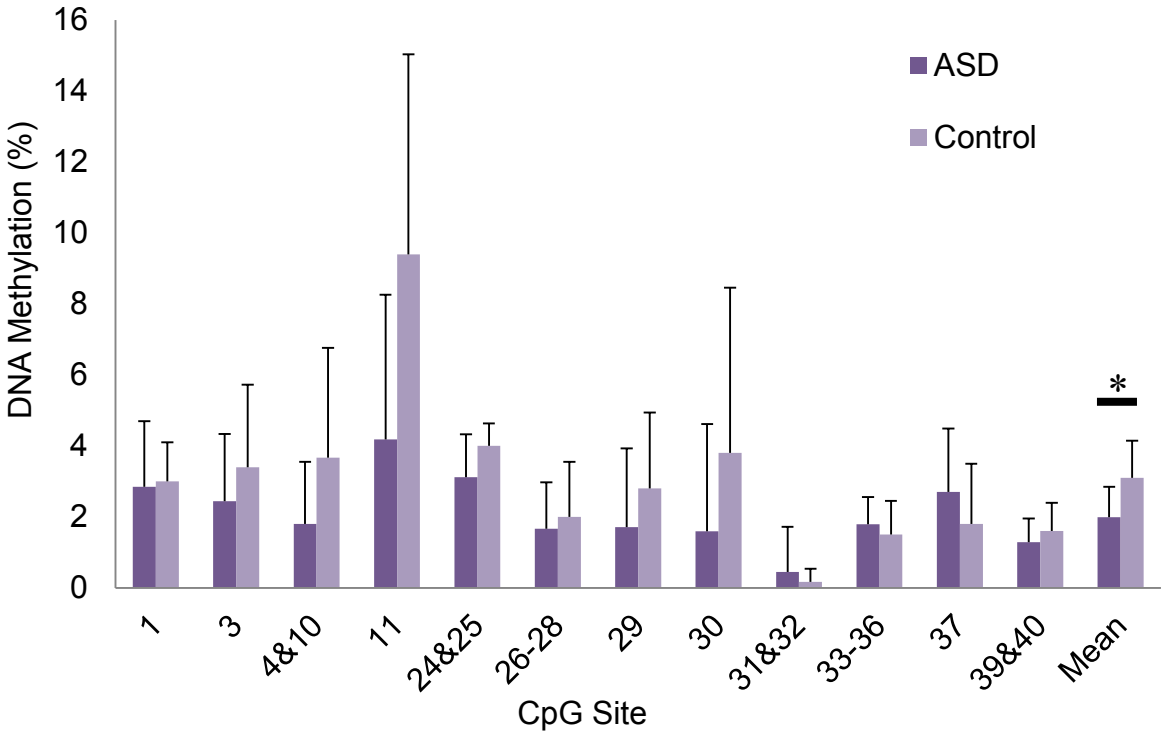

Figure 3.4. Average DNA methylation of the *SLC6A4* promoter for each target CpG site and DNA methylation across the 12 CpG regions in both the ASD (n=25) and control cohorts (n=6). The DNA methylation levels of the different cohorts were compared using the Mann-U Whitney test. Significance is denoted by * ($p<0.05$).

The target promoter region of the *SLC6A4* gene is located between 102bp and 506bp upstream of the gene (Figure 3.3) and showed low levels of DNA methylation. The methylation levels range from 0% to a maximum of 15% (CpG 11) for a control participant and 14% (CpG 30) for a participant with ASD. Methylation levels varied among sites and among individuals. Average DNA methylation across all 12 CpG

regions (%±SD) was significantly higher (p=0.011) in the control cohort (3.10±1.15%) compared to the ASD cohort (1.99±0.88%), as shown in Figure 3.4. However, differences in the levels of individual CpG site DNA methylation between the ASD and control cohorts were not significant.

Closer examination of the DNA methylation levels revealed variability in percentage across all sites, ranging from 0.39±1.16% (CpG 31/32) to 5.34±5.10% (CpG 11). However, the EpiTYPER assay results of 14.17% for the 25% technical control, 53.58% for the 75% technical control and 91.5% for the 100% technical control report an underrepresentation of the DNA methylation percentages.

SLC6A4 Promoter DNA Methylation in the ASD endophenotypes

The DNA methylation levels for these CpG sites were then compared across the ASD endophenotypes and the significant differences for both average and individual CpG site DNA methylation are shown in the figures below. First, the ADOS-2 comparison score-generated endophenotypes of ASD-high and ASD-moderate were compared to the control cohort. Average DNA methylation of the control cohort (3.10±1.15%) was significantly higher (p=0.004) than that observed for the ASD-moderate endophenotype (1.87±0.77%), as seen in Figure 3.5. However, the average DNA methylation of the control cohort was not significantly higher than that observed for the ASD-high endophenotype (2.50±1.19%, n=5, p=0.457). The other non-significant comparisons, including the ASD-high versus control comparison, are reported in Appendix Table D.

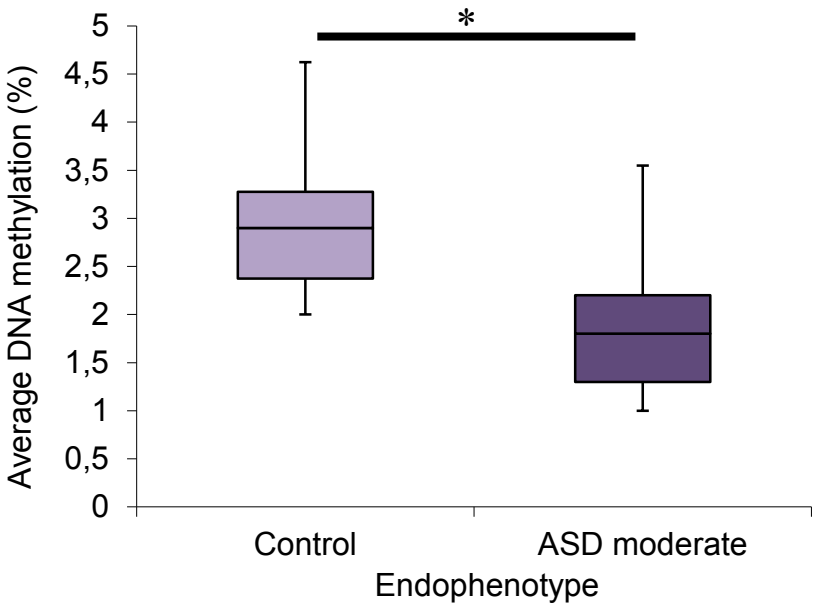

Figure 3.5. Average DNA methylation across all 12 target CpG regions of the *SLC6A4* promoter for the control cohort (n=6) and the ASD-moderate endophenotype (n=20). The DNA methylation level of the ASD-moderate endophenotype was compared to the control cohort using the Mann-U Whitney test. Significance is denoted by * ($p<0.05$).

When using language impairment as a way to partition the ASD cohort and then compare to the control, differences in DNA methylation levels were observed (Figure 3.6). Significance was observed across the CpG sites, but not at specific CpG sites (Appendix Table D). The control cohort displayed significantly higher average DNA methylation across the CpG sites (3.10±1.15%) compared to the participants who are pre-verbal (1.81±0.51%, $p=0.004$), used phrase speech only (1.83±1.04%, $p=0.026$) and the combined pre-verbal/phrase speech endophenotype (1.81±0.72%, $p=0.003$). No significant difference was observed between the participants who are pre-verbal and those with phrase speech only ($p=0.557$; Appendix Table D).

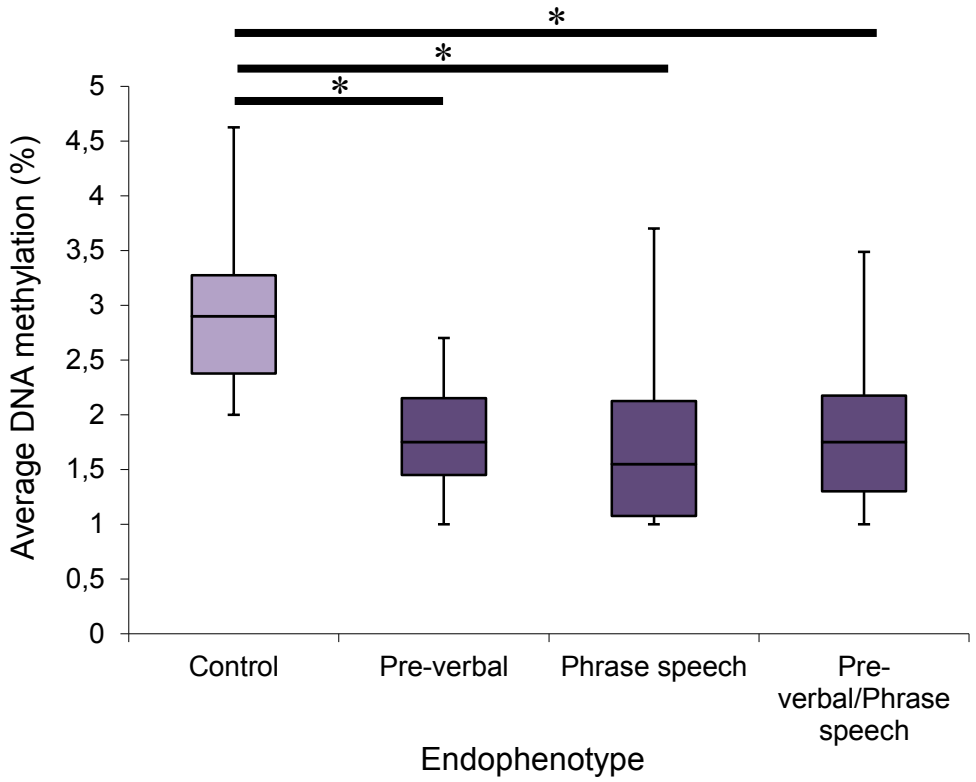

Figure 3.6. Average DNA methylation across all 12 target CpG regions of the *SLC6A4* promoter for the control cohort (n=6), the pre-verbal endophenotype (n=14), phrase speech endophenotype (n=8) and the combined pre-verbal/phrase speech endophenotype (n=22). The DNA methylation level of the three endophenotypes was compared to the control cohort using the Mann-U Whitney test. Significance is denoted by * ($p<0.05$).

The comparisons of DNA methylation based on ASD-trait severity, specifically for SA and RRB against the control cohort also yielded significant results. These data are presented in Figure 3.7. Both endophenotypes displayed significant differential DNA methylation across the CpG sites. Figure 3.7 shows that participants with more severe SA symptoms showed significantly decreased DNA methylation levels ($p=0.006$) across the CpG sites (1.90±0.88%) compared to the average DNA methylation level of the control cohort (3.10±1.15%). The control cohort also displayed significantly increased DNA methylation ($p=0.045$) compared to the high RRB endophenotype (1.95±0.58%). The remaining comparison results for individual CpG sites for the SA and RRB endophenotypes are shown in Appendix Table D.

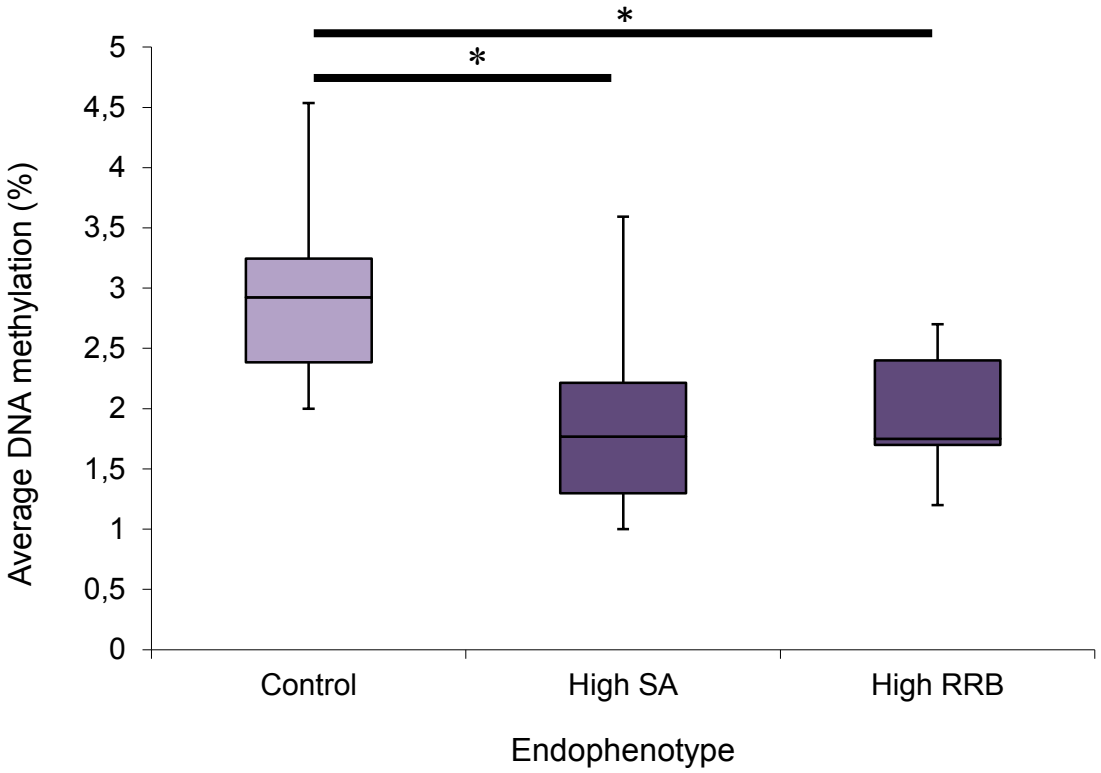

Figure 3.7. Average DNA methylation across all 12 target CpG regions of the *SLC6A4* promoter for the high SA endophenotype (n=23), the high RRB endophenotype (n=6) and the control cohort (n=6). The DNA methylation levels of the two endophenotypes were compared to the control cohort using the Mann-U Whitney test. Significance is denoted by * ($p<0.05$).

When compared to the control cohort, differential DNA methylation was also observed at the CpG 1 site in the high RRB endophenotype. At this site of the *SLC6A4* promoter region, significantly increased methylation ($p=0.045$) was observed in the control cohort (3±1.22%) compared to the high RRB endophenotype (1.67±1.15%; Figure 3.8).

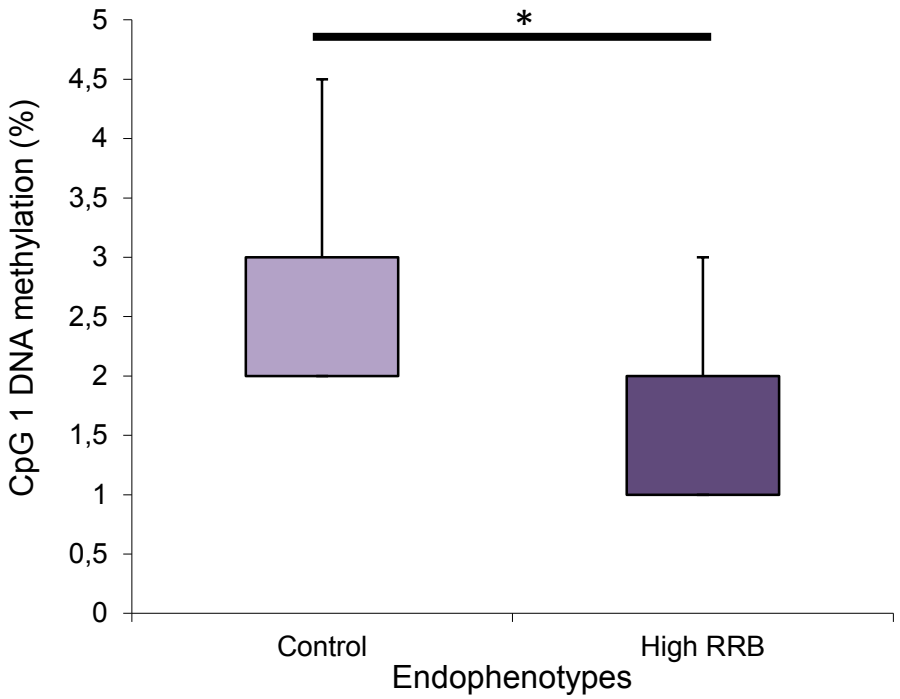

Figure 3.8. Average DNA methylation for the CpG 1 site of the *SLC6A4* promoter for the control cohort (n=6) and High RRB endophenotype (n=6). DNA methylation levels of the two groups were compared using the Mann-U Whitney test. Significance is denoted by * ($p<0.05$).

However, when the high RRB endophenotype was compared to the low RRB endophenotype, average DNA methylation levels at the CpG 30 site were found to be significantly increased ($p=0.001$) in the more affected (high) RRB endophenotype (3.00±1.22%) compared to the less affected RRB endophenotype (1.17±3.37%), observed in Figure 3.9. This is the only significant result of an endophenotype or subgroup of ASD that shows increased methylation with an increase in severity of symptoms or ASD. The control cohort and therefore less severe ASD traits have been associated with higher DNA methylation levels in all previous analyses.

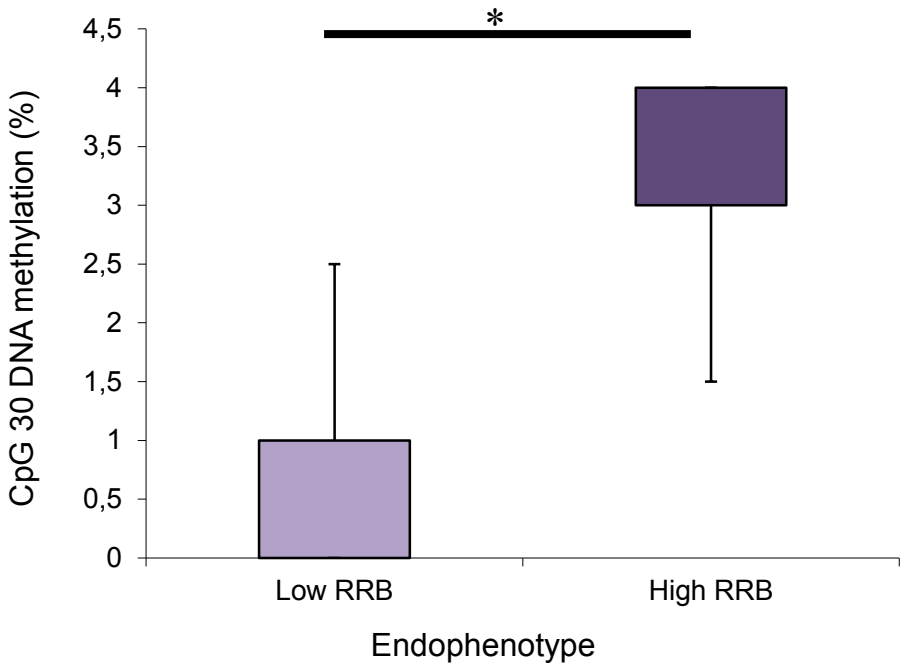

Figure 3.9. Average DNA methylation for CpG 30 of the *SLC6A4* promoter for the high RRB endophenotype (n=6) and the low RRB endophenotype (n=25). The DNA methylation level of the two endophenotypes was compared using the Mann-U Whitney test. Significance is denoted by * ($p<0.05$).

In an attempt to identify any association between the levels of DNA methylation at the *SLC6A4* promoter and the 5-HTTLPR, the methylation levels of the 5-HTTLPR groupings were compared. No significant average DNA methylation differences across all 12 CpG regions were observed between the LL participants and the LS/SS participants ($p=0.445$) or between the high and low *SLC6A4* expressers ($p=0.557$).

DNA methylation at the *SLC6A4* promoter was also compared across the different ancestry groups and according to age to account for any ancestry-or age-dependent changes to the DNA methylation. Average DNA methylation across all 12 CpG regions was not found to be significantly different in the two ancestry groups ($p=0.080$, see Appendix Table D). Only CpG24/25 showed significantly increased methylation ($p=0.008$), in participants of European ancestry (4.00±0.91%) compared to those of African ancestry (2.50±0.97%, see Appendix Figure D). No significant age-differences in average DNA methylation across all12 CpG regions ($p=0.358$) and for individual sites was observed (Appendix Table D).

CHAPTER 4
Discussion

Serotonergic signalling and serotonin levels have long been associated with the aetiology of ASD (McDougle *et al*., 1996; Makkonen *et al*., 2008; Murphy *et al*., 2006; Huang *et al*., 2008; Guhathakurta *et al*., 2009; Hranilovic *et al*., 2010; Kistner-Griffin *et al*., 2011). The most promising, yet ambiguous evidence, of this association is the presence of hyperserotonemia in up to 35% of ASD cases (Hranilovic *et al*., 2007). One of the potential disruptions to serotonergic signalling involves altered regulation of the serotonin transporter and therefore impaired reuptake of serotonin (Heils *et al*., 1996; Lesch *et al*., 1996). Changes in regulation of this transporter often attributed to the well-studied repeat polymorphism (5-HTTLPR) have proven to be inconclusive (Canli and Lesch, 2007; Lesch *et al*., 1996; Kistner-Griffin *et al*., 2011; Arieff *et al*., 2010; Cho *et al*., 2007; Yirmiya *et al*., 2001; Ma *et al*., 2010; Persico *et al*., 2002; Ramoz *et al*., 2006). However, recent evidence has emerged suggesting a stronger role of environmental factors working via epigenetic mechanisms, such as DNA methylation, in ASD aetiology (Wong *et al*., 2014; Melnyk *et al*., 2012; Nguyen et al, 2010). Furthermore, literature suggests that a potential interplay between the repeat polymorphism and DNA methylation at the *SLC6A4* promoter may be involved in conferring susceptibility to ASD (Philibert *et al*., 2007; van Ijzendoorn *et al*., 2010; Vijayendran *et al*., 2012). Both the repeat polymorphism and *SLC6A4* promoter DNA methylation were investigated in this study.

This study tested the hypothesis that the regulation of the serotonin transporter is different between an ASD and a control cohort, as well as between certain endophenotypes of ASD. Both regulatory mechanisms, the 5-HTTLPR and DNA methylation data presented here, support the involvement of serotonergic signalling within ASD aetiology. This is true for both ASD as a whole, as well as for certain traits associated with ASD, such as language impairments, SA and RRB. Furthermore, the importance of serotonergic signalling pathways in ASD is supported by significant differences in the 5-HTTLPR genotypic variant distributions within the study ASD cohort compared to the control cohort. Significant differential

DNA methylation patterns are also observed in the same ASD cohort and across ASD endophenotypes.

4.1. Reducing phenotypic heterogeneity

In order to effectively reduce the phenotypic heterogeneity that confounds many studies on ASD, smaller, homogenous endophenotypes were created. All participants, including the control cohort, underwent an ADOS-2 assessment to be placed into particular ASD endophenotypes. The endophenotypes created from the ADOS-2 assessments included varying ASD trait groups. This includes subgroups with varying language ability and subgroups based on severity of SA behaviours and the severity of RRB symptoms (Table 3.1). These endophenotypes are representative of the core domain limitations that characterise ASD in both the previous DSM-IV and current, revised DSM-V (APA, 2013).

To empirically explore a number of traits or categories of ASD, nine endophenotypes were created. The smallest grouping classified according to the ADOS-2 comparison scores was the ASD-high subgroup. This subgroup included five participants. However, this represents a clean, well-defined subgroup of ASD and therefore was still used in comparisons. The ASD-high subgroup refers to individuals with severe core ASD domain impairments compared to other children with ASD of similar developmental ages, and therefore could potentially show the biggest difference in regulation of the serotonin transporter gene compared to a control cohort.

Additionally, it is important to note that the entire ASD cohort is also representative of a clean ASD phenotype because of the strict exclusion criteria. Indeed, a number of potential participants with ASD were excluded from the study through a number of criteria. These criteria included another diagnosis, confirmed or suspected, of other chromosomal disorders, if the participant already had a sibling in this study, to exclude familial genetic associations, or if participants did not score on the ADOS-2 in spite of an ASD diagnosis. Noteworthy, any control participants were excluded if they scored on ADOS-2 or displayed severe speech delay. The latter highlights the value of excluding control participants who may have —ASD-like" traits and to ensure that all the behaviours used to characterise ASD are indeed qualitative traits. The exclusion criteria used in this study led to the creation of the clean endophenotypes

with reduced phenotypic heterogeneity. The molecular regulation of the serotonin transporter was then compared across these more homogenous endophenotypes. However, any significant change in *SLC6A4* regulation that may result from ancestry or age was first determined.

Notably, the 5-HTTLPR genotypic and allelic frequency distributions or DNA methylation levels did not significantly change across the ancestry groups or according to age (See Appendix Tables C and D); with the exception of the CpG 24/25 site (Appendix Figure D). This supported the comparisons of the control cohort, comprised of participants from European ancestry only. However, due to sample size constraints of the cohort, the effects of ancestry could not be rigorously tested. The clear limitation of sample size also partially extends to the endophenotypes. It is compounded by the fact that only a fraction of the starting sample cohort was used in the final analysis of this study. This is the result of the strict exclusion criteria and ADOS-2 thresholds required for the building of a clean ASD and control phenotype, as well as clean endophenotypes. The cohort building is a time consuming, arduous task, especially so in a South African context, where no established ASD databank exists. This is in contrast for other studies where, for example in the United States, there is access to AGRE (Autism Genetic Research Exchange, USA) or Simons Simplex Collection (SSC), a resource of the Simons Foundation Autism Research Initiative (SFARI). However, even with a relatively small sample size, significant data for both the 5-HTTLPR and *SLC6A4* DNA methylation were obtained. This indicates the potential utility of this region for further analyses with even larger sample sizes to validate these findings, as well as to identify further significant links to endophenotypes.

4.2. Repeat polymorphism in the ASD and control cohorts

Due to the long-standing association of ASD and ASD-traits with the serotonin transporter repeat polymorphism (Canli and Lesch, 2007; Lesch *et al*., 1996; Lin, 2007; McDougle *et al*., 1998; Sutcliffe *et al*., 2005; Kistner-Griffin *et al*., 2011; Arieff *et al*., 2010; Cho *et al*., 2007; Yirmiya *et al*., 2001), it was hypothesised in this study that there would be differential genotypic and allelic frequency distributions of the 5-HTTLPR in the ASD cohort, as well as across ASD-trait endophenotypes. This hypothesis is supported by the 5-HTTLPR results. Analysis of the repeat

polymorphism revealed that the presence of the L allele in a population significantly reduces the risk of ASD (RR=0.78; 95%, confidence interval: 0.622-0.997). In addition, a significantly increased presence of S allele-containing genotypes was observed in the ASD cohort compared to the control (Figure 3.1A, p=0.049). This enrichment of the S allele within an ASD population supports the study hypothesis of an increase in the S allele in the ASD cohort. This suggests that impaired serotonin transporter function may be implicated in ASD aetiology. Furthermore, this is supported by previous literature implicating the short allele in ASD development (Kistner-Griffin *et al*., 2011). These authors reported over-transmission of the S allele of the 5-HTTLPR in an AGRE sample cohort. The S-allele containing genotypic enrichment in this South African ASD population could potentially result in lowered *SLC6A4* mRNA transcription and therefore disruption to the serotonin transporter function and serotonergic system.

The 5-HTTLPR L allele is generally associated with higher levels of serotonin transporter mRNA transcription, as well as an increased level of serotonin reuptake (Heils *et al*., 1996; Lesch *et al*., 1996). Therefore, participants homozygous for the S allele would potentially have reduced reuptake of serotonin. This could result in the hyperserotonemia observed at higher frequencies in ASD individuals. The LS heterozygous genotype acts similarly to the SS homozygous genotype, because of the dominant effect of the S allele of the 5-HTTLPR (Hu *et al*., 2006). Therefore, the significant increase in LS/SS genotypes in the ASD cohort may contribute to ASD aetiology by impacting upon the serotonin reuptake mechanism. However, when the effect of the rs25531 SNP of the L allele is incorporated into the genotypic frequency calculations, the significance of this association is lost (Figure 3.2, p=0.3). The genotypic frequency distribution changes as the L_G allele would now be grouped similarly to the S allele. The loss of significance may be due to the presence of the G variant in only two LS heterozygous ASD participants. This L_GS genotype would remain in the "low expression" group and did therefore not affect the genotypic frequency calculations of the ASD cohort. The genotypic frequency calculations are only changed with the presence of the G variant on the L allele of LL homozygous participants, as observed in one control participant. Therefore, the control population were more likely to be affected, because of the enrichment of the LL genotype within

this cohort. However, the genotypic frequencies vary across other populations as well (Esau *et al*., 2008; and references within).

Due to the variation across populations, the data in this study were compared to previous results obtained within a South African population. One research group calculated the allelic and genotypic frequencies within a South African control population (Esau *et al*., 2008) and a South African ASD population (Arieff *et al*., 2010). The populations in these two studies overlapped the same Western Cape population pool as used in this study. ASD participants were recruited from the same local Cape Town government school for learners with Autism. However, both Esau *et al*. (2008) and Arieff *et al*. (2010) did not utilize ADOS-2 or any other ASD assessment to characterize their cohorts, which is the power of this current study.

With regard to Esau *et al*. (2008), comparable genotypic frequencies of their total control population (all ancestry groups) were reported compared to the control cohort of this study. The frequencies of 61.40% for the LL homozygous genotype and 38.60% for the LS/SS genotypes of the 5-HTTLPR only differ by 5.30% to those observed in this study (66.70% for LL and 33.30% for LS/SS). Therefore, this demonstrates that even with a small sample size of six participants without ASD, the control cohort in this study is representative of another control population within a South African context.

In addition, the result of this study is comparable to the data on the frequency of the 5-HTTLPR genotypes in another South African ASD population (Arieff *et al*., 2010). Their reported 52% frequency of homozygous LL individuals is 2.5-fold greater than that observed in this study (20.80%), corresponding to a 1.65-fold decrease in S-allele containing genotypes (48.00% in Arieff *et al*. 2010 and 79.20% in this study). Although Arieff *et al*. (2010) reported an increased presence of the S allele in the ASD population compared to a control population; this difference is more pronounced within the ASD cohort in this study. This may be due to the use of ADOS-2 and strict exclusion criteria leading to a cleaner ASD phenotype; one that is not affected by any genetic alterations which may account for any co-morbidities.

Repeat polymorphism in the ASD endophenotypes

The ASD endophenotypic comparisons for the 5-HTTLPR component of this study did not yield significant results. An increase in the S allele-containing genotypes is observed in the ASD-high endophenotype (n=5) compared to the control cohort, but this is only approaching significance (0.061). This may be a true association though, and should be repeated with an increased sample size for the ASD-high endophenotype. The lack of significance indicates that the 5-HTTLPR may play a role in ASD aetiology as a whole, but cannot be linked to conferring a specific trait of ASD or severity of this trait. Perhaps, with increased samples sizes of the endophenotypes, these comparisons should be repeated to determine whether the lack of significant association is continued. It is also possible that this locus of the serotonin transporter gene is not responsible for conferring susceptibility to specific traits of ASD, which may be more influenced by differential DNA methylation of the *SLC6A4* promoter.

4.3. SLC6A4 promoter DNA hypo-methylation

Interestingly, investigation of the DNA methylation status of the SLC6A4 promoter yielded the more supportive evidence for an ASD aetiology link to serotonin transporter regulation. DNA methylation can be determined through a variety of different methods. In this study, the DNA methylation status of *SLC6A4* promoter region, 102bp and 506bp upstream of the gene (Figure 3.3) was determined using bisulphite DNA sequencing and the EpiTYPER MassARRAY assay. A native sequence of this region was first analysed to confirm that no polymorphisms would interfere with DNA methylation status. Only one SNP, rs25533, with a *T-C* change in nucleotide, was identified in one control participant and one participant with ASD. Interestingly, the SNP identified in this study has been linked to allelic heterogeneity at *SLC6A4* to autism susceptibility and rigid-compulsive behaviours (Sutcliffe *et al*., 2005). However, according to the ADOS-2 results, the control participant heterozygous for this SNP does not present with any social communication or reciprocal social interaction impairments or repetitive and restricted behaviour. Therefore, the results of this study do not support an association of the SNP with any particular ASD trait. Notably, this SNP did not overlap with any CpG sites and therefore did not affect the DNA methylation analysis of the *SLC6A4* promoter region.

DNA methylation analysis was carried out using two different methodologies, bisulphite sequencing and EpiTYPER massARRAY assay analysis. Using the EpiTYPER assay, the *SLC6A4* promoter region (Figure 3.3) was determined to be hypo-methylated (Figure 3.4). DNA methylation was observed to range from 0% to 15%, depending on the CpG site. This observed hypo-methylation may be responsible for the lack of evidence for DNA methylation from the qualitative method of bisulphite sequencing. On close inspection, the chromatogram in Appendix Figure B shows a small cytosine peak at CpG 4. This may be indicative of the expected low level of DNA methylation (2.23±2.32% from the EpiTYPER assay), but is not unequivocally identified because of the background noise observed in the chromatogram. This background noise includes small cytosine peaks that may be mistaken for the presence of DNA methylation.

DNA methylation is generally observed at 60-90% of CpG sites in most mammals (Hendrich and Tweedie, 2003; Kinde *et al*., 2015). However, because CpG islands are specifically associated with hypo-methylation (Larson *et al*., 1992), the hypo-methylation level of the *SLC6A4* promoter region is not unexpected. The low levels of DNA methylation can also be explained by the presence of a number of putative Sp1 binding sites across the entire region investigated, with the exception of two CpG sites (Figure 3.3). Proteins bound to these sites can hinder access of DNA methyltransferases to the CpG sites. Removal of Sp1 binding sites which flank a CpG island can lead to *de novo* methylation during development (Jones and Takai, 2001, and references within). This suggests that Sp1 transcription factors may impede DNA methyltransferases in gaining access to specific CpG sites. Indeed, this may also be true for CpG sites 3 and 37, which are not covered by Sp1 sites, as flanking Sp1 sites can also impede DNA methyltransferase action.

In addition, the lower end (0%) of DNA methylation levels observed at the *SLC6A4* promoter could be attributed to an underrepresentation of true DNA methylation. The technical controls of 25% and 75% methylated DNA in the EpiTYPER assay were approximately 10% below the expected percentages (see Appendix Figure C). However, this 10% decrease likely results from an error in the percentage of methylated DNA in the technical controls, given the sensitivity of the EpiTYPER

assay, and the expected low DNA methylation levels. Zhao *et al*. (2013) also observed low levels of DNA methylation at *SLC6A4*, ranging from 3% to 16.7%, consistent with the methylation data collected in this study. Given the expected 60-90% methylation of CpG sites (Hendrich and Tweedie, 2003; Kinde *et al*., 2015), the hypo-methylation status of the serotonin transporter promoter more than likely has a functional role. Therefore, any changes in the mechanisms controlling this hypo-methylation and leading to alterations of the methylation levels may play a role in altering the serotonin transporter function. Only small differences are observed between the ASD cohort and control cohort, as well as across the endophenotypes, but these small differences may still play significant roles. Zhao *et al*. (2013) found significant phenotypic changes in obesity with only a one percentage change in DNA methylation at the *SLC6A4* promoter. Therefore, a one percentage change in DNA methylation, as observed in this study, may indeed lead to significant phenotypic changes.

Unfortunately, only 12 CpG regions and 21 out of 40 CpG sites in the region were investigated by the EpiTYPER assay (Figure 3.3), given the technical limitations of the assay. DNA methylation at CpG sites, 3 to 10, has previously been linked to *SLC6A4* mRNA expression levels (Philibert *et al*., 2007). Within this region, only CpG 3, CpG 4 and CpG 10 are covered by the assay. These missing CpG sites may well contribute to ASD aetiology. When combined with the data from this study, they may add to the significance of the differential DNA methylation between ASD and control cohorts analysed in future studies. The CpG rich status of the *SLC6A4* promoter region investigated resulted in many fragments of equal mass, as well as an inadequate number of thymine residues for appropriate cleavage in the EpiTYPER assay. Therefore, perhaps a different technique, such as the MiSeq Sequencing System on an Illumina platform may prove to be more ideal. This would allow for a quantitative calculation of every CpG site, while concurrently allowing for SNP analysis.

Differential SLC6A4 promoter DNA methylation in the ASD cohort

One of the hypotheses of this study states that differential DNA methylation at the *SLC6A4* promoter is expected between the ASD and control cohorts. This hypothesis is supported by the methylation data obtained in this study. The ASD

cohort presents with differential DNA methylation, with significantly lower DNA methylation levels across all 12 CpG regions over the *SLC6A4* promoter compared to the control cohort (Figure 3.4, p=0.011). In spite of significant differences in mean methylation levels, significant differences were not largely observed at individual CpG sites. This is partly explained by the large variation in DNA methylation percentage observed at each site (Figure 3.4). In summary, DNA methylation was higher at nine of the twelve CpG sites in the control cohort compared to the ASD cohort. Additionally, the CpG 31/32 sites were almost completely non-methylated for both cohorts (0.39±1.17%). This may result from the presence of overlapping and flanking Sp1 binding sites, as well as the overlapping binding site of C-Ets-1 (Figure 3.3). The transcription factor, C-Ets-1, is expressed in neural crest cells and is implicated in embryonic development. This is completed via RNA polymerase II transcription factor activity, sequence-specific DNA binding and histone acetyltransferase binding. The C-Ets-1 protein also interacts with other transcription factors and forms multi-protein complexes. It is thus implicated in typical functioning of the serotonin transporter by impeding the action of DNA methyltransferases from accessing this CpG site (Furlan and Pourtier, 2015; Li *et al*., 2000).

Differential SLC6A4 promoter DNA methylation in the ASD endophenotypes

This study also hypothesised that there would be differential DNA methylation at the *SLC6A4* promoter between the ASD endophenotype comparisons and control cohort. This is indeed observed. Compared to the control cohort or less affected ASD-trait endophenotypes, significantly reduced *SLC6A4* promoter DNA methylation was observed in the ASD-moderate endophenotype, the more severely impaired language endophenotypes (pre-verbal and phrase speech only) and the more severely affected SA and RRB endophenotypes (Figures 3.5-3.8). This corresponds to the significantly reduced DNA methylation levels within the ASD cohort.

On the contrary, the only ASD endophenotype characterised by increased *SLC6A4* DNA methylation with increased severity of symptoms is the repetitive and restricted behaviour endophenotype (Figure 3.9) at the CpG 30 site. The repetitive and restricted behaviour (RRB) endophenotype was of particular interest due to the already identified link between *SLC6A4* and rigid and compulsive behaviours (Lin, 2007; McDougle *et al*., 1998; Sutcliffe *et al*., 2005). Noteworthy, the CpG 30 site also

overlaps a Sp1 binding site and therefore increased DNA methylation at this site may hinder the binding of Sp1. This may then, in turn, hinder transcription and therefore may impact upon *SLC6A4* mRNA expression. The repetitive and restricted behaviour endophenotype also encompasses a wide range of behaviours, including stereotypic speech and restricted interests to unusual sensory seeking behaviours. Some of these behaviours are seen as coping mechanisms employed by a child with ASD in an uncomfortable environment as a calming influence. These repetitive behaviours are usually motivated by sensory seeking which can provide a child with a calm state (Joosten *et al*., 2009). Therefore, these behaviours may be a side effect of ASD and do not necessarily result from any change in molecular regulation. However, the association of significantly increased DNA methylation at the lone CpG 30 site with an increase in severity of repetitive and restricted behaviours implicates this site in the development of these behaviours.

To date, there are very few studies that have looked at DNA methylation at the *SLC6A4* promoter and therefore predicted differences in DNA methylation status are unexplored. Park *et al*. (2015) reported elevated DNA methylation at the *SLC6A4* promoter with increased ADHD symptoms, while van Miln *et al*. (2014) identified increased ADHD symptoms with lower DNA methylation levels. This is consistent with a hypothesis that overall dysregulation of SLC6A4 rather than increased or decreased function may result in increased severity of certain traits and symptoms associated with neuropsychiatric disorders.

Interestingly, specific CpG sites were associated with differential DNA methylation for certain endophenotypes. These include CpG 30 and CpG1 (both for RRB), while CpG 24/25 was found to be associated with ancestry groups. All of these sites overlap with Sp1 binding sites, while CpG24/25 is also located within a binding site for activating enhancer binding Protein 2 alpha (AP2-alpha). This protein is involved in transcription machinery recruitment (Williams and Tijan, 1991). It is expressed in the ectoderm, particularly the neural crest cell lineages and is involved in early cell differentiation and development, thus playing a role in neurodevelopment (Hilger-Eversheim *et al*., 2000). Differential DNA methylation at this site could impair the binding of AP2-alpha and therefore alter the expression of *SLC6A4* mRNA. However, CpG 24/25 is associated with differential DNA methylation within the

ancestry groups and not ASD. Even though only one CpG site (CpG 24/25) in this study exhibited ancestry-related differential methylation, it is still imperative to establish cohorts with equal distribution of the different ancestry groups. This will reduce the impact that molecular alterations based on ancestry may have on data acquired in further studies.

Additionally, there appears to be a robust association between lower DNA methylation levels and impaired language (Figure 3.4). This may result from differential regulation of the serotonin transporter and therefore atypical levels of serotonin in certain regions of the brain. For instance, Johnson *et al.* (2016) found an association between language problems in preschool-aged children and prenatal exposure to serotonin reuptake inhibitors (SRIs). This implicates typical functioning of the serotonin transporter in language development. Furthermore, inhibited levels of serotonin in serotonergic neuron rich brain regions in mice have been associated with short term memory loss (Banik and Lahiri, 2004). Also, impaired or a loss of short term memory has long been associated with limited language. Literature shows that children with language impairments do significantly worse on short term and working memory tests (Conti-Ramsden *et al*., 2015 and references within). Therefore, this implicates atypical levels of serotonin in the development of language impairments, which may stem from affected short term memory. However, another drawback of this study is the lack of a specific language test to categorize the language endophenotypes. The use of the modules is sufficient as an investigatory approach, but further comparison statistics need to be completed on language endophenotypes reclassified according to specific language testing.

4.4.Potential model of 5-HTTLPR and SLC6A4 promoter DNA methylation interplay

When attempting to reconcile the 5-HTTLPR and the DNA methylation data for the *SLC6A4*, it appears to be contradictory. Although no significant association was found between the 5-HTTLPR status and DNA methylation levels, there may still be a link between the lowered DNA methylation levels and the S-allele containing genotypes. These S allele-containing genotypes are significantly enriched in the ASD cohort (Figure 3.1A, p=0.049), which also has significantly lowered DNA methylation levels compared to the control cohort (Figure 3.4, p=0.011). The S allele

of the 5-HTTLPR is known to lower SLC6A4 mRNA expression, while lower DNA methylation of the serotonin transporter promoter leads to an increase in *SLC6A4* mRNA expression. Therefore, the decreased DNA methylation may be a ―corrective" strategy during development in an attempt to regain homeostasis and remedy the impaired function of *SLC6A4*. The same idea can be applied for the control cohort. The higher levels of DNA methylation within the control cohort could be a result of the enrichment of the L allele, in order to achieve appropriate levels of *SLC6A4* mRNA expression. This potential ―homeostasis hypothesis" of how both the 5-HTTLPR and DNA methylation of the promoter region of *SLC6A4* had a combined effect on serotonin mRNA expression is illustrated in Figure 4.1.

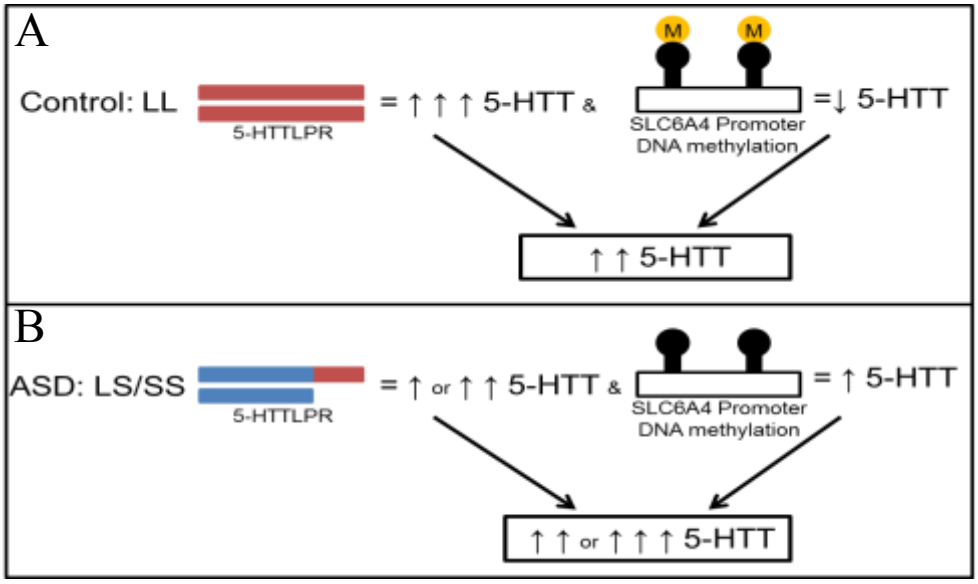

Figure 4.1. Diagram illustrating potential combined effects of the 5-HTTLPR and DNA methylation of the *SLC6A4* promoter on serotonin transporter mRNA expression. Panel A shows the potential action in the control cohort, with enriched presence of the LL genotype and generally higher methylation, leading to intermediate levels of 5-HTT. Panel B demonstrates the potential action in the ASD cohort with enriched presence of the S allele and generally lower methylation resulting in intermediate/higher levels of 5-HTT. Levels of expression are indicated using arrows (↑ is low or increased, ↑↑ is intermediate and ↑↑↑ is high, while ↓ is decreased).

However, changes in *SLC6A4* mRNA expression through differential methylation could also result from changes in the binding of the prominent methyl-DNA-binding protein, the methyl-CpG-binding protein 2 (MeCP2) to the *SLC6A4* promoter. MeCP2 binds to methylated cytosine residues in gene promoter regions and can then repress transcription via interactions with protein complexes (Lee *et al*., 2015).

Mutations in MeCP2 and subsequent altered gene expression have been linked to Rett syndrome, an X-linked neurological disorder, bipolar disorder, schizophrenia, as well as infantile ASD (Kinde *et al*., 2015; Lee *et al*., 2015). The maintenance of neuronal DNA methylation is essential for typical development of the mammalian brain and any molecular alterations of MeCP2 or its binding can lead to neurodevelopmental disorders, such as ASD (Kinde *et al*., 2015). Further analysis of this protein and where it binds could pinpoint or shed light on the mechanism responsible for differential DNA methylation at the *SLC6A4* promoter.

In conclusion, the data presented in this study suggest a role for atypical or altered serotonergic signalling within ASD aetiology. The data observed supports the hypothesis that there is differential molecular regulation of *SLC6A4* mRNA expression between an ASD cohort and a control cohort, which extends to ASD endophenotypes. This is found for both genetic regulation, as well as epigenetic regulation of the serotonin transporter. However, the stronger support is for the involvement of the epigenetic mechanism within the development of ASD traits. This mechanism would likely provide the link between environmental factors and molecular alterations involved in ASD aetiology.

Although the specific mechanisms of how DNA methylation or the repeat polymorphism impact directly on SLC6A4 mRNA expression are unknown, the combined effect of differential genetic and epigenetic regulation of the serotonin transporter may play a role in the aetiology of ASD. Differential regulation of the serotonin transporter within this ASD cohort compared to a control cohort is found. Therefore, an analysis of *SLC6A4* mRNA transcript levels within these cohorts should be completed. This will allow for the establishment of how the contradictory actions of the 5-HTTLPR and *SLC6A4* promoter DNA methylation within this ASD cohort affect *SLC6A4* mRNA expression compared to the control cohort.

The associations of differential regulation of the serotonin transporter to ASD and endophenotypes are an important starting point in determining to what an extent the development of ASD traits is linked to the 5-HTTLPR and *SLC6A4* promoter DNA methylation. These two molecular regulatory mechanisms are potential biomarkers that may eventually aid in early identification or intervention of ASD traits.

BIBLIOGRAPHY

Abdallah, M.W., Greaves-Lord, K., Grove, J., Nørgaard-Pedersen, B., Hougaard, D.M. & Mortensen. E.L. 2011. Psychiatric comorbidities in autism spectrum disorders: findings from a Danish Historic Birth Cohort. *European Child & Adolescent Psychiatry*. **20**:599–601.

Aljanabi, S.M. & Martinez, I. 1997. Universal & rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*. **25**(22):4692-4693.

American Psychiatric Association. 2013. DSM-V. Diagnostic & Statistical annual of Mental Disorders. 5th edition. Washington DC: APA. 2013 http://dx.doi.org/10.1176/appi.books.9780890425596.

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. & Hamosh, A. 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIMR), an online catalog of human genes & genetic disorders. *Nucleic Acids Research*. **43** Database issue.

Arieff, Z., Kaur, M., Gameeldien, H., van der Merwe, L. & Bajic, VB. 2010. 5-HTTLPR polymorphism: analysis in South African autistic individuals. *Human Biology*. **82**(3):291–300.

Banik, S. & Lahiri, T. 2005. Decrease in brain serotonin level & short term memory loss in mice: a preliminary study. Environmental Toxicology & Pharmacology 19:367–370.

Beach, S.R., Brody, G.H., Todorov, A.A., Gunter, T.D. & Philibert, R.A. 2010. Methylation at SLC6A4 is linked to family history of child abuse: an examination of the Iowa adoptee sample. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. **153**B:710-713.

Beach, S.R., Brody, G.H., Todorov, A.A., Gunter, T.D. & Philibert, R.A., 2011. Methylation at 5HTT mediates the impact of child sex abuse on women's antisocial behavior: an examination of the Iowa adoptee sample. *Psychosomatic Medicine*. **73**:83-87.

Betancur, C. 2011. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic & genomic disorders & still counting. *Brain Research.* **1380**:42–77.

Boeckers, T.M., Bockmann, J., Kreutz, M.R. & Gundelfinger, E.D. 2002. ProSAP/Shank proteins - a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *Journal of Neurochemistry.* **81**(5):903–910.

Bruder, C.E., Piotrowski, A., Gijsbers, A.A., &ersson, R., Erickson, S. & de Stahl, D.T. 2008. Phenotypically concordant & discordant monozygotic twins display different DNA copy-number-variation profiles. *American Journal of Human Genetics*. **82**(3):763–771.

Brugha, T.S., McManus, S., Bankart, J., Scott, F., Purdon, S., Smith, J., Bebbington, P., Jenkins, R. & Meltzer, H. 2011. Epidemiology of autism spectrum disorders in adults in the community in England. *Archives of General Psychiatry*. **68**(5):459-466.

Canli, T. & Lesch, K-P. 2007. Long Story Short: The Serotonin Transporter in Emotion Regulation & Social Cognition. *Nature Neuroscience*. **10**:1103 – 1109.

Caspi, A., Sugden, K., Moffitt, T.E., Taylor, A., Craig, I.W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A. & Poulton, R. 2003. Influence of life stress on depression: moderation by a polymorphism in the 5- HTT gene. *Science*. **301**:386-389.

CDC. 2012. Prevalence of autism spectrum disorders – autism & developmental disabilities monitoring network, 14 sites, United States, 2008. *MMWR Surveillance Summary*. **61**:1–19.

Cho, I.H., Yoo, H.J., Park, M., Lee, Y.S. & Kim, S.A. 2007. Family-based association study of5-HTTLPR & the 5-HT2A receptor gene polymorphisms with autism spectrum disorder in Korean trios. *Brain Research*. **1139**:34–41.

Cook, E.H.Jr., Arora, R.C., Anderson, G.M., Berry-Kravis, E.M., Yan, S.Y., Yeoh, H.C., Sklena, P.J., Charak, D.A. & Leventhal, B.L. 1993. Platelet serotonin studies in hyperserotonemic relatives of children with autistic disorder. *Life Science*. **52**:2005–2015.

de Leon-Guerrero, S.D., Pedraza-Alva, G. & Perez-Martınez, L. In sickness and in health: the role of methyl-CpG binding protein 2 in the central nervous system. 2011. *European Journal of Neuroscience*. **33**:1563–1574.

De Vries, E., van Driel, W., van den Heuvel, S.J. & van der Vliet, P.C. 1987. Contactpoint analysis of the HeLa nuclear factor I recognition site reveals symmetrical binding at one side of the DNA helix. *The EMBO Journal*. **6**(1):161–168.

Delbruck, S.J., Wendel, B., Grunewald, I., S&er, T., Morris-Rosendahl, D., Crocq, M.A., Berrettini, W.H. & Hoehe, M.R. 1997. A novel allelic variant of the human serotonin transporter gene regulatory polymorphism. *Cytogenetics & Cell Genetics.* **79**:214–220.

Dhingra, K.M. & Sarma, G.S. 2014. Analytical Techniques for DNA Methylation – An Overview. *Current Pharmaceutical Analysis*. **10**:71-85.

Dolen, G., Darvishzadeh, A., Huang, K.W. & Malenka, R.C. 2013. Social reward requires coordinated activity of nucleus accumbens oxytocin & serotonin. *Nature*. **501**(7466):179–184.

Dong, E., Dzitoyeva, S.G., Matrisciano, F., Tueting, P., Grayson, D.R. & Guidotti, A. 2015. Brain-Derived Neurotrophic Factor Epigenetic Modifications Associated with Schizophrenia-like Phenotype Induced by Prenatal Stress in Mice. *Biological Psychiatry.* **77**(6):589-596.

Dukal, H., Frank, J., Lang, M., Treutlein, J., Gilles, M., Wolf, I.A.C., Krumm, B., Massart, R., *et al*. 2015. New-born females show higher stress- & genotype-independent methylation of SLC6A4 than males. *Borderline Personality Disorder & Emotion Dysregulation*. **2**:8.

Elsabbagh, M., Divan, G., Koh, Y.J., Kim, Y.S., Kauchali, S., Marcin, C., Montiel-Nava, C., Patel, V., Paula, C.S., Wang, C., *et al*. 2012. Global prevalence of autism & other pervasive developmental disorders. *Autism Research*. **5**(3):160–179.

Esau, L., Kaur, M., Adonis, L. & Arieff, Z. 2008. The 5-HTTLPR Polymorphism in South African Healthy Populations: A Global Comparison. *Journal of Neural Transmission*. **115**:755-760.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. & Paul, C.L. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *PNAS*. **89**(5):1827-1831.

Furlan, A. & Pourtier, A. 2015. Ets-1 activation, when tumors crosstalk with their Microenvironment. *Cancer Cell & Microenvironment*. **2**:494.

Georgiades, S., Szatmari, P., Boyle, M., Hanna, S., Duku, E., Zwaigenbaum, L., Bryson, S., Fombonne, E., *et al*. 2013. Investigating phenotypic heterogeneity in children with autism spectrum disorder: a factor mixture modeling approach. *Journal of Child Psychology & Psychiatry*. **54**:206–215.

Geschwind, D.H. 2011. Genetics of autism spectrum disorders. *Trends in Cognitive Sciences*. **15**(9):409-416.

Giallongo, A., Oliva, D., Cali, L., Barba, G., Barbieri, G. & Feo, S. 1990. Structure of the human gene for α-enolase. *European Journal of Biochemistry*. **190**:567-573.

Giulian, D., Pohorecky, L.A. & McEwen, B.S. 1973. Effects of gonadal steroids upon brain 5-hydroxytryptamine levels in the neonatal rat. *Endocrinology*. **93**:1329–1335.

Gordon, C.T., State, R.C., Nelson, J.E., Hamburger, S.D. & Rapoport, J.L. 1993. A doubleblind comparison of clomipramine, desipramine, & placebo in the treatment of autistic disorder. *Archives of General Psychiatry*. **50**(6):441–447.

Grayson, D.R. & Guidotti, A. 2013. The dynamics of DNA methylation in schizophrenia and related psychiatric disorders. *Neuropsychopharmacology Reviews.* **38**:138–166.

Gropman, A.L. & Batshaw, M.L. 2010. Epigenetics, copy number variation & other molecular mechanisms underlying neurodevelopmental disabilities: new insights & diagnostic approaches. *Journal of Developmental & Behavioral Pediatrics.* **31**(7):582–591.

Guhathakurta, S., Singh, A.S., Sinha, S., Chatterjee, A., Ahmed, S., Ghosh, S. & Usha, R. 2009. Analysis of serotonin receptor 2A gene (HTR2A): association study with autism spectrum disorder in the Indian population & investigation of the gene expression in peripheral blood leukocytes. *Neurochemistry International*. **55**(8):754–759.

Hallmayer, J., Clevel&, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., Miller, J., Fedele, A., Collins, J., Smith, K., *et al*. 2011. Genetic heritability & shared environmental factors among twin pairs with autism. *Archives of General Psychiatry*. **68**(11):1095–1102.

Hart, A. 2001. Mann-Whitney test is not just a test of medians: differences in spread can be important. *BMJ : British Medical Journal*, **323**(7309), 391–393.

Heils, A., Teufel, A., Petri, S., Stober, G., Riederer, P., Bengel, D. & Lesch, K.P. 1996. Allelic variation of human serotonin transporter gene expression. *Journal of Neurochemistry*. **66**:2621-2624.

Hendrich, B. & Tweedie, S. 2003. The methyl-CpG binding domain & the evolving role of DNA methylation in animals. *Trends in Genetics*. **19**(5): 269-277.

Hessl, D., Tassone, F., Cordeiro, L., Koldewyn, K., McCormick, C., Green, C., Wegelin, J., Yuhas, J. & Hagerman, R.J. 2008. Brief Report: Aggression & Stereotypic Behavior in Males with Fragile X Syndrome—Moderating Secondary Genes in a ‗Single Gene‘‘ Disorder. *Journal of Autism & Developmental Disorders*. **38**:184-189.

Hilger-Eversheim, K., Moser, M., Schorle, H. & Buettner, R. 2000. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis & cell-cycle control. *Gene*. **260**(1–2):1-12.

Hirtz, D., Thurman, D.J., Gwinn-Hardy, K., Mohamed, M., Chaudhuri, A.R. & Zalutsky, R. 2007. How common are the ―common‖ neurologic disorders? *Neurology*. **68**(5):326–337.

Howlin, P. & Asgharian, A. 1999. The diagnosis of autism and Asperger syndrome: findings from a survey of 770 families. *Developmental Medicine & Child Neurology.* **41**:834–839.

Hranilovic, D., Bujas-Petkovic, Z., Vragovic, R., Vuk, T., Hock, K. & Jernej, B. 2007. Hyperserotonemia in adults with autistic disorder. *Journal of Autism Developmental Disorder*. **37**(10):1934–1940.

Hranilovic, D., Blazevic, S., Babic, M., Smurinic, M., Bujas-Petkovic, Z. & Jernej, B. 2010. 5-HT2A receptor gene polymorphisms in Croatian subjects with autistic disorder. *Psychiatry Research*. **178**(3):556–558.

Hu, X.Z., Lipsky, R.H., Zhu, G., Akhtar, L.A., Taubman, J., Greenberg, B.D., Xu, K., Arnold, P.D., *et al*. 2006. Serotonin transporter promoter gain of- function genotypes are linked to obsessive-compulsive disorder. *American Journal of Human Genetics*. **78**(5):815–826.

Hu, V. W., Sarachana, T., Kim, K. S., Nguyen, A., Kulkarni, S., Steinberg, M. E., Luu, T., Lai, Y. & Lee, N. H. 2009. Gene expression profiling differentiates autism case–controls & phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. *Autism Research*. **2**:78–97.

Hu, V. W. & Steinberg, M. E. 2009. Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. *Autism Research*. **2**:67–77.

Huang, C.H. & Santangelo, S.L. 2008. Autism & serotonin transporter gene polymorphisms: a systematic review & meta-analysis. *American journal of medical genetics Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics*. **147B**(6):903–913.

IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Ingram, V.M. 1957. Gene mutations in human haemoglobin: The chemical difference between normal and sickle cell haemoglobin. *Nature*. **180**:326-328.

Jiang, Y.H., Sahoo, T., Michaelis, R.C., Bercovich, D., Bressler, J., Kashork, C.D., Liu, Q., Shaffer, L.G., Schroer, R.J., Stockton, D.W., *et al*. 2004. A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for UBE3A. *American Journal of Medical Genetics*. **131A**:1–10.

Johnson, K.C., Smith, A.K., Stowe, Z.N., Newport, D.J. & Brennan, P.A. 2016. Preschool outcomes following prenatal serotonin reuptake inhibitor exposure: Differences in language & behavior, but not cognitive function. *Journal of Clinical Psychiatry*. **77**(2):176–182.

Jones, P.A. & Takai, D. 2001. The role of DNA methylation in mammalian epigenetics. *Science.* **293**(5532):1068-1070.

Joosten, A.V., Bundy, A.C. & Einfeld, S.L. 2009. Intrinsic & extrinsic motivation for stereotypic & repetitive behaviour. *Journal of Autism & Developmental Disorders*. **39**:521–531.

Katsuragi, S., Kunugi, H., Sano, A., Tsutsumi, T., Isogawa, K., Nanko, S. & Akiyoshi, J. 1999. Association between serotonin transporter gene polymorphism & anxiety-related traits. *Biological Psychiatry.***45**:368–370.

Kim, J-K., Stewart, R., Kang, H-J., Kim, S-W., Shin, I-S., Kim, H-R., Shin, M-G., Kim, J-T., Park, M-S., Cho, K-H. & Yoon, J.S. 2013. A longitudinal study of SLC6A4 DNA promoter methylation & poststroke depression. *Journal of Psychiatric Research*. **47**(9):1222-1227.

Kinde, B., Gabel, H.W., Gilbert, C.S., Griffith, E.C. & Greenberg, M.E. 2015. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, & MeCP2. *PNAS.* **112**(22):6800–6806.

Kinnally, E.L., Capitanio, J.P., Leibel, R., Deng, L., LeDuc, C., Haghighi, F. & Mann, J.J. 2010. Epigenetic regulation of serotonin transporter expression & behavior in infant rhesus macaques. *Genes Brain Behaviour*. **9**:575-582.

Kinnally, E.L., Feinberg, C., Kim, D., Ferguson, K., Leibel, R., Coplan, J.D. & Mann, J. 2011. DNA methylation as a risk factor in the effects of early life stress. *Brain Behaviour & Immunology*. **25**:1548-1553.

Kistner-Griffin, E., Brune C.W., Davis, L.K., Sutcliffe, J.S., Cox, N.J. & Cook, E.H.Jr. 2011. Parent-of-origin effects of the serotonin transporter gene associated with autism. *American Journal of Medical Genetics B: Neuropsychiatry Genetics.* **156**(2):139–144.

Ko, C.H. & Takahashi, J.S .2006. Molecular components of the mammalian circadian clock. *Human Molecular Genetics*. 2(2):271–277.

Koenen, K.C., Uddin, M., Chang, S.C., Aiello, A.E., Wildman, D.E., Goldmann, E. & Galea, S. 2011. SLC6A4 methylation modifies the effect of the number of traumatic events on risk for posttraumatic stress disorder. *Depression and Anxiety* **28**:639-647.

Kolevzon, A., Mathewson, K.A. & Holl&er, E. 2006. Selective serotonin reuptake inhibitors in autism: a review of efficacy & tolerability. *Journal of Clinical Psychiatry*. **67**(3):407–414.

Lamberti, G., Cascone, S. & Titomanlio, G. 2012. An engineering approach to biomedical sciences: advanced testing methods & pharmacokinetic modeling. *Translational Medicine @ Unisa.* **4**(4):34-38.

Landa, R. 2007. Early communication development & intervention for children with autism. *Mental Retardation & Developmental Disabilities Research Reviews.* **13**:16–25.

LaSalle, J.M. 2011. A genomic point-of-view on environmental factors influencing the human brain methylome, *Epigenetics*. **6**:7, 862-869.

Larsen, F., Gundersen, G., Lopez, R. & Prydz, H. 1992. CpG islands as gene markers in the human genome. *Genomics*. **13**(4):1095-1107.

Lee, B-Y., Park, S-Y., Ryu, H-M., Shin, C-Y., Ko, K-N., Han, J-Y., Koren, G. & Cho, Y-H. 2015. Changes in the methylation status of DAT, SERT & MeCP2 gene promoter in the blood cell in families exposed to alcohol during the periconceptional period. *Alcoholism: Clinical & Experimental Research*. **39**(2):239-250.

Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S., Benjamin, J., Muller, C.R., Hamer, D.H. & Murphy, D.L. 1996. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science.* **274**:1527-1531.

Lin, E., Chen, P.S., Chang, H.H., Gean, P-W., Tsai, H.C., Yang, Y.K. & Lu, R-B. 2009. Interaction of serotonin-related genes affects short-term antidepressant response in major depressive disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. **33**:1167–1172.

Lin, P-Y. 2007. Meta-analysis of the association of serotonin transporter gene polymorphism with obsessive–compulsive disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. **31**:683–689.

Lintas, C., Sacco, R. & Persico, A.M. 2012. Genome-wide expression studies in Autism spectrum disorder, Rett syndrome, & Down syndrome. *Neurobiology of Disease***. 45**(1):57–68.

Lopez-Rangel, E. & Lewis, M.E. 2006. Further evidence for epigenetic influence of MECP2 in Rett, autism & Angelman's syndromes. *Clinical Genetics*. **69**:23–25.

Lowe, R., Gemma, C., Beyan, H., Hawa, M.I., Bazeos, A., Leslie, R.D., Montpetit, A., Rakyan, V.K. & Ramagopalan, S.V. 2013. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics.* **8**(4).

Ma, D.Q., Rabionet, R., Konidari, I., Jaworski, J., Cukier, H.N., Wright, H.H., Abramson, R.K., Gilbert, J.R., Cuccaro, M.L., Pericak-Vance, M.A., *et al*. 2010. Association & gene-gene interaction of SLC6A4 & ITGB3 in autism. *American Journal of Medical Genetics B: Neuropsychiatry Genetics*. **153B**(2):477–483.

Madden, A.M.K. & Zup, S.L. 2014. Effects of developmental hyperserotonemia on juvenile play behavior, oxytocin & serotonin receptor expression in the hypothalamus are age & sex dependent. *Physiology & Behavior*. **128**:260–269.

Makkonen, I., Riikonen, R., Kokki, H., Airaksinen, M.M. & Kuikka, J.T. 2008. Serotonin & dopamine transporter binding in children with autism determined by SPECT. *Developmental Medicine & Child Neurology*. **50**(8):593–597.

Malcolm-Smith, S., Hoogenhout , M., Ing, N., Thomas, K.G.F. & de Vries, P. 2013. Autism spectrum disorders—Global challenges & local opportunities. *Journal of Child & Adolescent Mental Health*. **25**(1):1-5.

Mandy, W., Charman, T., Puura, K. & Skuse, D. 2014. Investigating the cross-cultural validity of *DSM-5* autism spectrum disorder: Evidence from Finnish & UK samples. *Autism.* **18**:45-54.

Marian, A.J. 2002. Molecular genetic studies of complex phenotypes. *Translational Research.* **159**:2:64-79.

McDougle, C.J., Naylor, S.T., Cohen, D.J., Aghajanian, G.K., Heninger, G.R. & Price, L.H. 1996. Effects of tryptophan depletion in drug-free adults with autistic disorder. *Archive of General Psychiatry*. **53**(11):993–1000.

McDougle, C.J., Naylor, S.T., Cohen, D.J., Volkmar, F.R., Aghajanian, G.K., Heninger, G.R. & Price, L.H. 1996. A double-blind, placebo-controlled study of fluvoxamine in adults with autistic disorder. *Archive of General Psychiatry*. **53**(11):1001–1008.

McDougle, C.J., Epperson, C.N., Price, L.H. & Gelernter, J. 1998. Evidence for Linkage Disequilibrium between Serotonin Transporter Protein Gene (SLC6A4) & Obsessive Compulsive Disorder. *Molecular Psychiatry*. **3**:270–273.

McGowan, P.O., Sasaki, A., D'Alessio, A.C., *et al*. 2009. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. *Nature Neuroscience.* **12**(3):342–348.

Melnyk, S., Fuchs, G.J., Schulz, E., *et al*. 2012. Metabolic imbalance associated with methylation dysregulation & oxidative damage in children with autism. *Journal of Autism & Developmental Disorders.* **42**(3):367–377.

Mill, J., Tang, T.T., Kaminsky, Z., Khare, T., Yazdanpanah, S., Bouchard, L., Jia, P., Assadzadeh, A., Flanagan, J., Schumacher, A., Wang, S-C. & Petronis, A. 2008. Epigenomic profiling reveals DNA-methylation changes associated with major psychosis. *The American Journal of Human Genetics*. **82**(3):696-711.

Murphy, D.G., Daly, E., Schmitz, N., Toal, F., Murphy, K., Curran, S., Erl&sson, K., Eersels, J., Kerwin, R., Ell, P., *et al*. 2006. Cortical serotonin 5-HT2A receptor binding & social communication in adults with Asperger's syndrome: an in vivo SPECT study. *American Journal of Psychiatry*. **163**(5):934–936.

Nakamura, M., Ueno, S., Sano, A. & Tanabe, H. 2000. The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. *Molecular Psychiatry*. **5**:32-38.

Naslund, J., Studer, E., Pettersson, R., Hagsater, M., Nilsson, S., Nissbr&t, H. & Eriksson, E. 2015. Differences in anxiety-like behaviour within a batch of wistar rats are associated with differences in serotonergic transmission, enhanced by acute SRI administration, & abolished by serotonin depletion. *International Journal of Neuropsychopharmacology*. 1–9.

Neale, B.M & Sklar, P. 2015. Genetic analysis of schizophrenia & bipolar disorder reveals polygenicity but also suggests new directions for molecular interrogation, *Current Opinion in Neurobiology*. **30**:131-138.

Newschaffer, C.J., Croen, L.A., Daniels, J., Giarelli, E., Grether, J.K., Levy, S.E., M&ell, D.S., Miller, L.A., Pinto-Martin, J., Reaven, J., *et al*. 2007. The epidemiology of autism spectrum disorders. *Annual Review of Public Health*. **28**:235–258.

Nguyen, A., Rauch, T.A., Pfeifer, G.P. and Hu, V.W. 2010. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders & a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB Journal.* **24**(8):3036–3051.

Nyffeler, J., Walitza, S., Bobrowski, E., Gundelfinger, R. & Grünblatt, E. 2014. Association study in siblings & case-controls of serotonin- & oxytocin-related genes with high functioning autism. *Journal of Molecular Psychiatry.* **2**:1.

Olsson, C.A., Foley, D.L., Parkinson-Bates, M., Byrnes, G., McKenzie, M., Patton, G.C., Morley, R., Anney, R.J.L., Craig, J.M. & Saffery, R. 2010. Prospects for epigenetic research within cohort studies of psychological disorder: A pilot investigation of a peripheral cell marker of epigenetic risk for depression. *Biological Psychology*. **83**(2):159-165.

Park, S., Lee, J-M., Kim, J-W., Cho, D-Y., Yun, H-J., Han, D-H., Cheong, J-H & Kim, B-N. 2015. Associations between serotonin transporter gene (SLC6A4) methylation & clinical characteristics & cortical thickness in children with ADHD. *Psychological Medicine*. **45**:3009–3017.

Parracho, H.R.M.T., Bingham, M.O., Gibson, R., McCartney, A.L. 2005. Differences between the gut microflora of children with autistic spectrum disorders & that of healthy children. *Journal of Medical Microbiology*. **54**:987-991.

Persico, A.M., Pascucci, T., Puglisi-Allegra, S., Militerni, R., Bravaccio, C., Schneider, C., Melmed, R., Trillo, S., Montecchi, F., Palermo, M., *et al*. 2002. Serotonin transporter gene promoter variants do not explain the hyperserotoninemia in autistic children. *Molecular Psychiatry*. **7**(7):795–800.

Persico, A.M. & Bourgeron, T. 2006. Searching for ways out of the autism maze: genetic, epigenetic & environmental clues. *Trends in Neurosciences.* **29**(7):349–358.

Philibert, R., Madan, A., &ersen, A., Cadoret, R., Packer, H. & Sandhu, H. 2007. Serotonin transporter mRNA levels are associated with the methylation of an upstream CpG island. *American Journal of Medical Genetics B: Neuropsychiatry Genetics*. **144B**:101-105.

Philibert, R., Sandhu, H., Hollenbeck, N., Gunter, T., Adams, W. & Madan, A. 2008. The relationship of 5HTT (SLC6A4) methylation & genotype on mRNA expression & liability to major depression & alcohol dependence in subjects from the Iowa adoption studies. *American Journal of Medical Genetics B: Neuropsychiatry Genetics*. **147B**:543-549.

Ramoz, N., Reichert, J.G., Corwin, T.E., Smith, C.J., Silverman, J.M., Hollander, E. & Buxbaum, J.D. 2006. Lack of evidence for association of the serotonin transporter gene SLC6A4 with autism. *Molecular Psychiatry*. **60**(2):186–191.

Reaume, C.J. & Sokolowski, M.B. 2011. Conservation of gene function in behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences.* **366**(1574):2100–2110.

Rodriguez-Porcel, F., Green, D., Khatri, N., Swilley Harris, S., May, W.L., Lin, R.C., *et al*. 2011. Neo-natal exposure of rats to antidepressants affects behavioral reactions to novelty & social interactions in a manner analogous to autistic spectrum disorders. *The Anatomical Record*. **294**:1726–1735.

Rogers, S. & Vismara, L. 2008. Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child & Adolescent Psychiatry.* **37**:8–38.

Samaco, R.C., Hogart, A. & LaSalle, J.M. 2005. Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A & GABRB3. *Human Molecular Genetics*. **14**:483–492.

Schanen, N.C. 2006. Epigenetics of autism spectrum disorders. *Human Molecular Genetics* **15**(2):138–150.

Simpson, K.L., Weaver, K.J., de Villers-Sidani, E., Lu, J.Y., Cai, Z., Pang, Y. *et al*. 2011. Perinatal anti-depressant exposure alters cortical network function in rodents. *Proceedings of the National Academy of Sciences.* **108**:18465–18470.

Suchiman, H., Eka, D., Slieker, R.C., Kremer, D., Slagboom, P.E., Heijmans, B.T. & Wouter, T.E. 2015. Design, measurement & processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. *Frontiers in Genetics*. **6**.

Sun, X., Allison, C., Matthews, F.E., Sharp, S.J., Auyeung, B., Baron-Cohen, S. & Brayne, C. 2013. Prevalence of autism in mainland China, Hong Kong & Taiwan: a systematic review & meta-analysis. *Molecular Autism*. **4**(1):7.

Sutcliffe, J.S., Delahanty, R.J., Prasad, H.C., McCauley, J.L., Han, Q., Jiang, L., Li, C., Folstein, S.E. & Blakely, R.D. 2005. Allelic heterogeneity at the serotonin transporter Locus (SLC6A4) confers susceptibility to autism & rigid-compulsive behaviors. *American Journal of Human Genetics*. **77**:265–279.

Sykes, N.H. & Lamb, J.A. 2007. Autism: the quest for the genes. *Expert Reviews in Molecular Medicine.* **9**(24):1–15.

Szyf, M. 2015. Epigenetics, a key for unlocking complex CNS disorders? Therapeutic implications. *European Neuropsychopharmacology*. **25**(5):682-702.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology & Evolution*. **30**:2725-2729.

Titlestad, I. L., Kyvik, K. O., Kristensen, T. & Lillevang, S. 2002. HLA haplotypes in dizygotic twin pairs: are dizygotic twins more similar than sibs? *Twin Research*. **5**(4):287–288.

Tordjman, S., Somogyi, E., Coulon, N., Kermarrec, S., Cohen, D., Bronsard, G., Bonnot, O., Weismann-Arcache, C., *et al*. 2014. Gene × environment interactions in autism spectrum disorders: role of epigenetic mechanisms. *Frontiers in Psychiatry*. **5**:53.

Tsankova, N., Renthal, W., Kumar, A. & Nestler, E.J. 2007. Epigenetic regulation in psychiatric disorders. *Nature Reviews Neuroscience.* **8**:355-367.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. & Rozen SG. 2012. Primer3--new capabilities & interfaces. *Nucleic Acids Research*. **1**:40(15):115.

van Ijzendoorn, M.H., Caspers, K., Bakermans-Kranenburg, M.J., Beach, S.R. & Philibert, R. 2010. Methylation matters: interaction between methylation density & serotonin transporter genotype predicts unresolved loss or trauma. *Biological Psychiatry*. **68**:405-407.

Van Vliet, J., Oates, N.A. & Whitelaw, E. 2007. Epigenetic mechanisms in the context of complex diseases. *Cellular & Molecular Life Science.* **64**(12):1531–1538.

Vijayendran, M., Beach, S.R., Plume, J.M., Brody, G.H. & Philibert, R.A. 2012. Effects of genotype & child abuse on DNA methylation & gene expression at the serotonin transporter. *Frontiers in Psychiatry*. **3**:55.

Völkel, S., Stielow, B., Finkernagel, F., Stiewe, T., Nist, A. & Suske, G. 2015. Zinc finger independent genome-wide binding of Sp2 potentiates recruitment of histone-fold protein Nf-y distinguishing it from Sp1 & Sp3. *PLoS Genet*ics. **11**(3):100-102.

Voyiaziakis, E., Evgrafov, O., Li, D., Yoon, H-J., Tabares, P., Samuels, J., Wang, Y., Riddle, M.A., *et al*. 2011. Association of *SLC6A4* variants with obsessive-compulsive disorder in a large multicenter US family study**.** *Molecular Psychiatry*. **16:**108–120.

Wang, D., Szyf, M., Benkelfat, C., Provencal, N., Turecki, G., Caramaschi, D., Coˆte´, S.M., Vitaro, F., *et al*. 2012. Peripheral SLC6A4 DNA methylation is associated with in vivo measures of human brain serotonin synthesis & childhood physical aggression. *PLoS ONE*. **7**:39501.

Wendland, J., Martin, B., Kruse, M., Lesch, K. & Murphy, D. 2006. Simultaneous genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-HTTLPR & rs25531. *Molecular Psychiatry*. **11**:224–226.

Wiggins, L.D., Baio, J. & Rice, C. 2006. Examination of the time between first evaluation & first autism spectrum diagnosis in a population-based sample. *Journal of Developmental & Behavioral Pediatrics.* **27**:79–87.

Williams, J.G., Higgins, J.P. & Brayne, C.E. 2006. Systematic review of prevalence studies of autism spectrum disorders. *Archives of Diseases in Childhood*. **91**(1):8–15.

Williams, T. & Tjian, R 1991. Analysis of the DNA-binding & activation properties of the human transcription factor AP-2. *Genes & Development*. **5**(4):670–82.

Wilson, C.A., Pearson, J.R., Hunter, A.J., Tuohy, P.A. & Payne, A.P. 1986. The effect of neonatal manipulation of hypothalamic serotonin activity in the adult rat. *Biochemical Behavior*. **24**:1175–83.

Wong, C.C., Caspi, A., Williams, B., Craig, I.W., Houts, R., Ambler, A., Moffitt, T.E. & Mill, J. 2010. A longitudinal study of epigenetic variation in twins. *Epigenetics*. **5**:516-526.

Wong, C.C., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C. & Mill, J. 2014. Methylomic analysis of monozygotic twins discordant for autism spectrum disorder & related behavioural traits. *Molecular Psychiatry*. **19**(4):495–503.

Wotton, D., Ghysdael, J., Wang, S., Speck, N.A. & Owen, M.J. 1994. Cooperative binding of Ets-1 & core binding factor to DNA. *Molecular & Cellular Biology*. **14**:840-860.

Yeo, A., Boyd, P., Lumsden, S., Saunders, T., Handley, A., *et al*. 2004. Association between a functional polymorphism in the serotonin transporter gene & diarrhoea predominant irritable bowel syndrome in women. *Gut*. **53**:1452–1458.

Yirmiya, N., Pilowsky, T., Nemanov, L., Arbelle, S., Feinsilver, T., Fried, I. & Ebstein, R.P. 2001. Evidence for an association with the serotonin transporter promoter region polymorphism & autism. *American Journal of Medical Genetics*. **105**(4):381–386.

Zhao, J., Goldberg, J. & Vaccarino, V. 2013. Promoter methylation of serotonin transporter gene is associated with obesity measures: a monozygotic twin study. *International Journal of Obesity*. **37**:140 – 145.

Zhu, L., Wang, X., Li, X-L., Towers, A., Cao, X., Wang, P., Bowman, R., Yang, H., Goldstein, J., Li, Y-L. & Jiang, Y-H. 2014. Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders. *Human Molecular Genetics*. **23**(6):1563–1578.

APPENDIX

Table A. Primer and amplicon information

Amplicon purpose (size)	Forward primer (5'-3')	Reverse primer (5'-3')
Repeat polymorphism (512/469bp)	TCCTCCGCTTTGGCGCCTCTTCC	TGGGGGTTGCAGGGGAGATCCTG
Bisulphite DNA sequencing (201bp)	AGGTTTTAGGAAGAAAGAGAGAGT	CGAACGAAAAATCCTAACTTTCC
Native DNA sequencing (444bp)	GCCGGTCAGTCAGATAAACG	CGTCACTTTGAGGCGAATAAA
EpiTYPER MassARRAY (467bp)	TTGTTAGGTTTTAGGAAGAAAGAGAGA	CCCTCACATAATCTAATCTCTAAATAACC

Table B. Participant information for the ASD and control cohorts

	ASD cohort	Control cohort
Age	8.44±2.40	7.00±1.55
European ancestry	44% (n=11)	100% (n=6)
African ancestry	12% (n=3)	0% (n=0)
Mixed ancestry	44% (n=11)	0% (n=0)

Mean values are presented as values ± standard deviation

Table C. Two-tailed Fisher's exact test p-values for each comparison of allelic and genotypic frequency distributions across the endophenotypes

Comparison	L vs. S	LL vs. LS/SS	H vs. L expressers
ASD vs. control	0.100	0.049*	0.300
ASD-mod. vs. control	0.170	0.142	0.344
ASD-high vs. control	0.074	0.061	0.200
ASD-mod. vs. ASD high	0.267	0.544	0.539
Pre-verb. vs. control	0.260	0.141	0.344
Phrase speech. vs. control	0.121	0.136	0.329
Pre-verb./phrase speech vs. control	0.099	0.136	0.319
Pre-verb. vs. phrase speech	0.756	1.000	1.000
High SA vs. control	0.099	0.064	0.311
High RRB vs. control	0.074	0.061	0.182
High RRB vs. low RRB	0.475	0.545	0.545
Ancestry	0.414	0.440	0.226

Significance is denoted by *, set at $p<0.05$.

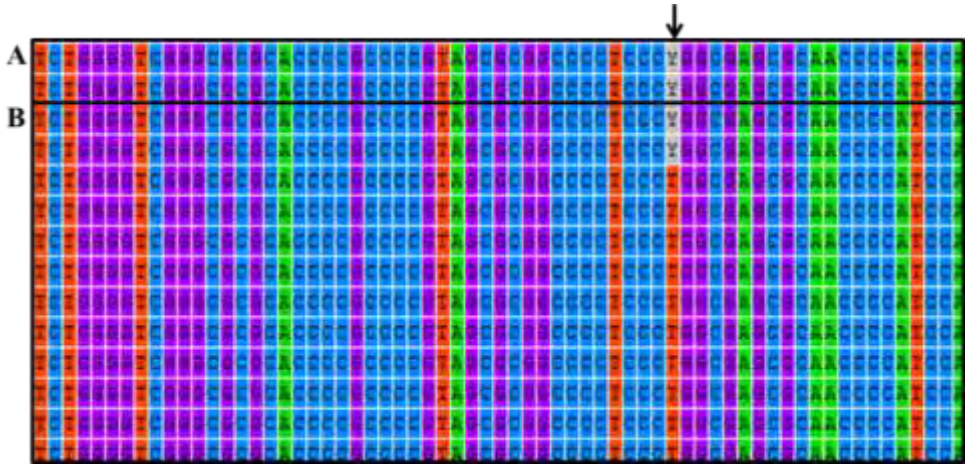

Figure A. Multiple sequence alignment in MEGA of a partial sequence of native DNA of the *SLC6A4* promoter, upstream of the gene. A represents the forward and reverse sequencing reactions of one ASD participant, while B shows the forward and reverse sequencing of all six control participants. The arrow indicates the presence of the rs25533 SNP, with Y demonstrating a heterozygote genotype of TC.

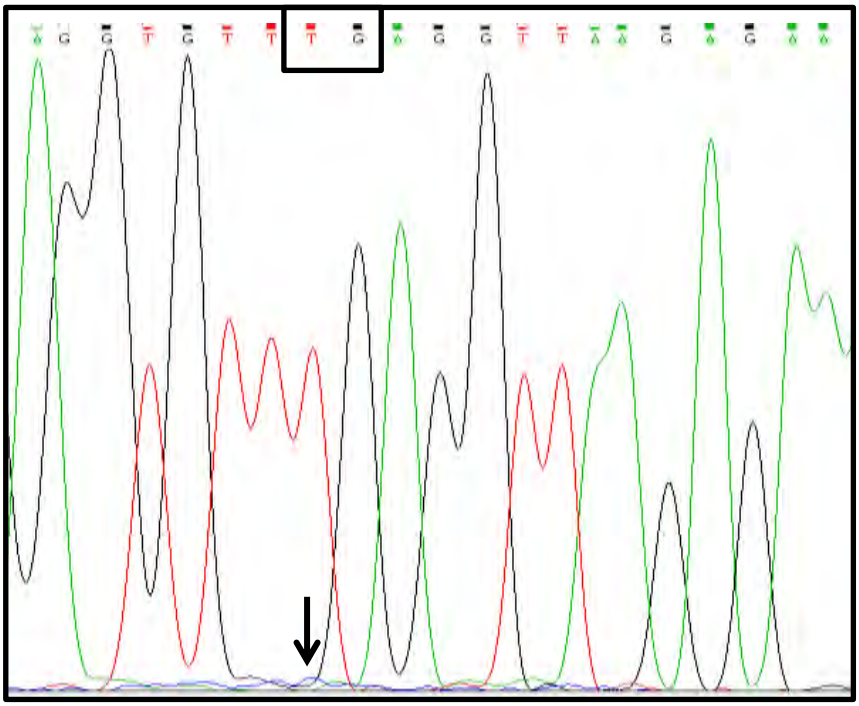

Figure B. Chromatogram showing partial sequence of the converted DNA of the *SLC6A4* promoter, upstream of the gene. The box contains CpG site 4, with the arrow indicating a slight cytosine peak.

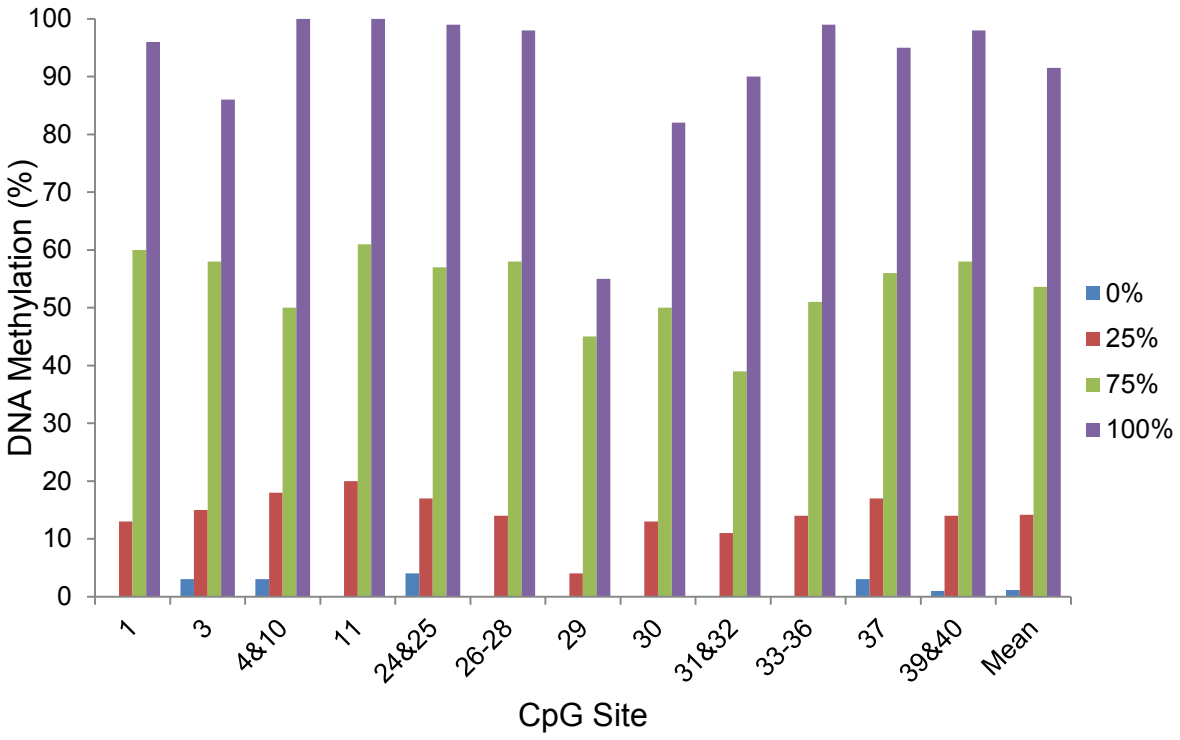

Figure C. Average DNA methylation of the duplicate technical controls of 0, 25, 75 and 100% methylated DNA for each target CpG site and DNA methylation across the 12 CpG regions. The technical controls were tested for every CpG site, with the mean methylation calculated.

Table D: Mann-U Whitney p-values for each comparison of CpG site and mean DNA methylation percentages across the endophenotypes

Comparison	1	3	4/10	11	24/25	26-28	29	30	31/32	33-36	37	39/40	Mean
ASD vs. control	0.564	0.670	0.330	0.146	0.318	0.459	0.594	0.222	0.798	0.274	0.465	0.702	0.011*
ASD-mod. vs. control	0.802	0.661	0.267	0.109	0.373	0.620	0.557	0.067	0.933	0.307	0.496	0.697	0.004*
ASD-high vs. control	0.160	0.812	0.823	0.628	0.292	0.320	0.898	0.636	0.697	0.333	0.593	0.974	0.457
ASD-mod. vs. high	0.777	0.856	0.253	0.498	0.831	0.310	0.966	0.000[a]	0.697	0.861	0.911	0.364	0.111
Pre-verb. vs. control	0.265	0.405	0.503	0.133	0.534	0.451	0.473	0.344	0.965	0.165	0.314	0.695	0.004*
Phrase speech. vs. control	0.553	0.924	0.315	0.240	0.160	0.522	0.772	0.161	1.000	0.933	0.628	1.000	0.026*
Pre-verb./phrase speech vs. control	0.303	0.615	0.369	0.121	0.300	0.424	0.525	0.192	0.825	0.340	0.648	0.685	0.003*
Pre-verb. vs. phrase speech	1.000	0.166	0.599	0.725	0.288	0.991	1.000	0.168	1.000	0.165	0.282	0.639	0.557
High SA vs. control	0.406	0.643	0.373	0.152	0.333	0.382	0.525	0.291	0.985	0.323	0.573	0.782	0.006*
High vs. low RRB	0.235	0.843	0.675	0.861	0.687	0.457	0.823	0.001*	0.983	0.745	0.249	0.636	0.816
High RRB vs. control	0.045*	0.675	0.574	0.286	0.290	0.359	0.610	0.935	1.000	0.526	0.860	0.654	0.045*
Ancestry	0.657	0.546	0.401	0.550	0.008*	0.216	0.684	0.659	0.640	0.792	0.852	0.755	0.080
LL vs. LS/SS	0.457	0.358	0.532	0.190	0.935	0.516	0.422	0.868	0.966	0.959	0.212	0.395	0.445
H vs. L expressers	0.538	0.773	0.552	0.386	0.972	0.717	0.366	0.556	0.633	0.619	0.246	0.212	0.557
Age	0.127	0.725	0.109	0.634	0.323	0.107	0.935	0.456	0.146	0.987	0.562	0.853	0.358

Significance denoted by *, set at $p<0.05$

Significance without homogeneity of variance is denoted by [a]

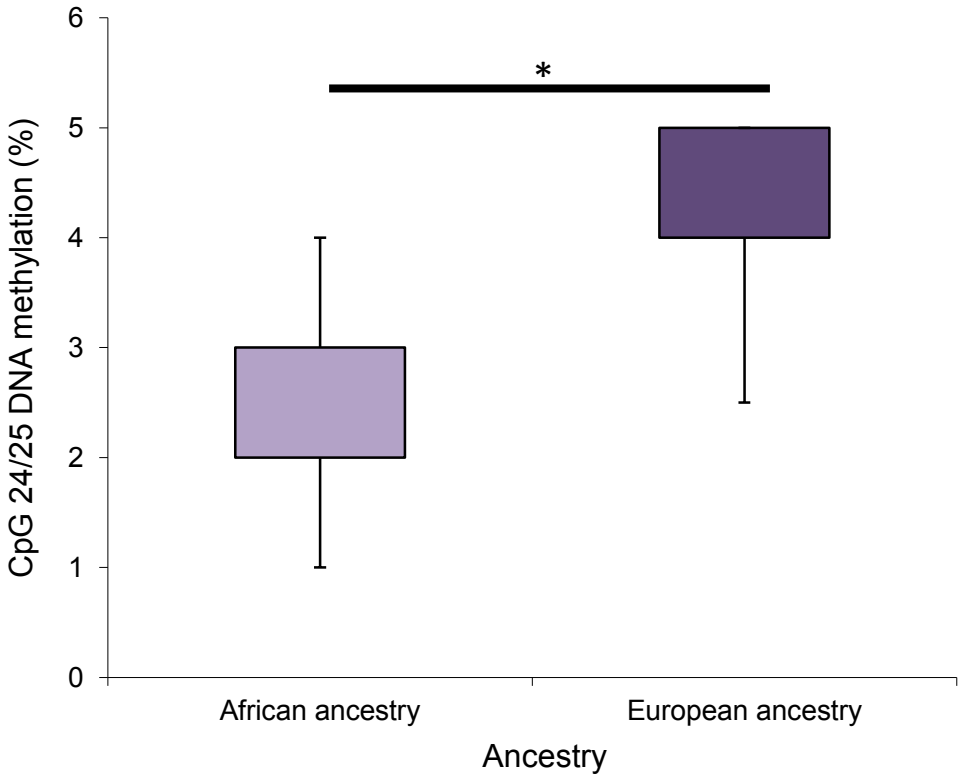

Figure D. Average DNA methylation for the CpG 24/25 sites of the *SLC6A4* promoter for participants of African ancestry (n=14) and European ancestry (n=11). DNA methylation levels of the two ancestry populations were compared using the Mann-U Whitney test. Significance is denoted by * ($p<0.05$).

CONSENT FORM FOR BIOLOGICAL AND PSYCHOLOGICAL RESEARCH IN ASD

PURPOSE

The Molecular Autism Research in South Africa (MARSA) group (Department of Molecular and Cell Biology, UCT) is conducting a research that investigates how genes, social communication deficits, restricted and repetitive behaviours and interests are implicated in Autism Spectrum Disorder (ASD). This multi-disciplinary initiative includes research by geneticists, psychologists, clinical practitioners, and other health practitioners. We are gathering phenotype and genetic information from children who have the above-mentioned deficits or characteristics, or who have a formal or tentative diagnosis of ADHD, Dyspraxia, Sensory integration difficulties, Aspergers or Autism Spectrum Disorder, and invite you to participate in our research.

PROCEDURES

The genetic aspect of the study: If you agree to participate in our study, we will ask for cheek cell swabs from your child. The cheek swabs are painless where we will gently rub a sterile cotton swab across the inside of each side of your child‘s cheeks for about 30 seconds. All samples will be taken to the MARSA laboratory where DNA will be extracted to examine the genetic make-up of a number of selected genes. The genes we will target are thought to be associated with a specific characteristic of interest.

The non-genetic aspects of the study will be an Autism Diagnostic Observation Schedule (ADOS-2) assessment to identify social deficits and restricted and behaviours. This assessment is play-based session that will take about 1 hour to complete.

RISKS

There are no medical or psychological risks associated with this research.

BENEFITS

Although you personally will not receive any direct benefit from this project, individuals (or their family members, and future generations) who might develop, or have the above mentioned diagnosis of ASD may in the future benefit if we can locate genes associated with, or that lead to such disorders, or locate genes that predict responses to different medications. We do not expect to discover any information of direct clinical relevance to your condition or your treatment during the next few years. In line with the current regulations, we will not make the results of the genetic tests available to subjects or their doctors. However, the results of the play-based scoring sessions, the ADOS-2, may be passed on to your neuro-paediatrician or psychologist on request.

COSTS AND COMPENSATION

There are no costs to participants and their families in this research project.

CONFIDENTIALITY

We will keep your name and any other personal information we learn about you confidential. This information will not be given out to the repository or to anyone else. We will take the following steps to ensure confidentiality: a research number will be assigned to you when you enter the study and thereafter your name will not be used on any of the data. The results from the analysis of your DNA will not be released or shared in any way

with your relatives, with insurance companies, or any third party not involved in research. When results of this study are published, your name will not be used.

WITHDRAWAL FROM THE RESEARCH PROJECT

You have the right to leave the study at any time without giving any reason, and without penalty. If you wish to leave the study, contact Dr Colleen O'Ryan.

CONTACTS

If you have any questions about the study, you may contact Dr O'Ryan at genetics.kidslab@gmail.com or 021 650 2457.

PARTICIPATION

I give consent for my child to participate in the following aspects of this study:

- Provide cheek swabs and participate in the play-based assessment (ADOS-2)* which I consent to have recorded on video tape/DVD** Yes / No
- Contact you for future research Yes /No

** A familiar educator/carer is given permission to be present during the ADOS if the learner is pre-verbal / uses single or phrase speech words only*

***The video will used for scoring the assessment, is kept strictly confidential and the children's identities will not be revealed;*

If you sign below, it means that you have read (or have had read to you) the information given in this consent form, and you would like to be a volunteer in this study.

Parent / Guardian Name:	Child's Name:
Email:	Tel:
Signature:	Date:
Signature of Principal Investigator:	Date:
Witness:	Date