UNIVERSITY OF CAPE TOWN

ASMT Gene Polymorphisms are Associated with Autism Spectrum Disorder (ASD) Symptom Severity in a South African Population

Margaretha de Waal (DWLMAR005)

University of Cape Town

Submitted: April 2016

Supervisor: Dr Colleen O'Ryan

Co-supervisor: Dr Laura Roden

Department of Molecular and Cell Biology

Submitted to the University of Cape Town in fulfilment of the requirements for the degree: Master of Science in Molecular and Cell Biology

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or noncommercial research purposes only. Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

I, Margaretha de Waal, know the meaning of plagiarism and declare that all of the work in the thesis, save for that which is properly acknowledged, is my own and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr Colleen O'Ryan and Dr Laura Roden, for designing, funding, and facilitating my project, as well as for their emotional support, and believing in me when I went through difficult times. I would like to thank the MCB Equity Bursary committee for funding part of my studies and living expenses. I am also extremely grateful to my lab mates Michaella Hulley and Gerrit van Wyk for standing by me, and for all the scientific and non-scientific discussions and support. You guys are awesome. I also extend my thanks to the schools, parents, and children who participated in this study. Without you this study would not be possible. And, finally, my love and thanks go out to my sisters, parents, grandmother and fiancé, for all the love, support, and endless cups of tea.

TABLE OF CONTENTS

Abstract 1
List of Tables 2
List of Figures 3
1 Introduction 4
1.1 Autism Spectrum Disorders 4
1.2 Circadian Cycles and the Role of Melatonin 6
1.3 Sleep Disruption in Autism Spectrum Disorders 9
1.4 Acetylserotonin Methyl Transferase: A Candidate Gene In Autism 11
1.5 Molecular Study Design 11
1.6 ASD Endophenotyping 12
1.7 The Present Study 13
2 Materials and Methods 14
2.1 Ethics clearance 14
2.2 Research Participants 14
2.3 Samples and Nucleic Acid extraction 15
2.4 Polymerase Chain Reaction 16
2.5 Sequencing 17
2.6 Single Nucleotide Polymorphism Genotyping 17
2.7 Statistical analysis 17
2.8 Phenotypic Grouping 18
3 Results 19
3.1 Characteristics of Participants in the Study Cohort 19
3.2 Endophenotypes 20
3.3 Putative Single Nucleotide Polymorphisms Associated with Autism 21
3.4 Haplotypes 24
3.5 Deleterious Single Nucleotide Polymorphisms 27
4 Discussion 32
4.1 Verbal Fluency and Autism Spectrum Disorders 32
4.2 Autism Risk Haplotypes 37
4.3 Conclusion and Future research 38
5 Bibliography 40
6 Supplementary Materials 45

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by behavioural and social impairments. ASD shows evidence of a genetic aetiology, with a large body of research linking ASD to polymorphisms in several different genes and gene families, including those involved in circadian rhythm generation and melatonin biosynthesis. Sleep disorders are highly comorbid with ASD in both children and adults, and range from sleep onset delay, phase shift and sleep disruption. These parasomnias can have a significant impact on the quality of life for persons with ASD and their families, and sleep deprivation can feed into the behavioural deficits in ASD. Melatonin supplementation is often prescribed to assist in alleviating the above mentioned sleep dysfunction. Melatonin is a hormone in the circadian clock system, and is a biochemical signal for darkness to synchronise peripheral cells to the master oscillator. Clinical trials reported that melatonin supplementation at night assists in sleep initiation. However both the mode of action of supplemental melatonin, as well as whether melatonin deficiency is common in ASD, remains unclear. Furthermore, any research on ASD is often hamstrung by the heterogeneous nature of the disorder, necessitating clear phenotyping. This study examines single nucleotide polymorphisms (SNPs) in the gene acetylserotonin methyl transferase (ASMT), which encodes an enzyme in melatonin biosynthesis, in a South African ASD cohort (n=28) and controls (n=6). All participants completed and Autism Diagnostic Observation Schedule-2 assessment that allowed partitioning of the ASD individuals into ASD endophenotypes, to reduce phenotyping heterogeneity. This study found SNPs previously associated with ASD in the promoter and intronic region. Additionally, this study found novel SNPs, and a SNP in a putative transcription factor binding site not previously associated with ASD. The associations found between SNPs and ASD endophenotypes, together with the positions of the SNPs, suggest a potential link between ASMT polymorphisms and ASD symptom severity. Further research, using language assessment tools as well as quantitative measures of melatonin and sleep disruption, may establish the role of melatonin in language impairment in ASD.

LIST OF TABLES

Table 1.1 Cut-Off Scores For Classification Of ASD Using The ADOS……………………………………13

Table 2.1 The Selection Criteria For The Cohort Used In This Study……………………………………15

Table 3.1 Selection Criteria For The Endophenotypes Used In This Study……………………………20

Table 3.2 Genotype Frequencies Of The Control Group And The ASD Cohort…………………….22

Table 3.3 Allele frequencies in the 1000 Genomes Project and this current study………………23

Table 3.4 SNPs Found To Differ In Frequency Between Endophenotypes……………………………24

Table 3.5 Haplotype Counts In The ASD Group vs. The Controls………………………………………25

Table 3.6 Haplotype Counts For The Melke 'Risk' Haplotype……………………………………………27

Table 3.7 Position And Functional Implication Of Non-Synonymous SNPs………………………….28

Table 3.8 Participant Data For SNPs Associated With An Endophenotype…………………………..31

LIST OF FIGURES

Figure 1.1 Circadian Cycle and the Role of Melatonin……………………………………………………..7

Figure 1.2 The Melatonin Biosynthetic Pathway…………………………………………………………..8

Figure 3.1 Linkage Disequilibrium Maps…………………………………………………………………......26

Figure 3.2 Promoter Sequence and Putative Transcription Factor Binding Sites………………...29

Figure 3.3 Intronic Sequence and Putative Transcription Factor Binding Sites……………………30

1 INTRODUCTION

1.1 AUTISM SPECTRUM DISORDERS

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by behavioural and social impairments. The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) consolidated the diagnoses of Asperger's Disorder, Pervasive Developmental Disorder, and Autism Disorder into Autism Spectrum Disorders (ASD) (American Psychiatric Association 2013). These disorders represent a sliding scale of impairments from mild to severe symptoms, including limitations in social communication, problems in interpersonal interaction, repetitive behaviours and restricted interests. ASD is frequently associated with behavioural comorbidities, such as anxiety or depressive disorders, as well as medical comorbidities including epilepsy and sleep disruption (Jonsson et al. 2010; Hu & Steinberg 2009, American Psychiatric Association 2013). ASD exhibits strong heritability, and is a life-long condition. The median prevalence is estimated to be 65 in every 10,000 people worldwide (Elsabbagh et al. 2012; Jonsson et al. 2010).

ASD manifests with an extremely heterogeneous symptom set; symptom severity ranges from mild to severe or debilitating. The DSM-4 diagnosis of Asperger's disorder represented the milder end of the spectrum, characterised by relatively average to above-average IQ and verbal fluency, yet still exhibiting intrinsic difficulty with interpersonal interaction (Hu & Steinberg 2009, American Psychiatric Association 2000). While such individuals can function independently and effectively in society, others with more severe impairments or comorbidities may remain largely dependent on parents or care-givers (Elsabbagh et al. 2012). Severely affected individuals may present with symptoms including cognitive impairment, compulsions, restricted verbal fluency, sensory aversions or savant skills (Hu et al. 2009; Jonsson et al. 2010; Courchesne et al. 2007).

ASD manifests in early childhood, and the first symptoms (such as delays in developmental milestones) appearing between one and two years of age (Courchesne et al. 2007). Some children may only show ASD symptoms after a period of typical development, sometimes losing previously acquired verbal or non-verbal communicative skills. This phenomenon is known as autistic regression, and occurs in approximately one in every three ASD cases (see

Giannotti *et al* (2008) and sources within). Clinical diagnosis of ASD or Pervasive Developmental Disorder- Not Otherwise Specified (PDD-NOS), the mildest phenotype in ASD (American Psychiatric Association 2000), is usually made between the age of one to four years of age (Courchesne et al. 2007; Szatmari et al. 2015). Currently, there is no quantitative medical test for ASD and diagnosis is made using standardised psychological diagnostic tools such as the Autism Diagnostic Observation Schedule-2 (ADOS-2) together with developmental history questionnaires such as the Autism Diagnostic Interview – Revised (ADI-R) (Lord et al. 1994; Lord et al. 1989).

While genetic tests for autism have not been developed yet, ASD shows strong evidence of a genetic aetiology, with a large body of research linking ASD to polymorphisms in several different genes and gene families (Glessner et al. 2009; Arieff et al. 2010; Bartlett et al. 2005; Jonsson et al. 2014; Nicholas et al. 2007; Pinto et al. 2010). However, genetic studies are hamstrung by the variable nature of the behaviours of the autism spectrum, making it extremely difficult to identify target genes from such a heterogeneous phenotype. Due to the variable nature of ASD it is essential to consider "subtypes" or endophenotypes of autism when looking at the disorder at a molecular level. A good example of this approach is the study by Hu et al. (2009) which used microarrays to examine thousands of genes for differential expression. This study divided phenotypic variants of ASD into three groups, to compare the gene expression of these groups to one another and to age-matched controls.

An unprecedented result of this study was the differences in gene expression that were observed both among the three groups of ASD participants and between the ASD group and the controls. This study emphasises the importance of clear phenotyping in molecular research, which includes sequencing studies examining single nucleotide polymorphisms (SNPs). In order to reduce the heterogeneity in a sample group, past studies would identify a particular trait, for example sleep disruption, or the degree of age-appropriate language acquisition in participants (Hu et al. 2009) (Veatch, Pendergast, et al. 2015; Jonsson et al. 2010; Melke et al. 2008).

1.2 Circadian Cycles and the Role of Melatonin

Molecular studies on neuropsychiatric disorders, including ASD, have implicated a variety of regulatory or responsive genes that form part of the human circadian clock system (Hu et al. 2009; Talarowska et al. 2014; Gałecki et al. 2010; Etain et al. 2012; Geoffroy et al. 2015; Li et al. 2012). Circadian rhythms are generated by an organism's internal clock, which regulates many physiological processes such as immunity, metabolism and sleep patterns. Differential expression of circadian genes has also been associated with endophenotypes within ASD. In a previous study, Hu *et al* (2009) subdivided their ASD cohort into three different groups ("savant", "severe language impairment" and "mild ASD"), and compared the gene expression among these groups, and to controls. They found that 15 circadian genes were differentially expressed specifically in the group with severe language impairment, and not in the other two ASD groups, when compared to the control group (Hu et al. 2009). This highlights not only the importance of endophenotyping in ASD studies, but also that circadian genes may be implicated in language development or symptom severity in ASD. Circadian cycles also regulate behavioural processes such as social timing, for example interactive turn-taking in conversation, or rapid shifts in attention (Wimpory et al. 2002). It has been proposed that such timing difficulties may be central to ASD, potentially even playing a causative role (Nicholas et al. 2007; Wimpory et al. 2002). However, the mechanism by which circadian disruption directly influences the development of ASD has not been elucidated.

In mammals, the circadian system is directed by the master oscillator, situated in the superchiasmatic nucleus (SCN) of the hypothalamus (Figure 1.1). This is the main biological clock and synchronises the clock system present in each cell (Jin et al. 1999). The biological clock is a core group of genes that form a system of transcription and translation feedback loops, creating an internal biochemical and electric cycle with a period of around 24 hours (Czeisler et al. 1999; Ko & Takahashi 2006). While these cycles are endogenous and self-sustaining, they can be influenced by environmental factors known as *zeitgebers,* such as light, sound or social cues (Wimpory et al. 2002; Wittmann et al. 2006; Roenneberg et al. 2003). The master oscillator is synchronised to the night/day cycle of the earth's cycle around the sun by light.

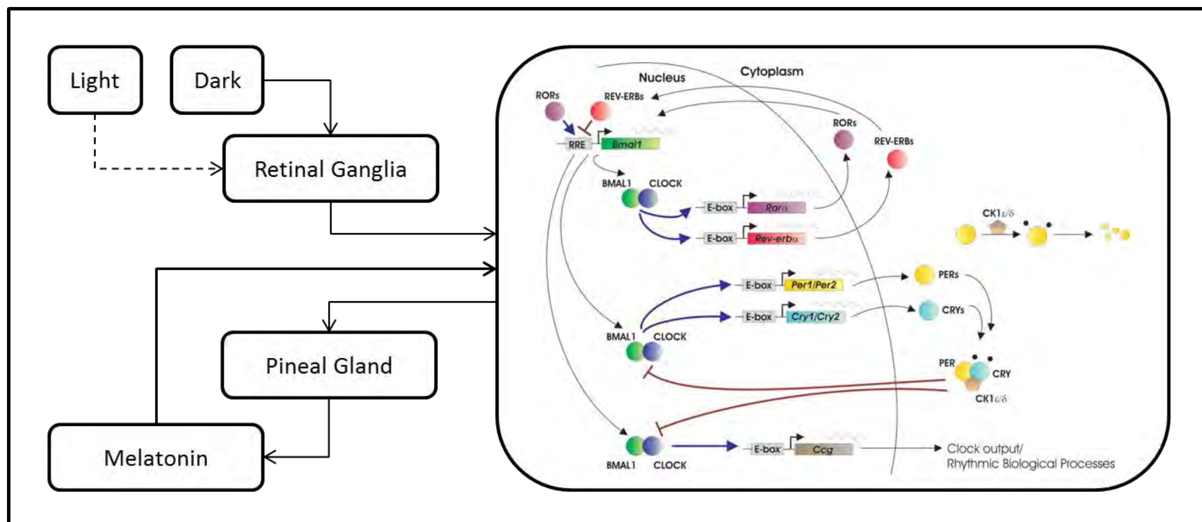

Figure 1.1 Circadian Cycle and Regulation of Melatonin: In mammals, the SCN houses the master oscillator, which governs all other oscillators in the body. The oscillator is comprised of several transcription/translation feedback loops that generate rhythmicity. This master oscillator is independent and self-sustaining, but can be reset by light. Light signals detected by the retinal ganglia of the eye cause a signal to be transduced to the SCN via the retinohypothalamic tract, indicated by arrows. The SCN signals the paraventricular nucleus, which stimulates the pineal gland. In the absence of light, melatonin is synthesised in the pineal gland. Melatonin thus acts as a signal for darkness for oscillators in the rest of the body, as well as on the SCN itself. Image of transcription/translation feedback loop taken from Ko & Takahashi (2006).

Light signals are detected by the rod and cone cells in the eye, as well as the retinal ganglia. These structures transduce the light signal to the SCN via the retinohypothalamic tract (see Figure 1.1). The SCN stimulates a further neurological signalling cascade, which ultimately stimulates the pineal gland through the paraventricular nucleus (Ko & Takahashi 2006). In the pineal gland, the melatonin biosynthetic pathway starts with the dietary amino acid L-tryptophan, which is converted to the hormone serotonin in two enzymatic steps (Figure 1.2). Only in the absence of light, the enzyme aryl-alkylamine N-acetyltransferase (A-ANAT) is transcribed and activated. This enzyme acetylases serotonin to N-acetylserotonin. The final enzyme, acetyl serotonin methyl transferase (ASMT), converts N-acetylserotonin into melatonin (Chaste et al. 2010; Macchi & Bruce 2004; Ko & Takahashi 2006). Because the expression of A-ANAT only occurs in darkness, melatonin acts as a biochemical cue for darkness from the central oscillator to oscillators in peripheral tissues (Macchi & Bruce 2004). Once melatonin has been synthesised, it is secreted into the bloodstream, and then passes to the cerebrospinal fluid through the choroid plexus (Pandi-Perumal et al. 2008). Melatonin is ultimately broken down by the enzyme cytochrome P450 1A2 (CYP1A2) in the liver (Arendt 1998 and sources within).

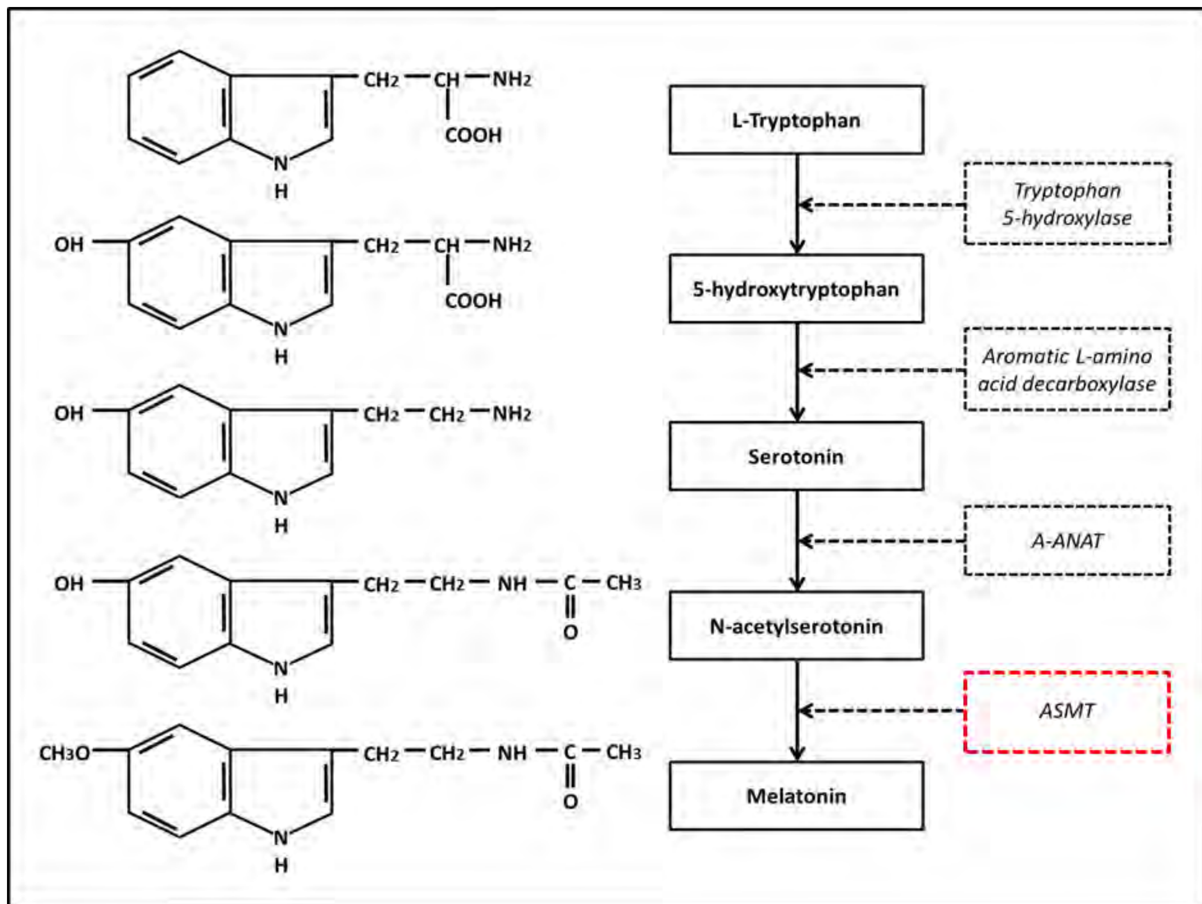

Figure 1.2 The melatonin biosynthetic pathway: The dietary amino acid L-tryptophan is taken up into the pineal gland. Here it is converted to another amino acid, 5-hydroxytryptophan, by the enzyme tryptophan 5-hydroxylase. The hormone serotonin is synthesised from 5-hydroxytryptophan through the action of the enzyme aromatic L-amino acid decarboxylase. With the onset of darkness, the enzyme aryl-alkylamine N-acetyltransferase (A-ANAT) is transcribed and activated in the pineal gland, where it acetylises serotonin. The product of this enzyme, N-acetylserotonin, is converted to melatonin by the enzyme acetylserotonin methyl transferase (ASMT) (Macchi & Bruce 2004).

Several studies have found associations between melatonin pathway polymorphisms and ASD (Jonsson et al. 2014; Melke et al. 2008; Toma et al. 2007). As melatonin is the hormone that synchronises the clock system in peripheral tissue to the main oscillator in the SCN, it has been suggested that melatonin may play a role in circadian disruption in ASD (Veatch, Pendergast, et al. 2015; Pagan et al. 2014). In addition to genetic polymorphisms and differential regulation, some studies have also found decreased melatonin levels, or decreased ASMT transcript levels (Melke et al. 2008; Hu et al. 2009; Malow et al. 2006; Goldman et al. 2014).

The role of melatonin in the sleep cycle, and sleep disruption, is well-characterised (Hardeland 2012). It is important, however, to note that sleep disruption or circadian

rhythm sleep disorders can take on several different forms, with different molecular underpinnings (Pandi-Perumal et al. 2008). Exogenous factors can influence melatonin secretion and timing, without any intrinsic polymorphism existing in the melatonin pathway or circadian clock. The timing or spectrum of light can induce melatonin suppression in a way that desynchronises melatonin expression with the natural light-dark cycle. This is often the case in shift work or jet lag, but can also be the effect of artificial light from electronic devices such as mobile phones or computer screens (Kojo et al. 2005; Wittmann et al. 2006; Cajochen et al. 2011). These factors can result in a delay in sleep onset. If this delay is chronic, it is referred to as Delayed Sleep Phase Syndrome (DSPS). Individuals with delayed sleep onset generally have normal sleep patterns and sleep quality, barring the inability to fall asleep at the desired time (Pandi-Perumal et al. 2008). This kind of sleep disruption can result from intrinsic factors, such as light sensitivity, or delayed melatonin secretion patterns in conjunction with other circadian disruptions (Shibui et al. 2003). Similarly, Advanced Sleep Phase Syndrome (ASPS) is characterised by persistent early sleep onset, as well as early awakenings. Individuals with this disorder also have a decrease in sleep quality. ASPS has been found to be associated with age-related earlier melatonin synthesis, as well as familial polymorphisms in clock genes such as Per2 (Pandi-Perumal et al. 2008).

Melatonin deficiency sometimes occurs in blind individuals, and results in a free-running sleep cycle, where the main oscillator is non-synchronised to the natural light/dark cycle. However, in some individuals, melatonin deficiency results in non-24 hour sleep cycles, rather than advances or delays in sleep cycle (Hardeland 2012). The most obvious cause for melatonin deficiency would be disruption in the translation or activation of ASMT and A-ANAT, or deficiency in tryptophan or serotonin (Hardeland 2012). Polymorphisms in the two melatonin receptors may also disrupt melatonin signalling (Pandi-Perumal et al. 2008).

1.3 Sleep Disruption in Autism Spectrum Disorders

Sleep disorders are very common in individuals with ASD (Veatch, Goldman, et al. 2015) and range from sleep onset delay, phase shift and sleep disruption (Malow et al. 2006). These parasomnias can have a significant impact on the quality of life of individuals with ASD and their families, and sleep deprivation exacerbates the behavioural deficits of ASD (Malow et al. 2006; Veatch, Goldman, et al. 2015). To alleviate these parasomnias, health practitioners

often prescribe melatonin supplementation to individuals with ASD. Studies on adults with endogenous sleep disruption, without psychiatric, social or developmental disorders, found that some individuals may benefit from exogenous melatonin administered at night (Brzezinski et al. 2005). Likewise, smaller studies have concluded that it's possible to entrain the sleep cycle of blind people with free-running sleep phase (Pandi-Perumal et al. 2008; Hardeland 2012). A meta-analysis of melatonin therapy on adults with sleep disorders that accompany shift work or jet lag, concluded that melatonin supplementation was not effective in these cases (Buscemi et al. 2006). Some studies have shown that individuals with ASD who experience sleep disruption benefit from receiving melatonin supplements at night, particularly in inducing sleep onset, with less effect in early or nocturnal waking (Jan & Freeman 2004; Rossignol & Frye 2011; Malow et al. 2012; Wright et al. 2011). A trial examining a cohort of children with neurodevelopmental disorders, including some children with autism, concluded that melatonin supplementation did not improve sleep efficiency, or early waking (Gringras et al. 2012). This finding is consistent with the role of melatonin in the circadian system, as previously discussed. A potential confounding factor when comparing results among these studies, and within the meta-analyses, is that there is little standardisation in the dose prescribed, the time of administration, or the form in which the melatonin was administered. It is also not stated what, if any, environmental factors such as nocturnal light or consistent bed-time were addressed, or whether the study participants had a melatonin deficiency.

A point to consider, however, is that melatonin is involved in several cellular processes in addition to its role in the sleep cycle (Hardeland et al. 2011). A study by Veatch *et al* (2015) found that even in ASD children with endogenous melatonin levels comparable to that of typically-developing children, responded to melatonin supplementation, and fell asleep faster. Based on this finding, they suggested that melatonin may have additional effects, not limited to its action as a chronobiotic, such as anxiolytic, hypnotic or sedative effects (Veatch, Goldman, et al. 2015; Yousaf et al. 2010; Campino et al. 2011).

1.4 Acetylserotonin Methyl Transferase: A Candidate Gene In Autism

Given that A-ANAT is the rate-limiting enzyme in melatonin synthesis, it was a logical putative gene in autism circadian research. However, results from A-ANAT studies have not concretely established this gene as a key component of ASD aetiology (Hu et al. 2009; Hu & Steinberg 2009). Furthermore, Melke et al (2008) reported elevated levels of n-acetylserotonin in conjunction with low levels of melatonin in individuals with ASD. Because n-acetylserotonin is the product of the action of A-ANAT, this indicates that the A-ANAT enzyme is functional. Therefore the levels of n-acetylserotonin in conjunction with low melatonin levels suggests that the melatonin deficiency in ASD could be the result of low enzymatic activity of ASMT (Melke et al. 2008; Cotton & Richdale 2010). Previous studies examining genetic variation of the ASMT gene found associations of polymorphisms with a myriad of psychiatric and circadian disturbances. These disorders include allelic imbalance, circadian disruption, acute myocardial infarction, intellectual impairment, Bipolar Mood Disorder, Major Depressive Disorder, and ASD (Jonsson et al. 2014; Wang et al. 2013; Melke et al. 2008; Etain et al. 2012; Pagan et al. 2011; Pagan et al. 2014; Pereira et al. 2013; Toma et al. 2007).

The ASMT gene is one of the 12 genes affected by deletions in the protease-activated receptor (PAR1) in ASD (Melke et al. 2008). Several single-nucleotide polymorphisms (SNPs) and splice site mutations in ASMT have been observed in both autism participants and participants with clinical depression (Gałecki et al. 2010; Etain et al. 2012). The ASMT gene is located on the pseudoautosomal region of the X and Y chromosomes GrCh38.p2 position 1595455-1643081 (NCBI). The gene is approximately 35kb in length, with several introns and exons (NCBI). In humans, the ASMT gene can be alternatively spliced to produce three different isoforms, all of which are involved in melatonin biosynthesis (NCBI).

1.5 Molecular Study Design

When integrating past studies on the ASMT gene (Jonsson et al. 2014; Wang et al. 2013; Melke et al. 2008; Pagan et al. 2014; Pereira et al. 2013; Toma et al. 2007), the polymorphisms associated with ASD clustered in a few regions of the gene. This study

focussed on two of these “cluster” regions: the region spanning from GRCh38.p2 position X:1614846-1615299; and the region spanning GRCh38.p2 position X:1629573-1630120.

The first region, X:1614846-1615299, contains the promoter and 5’-untranslated region (UTR) of the ASMT gene. Previous studies have identified an ASD “risk” haplotype in this region (Melke et al. 2008; Botros et al. 2013), which consists of the SNPs rs4446909, rs5989681, and rs6644635. However, other studies did not significantly implicate this haplotype in ASD aetiology (Toma et al. 2007; Wang et al. 2013; Veatch, Pendergast, et al. 2015). It is noteworthy that many of these studies do not examine specific traits within the ASD phenotype, for example sleep disruption (Veatch, Pendergast, et al. 2015), which could mean that the data is too heterogeneous to highlight specific genes or polymorphisms. Furthermore, it is possible that regional genetic variants within these cohorts confound comparison between the international cohorts. We selected this region to investigate the presence of the above mentioned risk haplotype, as well as whether regional genetic variants were present in a South African population. The second region spans the GrCh38.p2 position X: 1629573-1630120. This covers exon 5 of ASMT, and part of the intronic region on both the 5’ and 3’ end of exon 5. This region has not been as well-studied as the promoter region, but SNPs in this region were associated ASD (Wang et al. 2013; Pereira et al. 2013; Holt et al. 2010)

1.6 ASD Endophenotyping

In addition to the published clinical uses of ADOS (Autism Diagnostic Observation Schedule) assessments as a diagnostic tool, this study used ADOS assessment criteria as a phenotyping tool. The ADOS tool assigns numerical scores to behaviours observed during the assessment, in both structured and unstructured play. The behaviours are divided into three broad categories: Communication (for example, reporting of events, and use of gestures), Reciprocal Social Interaction (for example, eye contact and forming a rapport with the assessor), and Restrictive and Repetitive Behaviours (for example, stereotyped use of language, or stereotyped and repetitive movements). The sum of these scores (the Overall Total) is interpreted using different cut-offs, based on the age and verbal fluency of the individual (see Table 1.1), to establish the ADOS diagnosis/classification. Symptom Severity is established by assigning a Comparison Score, which ranges from 1 to 10, based on the

Overall Total Score, the chronological age of the participant, and the Module number with which they were assessed. A Comparison Score of 1-2 represents Minimal-to-No Evidence of spectrum-related symptoms, 3-4 represents a Low level of symptoms, 5-7 represents a Moderate level of symptoms, and 8-10 represents a High symptom severity.

1.7 The Present Study

This study will examine the genotypes of children with ASD, at the two genetic regions described in Section 1.5. The genotypes of the ASD cohort will be compared to that of typically developing age-matched controls. The goal is to establish if children on the Autism spectrum have any significant genetic changes or variants compared to typically developing children. The study will also sub-group the ASD cohort to look for genetic variability within the ASD group.

Table 1.1 Table indicating the cut-off scores for diagnosis/classification of ASD using the ADOS assessment, taking into account the verbal fluency and age-appropriate linguistic development of the individual.

ADOS Module	ADOS Classification	Developmental Marker	Overall Total Score
Module 1	Autism	Few to no words	16 or higher
		Some words	12 or higher
	Autism Spectrum	Few to no words	11 to 15
		Some words	8 to 11
	Non-Spectrum	Few to no words	10 or lower
		Some words	7 or lower
Module 2	Autism	Younger than 5 years	10 or higher
		Age 5 or older Simple phase speech	9 or higher
	Autism Spectrum	Younger than 5 years	7 to 9
		Age 5 or older Simple phase speech	Equal to 8
	Non-Spectrum	Younger than 5 years	6 or lower
		Age 5 or older Simple phase speech	7 or lower
Module 3	Autism	Fluent Speech	9 or higher
	Autism Spectrum	Fluent Speech	7 or 8
	Non-Spectrum	Fluent Speech	6 or lower

2 MATERIALS AND METHODS

2.1 ETHICS CLEARANCE

Ethics clearance was granted by the University of Cape Town, Faculty of Science Research Ethics Committee, **FSREC 076-2014**. Approval to approach schools was granted by Western Cape Research Services, **20141002-37506**.

2.2 RESEARCH PARTICIPANTS

Children with ASD were recruited from two schools for children with developmental or social disabilities, while typically developing children (controls) were recruited from mainstream primary schools and sports clubs in Cape Town, South Africa. After informed parental consent (sample of consent form Supplementary Figure 6.1), children were recruited to participate in in the study. This study forms part of a larger ASD genomic and epigenetic research programme, so the cohort is of a specific age range with age-matched controls. Participants were aged from 5 to 14 years old and were unrelated males from African (BA), European (WA) and Mixed ancestry (MA). All participants, including controls, were screened using the Autism Diagnostic Observation Schedule-2 (ADOS-2) (Lord et al. 1989) by certified, research reliable ADOS administrators.

Participants in the ASD group were required to have a prior diagnosis of ASD, as the ADOS assessments in this study were used for categorisation and phenotyping, not as part of a clinical diagnosis. Participants with known chromosomal disorders, such as Neurofibromatosis or Fragile X syndrome were excluded from the final data set. Typically developing children were only considered if they had no social or learning disability diagnoses, and were excluded from the final data set if they scored above zero on the ADOS final algorithm. The study recruited 50 children with previous diagnoses of ASD and 7 typically developing children as controls. Of the recruited children with ASD, 32 participants allowed DNA to be sampled using buccal swabs (see Table 2.1). Due to the exclusion criteria, the Final Study Sample consisted of 28 children with ASDs, and 6 controls; all with a stringent, “clean” phenotype. After informed parental consent had been obtained in writing, the research participants were assessed using the ADOS-2, and if they met the study criteria,

a genetic sample was obtained using buccal swabs. Noteworthy is that this study is part of a programme that is building an ASD cohort *de novo*, hence the limited sample size.

Table 2.1 Table indicating the selection criteria for the cohort used in this study, as well as the demographic breakdown of each phenotype, and the number of participants in each subgroup.

Grouping	Participants	Ancestry	Criteria
Whole Cohort	**All ASD** (*n*=32)	BA= 3 MA= 15 WA= 10	Prior ASD diagnosis; ADOS classification " Autism or Spectrum"
	Controls (*n*=7)	WA= 7	ADOS classification "Non-Spectrum"
Final Study Sample	**Autism** (*n*=28)	BA= 3 MA= 15 WA= 10	Prior ASD diagnosis; ADOS classification "Autism or Spectrum"; no known chromosomal comorbidities
	Controls (*n*=6)	WA= 6	ADOS classification "Non-spectrum"; and ADOS Severity "Minimal/No evidence"; No chromosomal comorbidities

BA= Black ancestry; MA= Mixed ancestry; WA= White ancestry

2.3 SAMPLES AND NUCLEIC ACID EXTRACTION

Buccal swabs were selected for this study because they are non-invasive and do not carry the same health risks as working with blood. Is has also been proposed that buccal cells are a better proxy for brain tissue than blood, given their developmental lineage, and that blood is a mixed cell population (Smith et al. 2015). There is also evidence that melatonin is produced in human salivary glands (Shimozuma et al. 2011).

Buccal cells were obtained using Isohelix DNA swabs (Cell Projects Ltd., Kent, UK). The swabs were introduced into the mouths of the participants, and rubbed in a firm circular motion on each inner cheek for at least 30 seconds per cheek. Thereafter the swabs were placed in lysis buffer (Ethylenediaminetetraacetic acid (EDTA), pH 8; 10 % (w/v) sodium dodecyl sulphate (SDS)) at 4°C, for up to a maximum of two weeks before extraction. The genomic DNA extraction method was based on a published protocol, but with some modifications (Aljanabi & Martinez 1997). In short, the swabs were allowed to return to room temperature, for at least one hour prior to incubation, with the addition of 10 mg ml-1 proteinase K (Thermo Scientific) for 1.5 hours. This was followed by the addition of

saturated (6M) NaCl to precipitate cell debris, and centrifuged at 12000 xg for 10 minutes. The resulting supernatant was aspirated off, and added to pure isopropanol, and kept at -8°C overnight. The DNA was pelleted by a further two centrifugation and ethanol wash steps. The extracted genomic DNA is stored at -20°C. The DNA was quantified using a Nanodrop ND- 1000 spectrophotometer (ThermoScientific).

2.4 Polymerase Chain Reaction

PCR of GRCh38.p2 position X:1614846-1615299

Each participant was genotyped for SNPs or other polymorphisms in two ASMT gene regions, namely the upstream promoter region, and the exon/intron splice junction at NC_000023.11 (16300kbp).

A 454 bp fragment of the promoter region for ASMT was amplified using ThermoScientific Phusion High-Fidelity DNA Polymerase (#F-530S) using the forward primer 5'-ACCTGCTCAATCCATAAGACGA-3' and the reverse primer 5'-ATTCCCCCTTGTCCCACAGA-3' designed in this study using Primer BLAST and OligoAnalyzer. PCR reagent concentrations and cycling conditions as per Supplementary Table 6.1 and Supplementary Table 6.2. This region contains the SNPs rs4446909, rs5989681, rs56690322, rs6644635, and rs17149149, previously associated with ASD (Melke et al. 2008; Wang et al. 2013; Toma et al. 2007; Veatch, Pendergast, et al. 2015). The success of the PCR, and the PCR product size, was confirmed by electrophoresing the PCR product on a 1.2 % (w/v) agarose gel in 1 X tris-acetate with EDTA (TAE) buffer stained with ethidium bromide.

PCR of GRCh38.p2 position X:1629573-1630120

A 548 bp fragment of the exon/intron boundary region for ASMT was amplified using ThermoScientific Phusion High-Fidelity DNA Polymerase (#F-530S) using the forward primer 5'- TCCCTTCCCATCCAGAGAAC-3' and the reverse primer 5'- GCAGAGGAGATGTTTGCTGA-3' designed in this study using Primer BLAST and OligoAnalyzer. PCR reagent concentrations and cycling conditions as per Supplementary Table 6.1 and Supplementary Table 6.2. This region contains the SNPs rs6588807, rs144935309, rs147969184, rs141937160, rs145804175, rs145494220, and rs28675287 previously associated with ASD (Holt et al.

2010; Wang et al. 2013). The success of the PCR, and the PCR product size, was confirmed by electrophoresing the PCR product on a 1.2 % (w/v) agarose gel in 1 X TAE buffer stained with ethidium bromide.

2.5 SEQUENCING

Sequencing of the two genomic PCR fragments of ASMT was completed by the Centre for Analytical Services (CAF) at the University of Stellenbosch, South Africa. CAF used the BigDye Terminator V3.1 sequencing kit (Applied Biosystems) in an Applied Biosystems 3730xl DNA Analyzer, using the manufacturers protocol with slight modifications. CAF performed data collection and sequence analysis using the Applied Biosystems DATA Collection v3.0 and Applied Biosystems Sequence Analysis v5.3.

2.6 SINGLE NUCLEOTIDE POLYMORPHISM GENOTYPING

The resultant sequences were edited and aligned using Chromas v2.4 (Technelysium Pty Ltd) and Mega 6.06 (Build no.: 6140226). The sequences were aligned with the *Homo sapiens* chromosome X, GRCh38.p2 Primary Assembly (NCBI Reference Sequence: NC_000023.11) Add download date from NCBI Nucleotide. The forward and reverse sequences for each individual overlapped, therefore SNPs were only confirmed if they were present in both reads.

2.7 STATISTICAL ANALYSIS

Statistical analysis on both the genotype and allele quantitative data was performed using RStudio (Version 0.99.489 – © 2009-2015 RStudio, Inc.). Given the sample sizes, a two-tailed Fisher's Exact test was used to compare subsets of the data grouped based on phenotype (using the fisher.test function in RStudio). Hardy-Weinberg exact tests (GenePop, web version 4.2, http://genepop.curtin.edu.au/genepop_op1.html) were used to test for Hardy-Weinberg equilibrium. Haplotyping of the genetic data was done using Haploview (Barrett et al. 2005). SNP functional implications were assessed using Ensembl Variant Effect Predictor. All p-values were reported using the standard $p<0.05$ significance for biological data.

2.8 PHENOTYPIC GROUPING

The first line of comparison was to compare the entire ASD cohort based on the ADOS module with which they were assessed. Given that the choice of ADOS module is selected based on the verbal fluency of the child, the module number was used as a proxy for language development. Module 1 is used for participants with no language/limited single word usage, Module 2 for simple phrase speech and Module 3 for fluent speech. The second comparison used on the entire cohort was based on ancestry, to establish whether there is differential SNP frequency based on population stratification. However, this is a poor approximation of the whole South African population, as discussed later. Comparisons between sub-phenotypes within the ASD cohort and the controls were based on the ADOS Classification and Symptom Severity. ADOS Classification falls into three categories (Autism, Autism Spectrum, and Non-Spectrum), and Symptom Severity falls into four categories (High, Moderate, Low, and Minimal-to-No Evidence).

3 RESULTS

3.1 CHARACTERISTICS OF PARTICIPANTS IN THE STUDY COHORT

A total of 50 children with a prior ASD diagnosis were screened using the ADOS-2 assessment to obtain a research cohort of 32 children with ASD and seven children as controls. All study participants were male, aged 5 to 14 years old. Genomic DNA was sequenced for target regions in ASMT and genotyping of both gene regions was successful for *n*=36 individuals, with the remaining three individuals being genotyped for only one of the two gene regions. However, for the sake of a "clean" phenotype, all individuals with medical comorbidities were excluded from the final study sample (Table 2.1). One individual was excluded from the control group based on their ADOS score being above the agreed cut-off of a total ADOS score over zero. This stringent cut-off was selected to ensure a clean phenotype of true control typically developing candidates. Thus the final study sample consisted of 28 ASD cases and six controls, with a total of 32 individual sequences for both target regions, and two partial sequences.

Within a South African context, ethnicity is difficult to establish, as the general population is made up of multiple nationalities and heritages. Thus participants were divided into three broad groups based on self-reported ancestry: Black Ancestry (BA), Mixed Ancestry (MA), and White Ancestry (WA) (Table 3.1). The BA group was underrepresented in both the ASD and typically developing control cohorts, predominantly due to the demographic range in the two schools. This is partly a result of the fact that administrators for this study are English speakers and had been accredited to administer the assessment in English only. A further limitation is that the typically developing control group comprised entirely of individuals of WA. Hardy-Weinberg testing of both ASD and control groups showed that both populations were in Hardy-Weinberg equilibrium for all loci (Supplementary Table 6.7 and Supplementary Table 6.8).

3.2 ENDOPHENOTYPES

Within the final study sample, the following ADOS Symptom Severity scores were observed: High, n=5, Moderate, n= 20 and Low, n= 3. The six individuals in the control group were categorised as having Minimal/No Evidence of ASD symptoms.Participants were screened using an ADOS module appropriate for their spontaneous, expressive language. This study used Modules 1 to 3 for assessment of participants because of the restricted age range used; the oldest participants were two 14 year old boys with ASD who were both pre-verbal. Module 1 and 2 were used for the majority of the ASD participants, and Module 3 was used for all controls as well as four ASD participants (see Table 3.1). For a detailed breakdown of all participants' demographic data, see Supplementary Table 6.3.

Table 3.1 Table indicating the selection criteria for the endophenotypes used in this study, as well as the demographic breakdown of each endophenotype, and the number of participants in each subgroup. Symptom Severity is based on the overall ADOS score of a participant, their age, and the Module with which they were assessed. The ADOS Module choice is based on the verbal fluency of a participant: Module 1 is used for children with no speech or limited, single words; Module 2 is used for children who use simple, meaningful phrase speech; and Module 3 is used for children with verbal fluency

Grouping	Endophenotypes	Ancestry	Criteria
Symptom Severity	**ASD High** (*n*=5)	BA=1 MA=2 WA=2	ADOS classification "Autism" or "ASD"; ADOS symptom severity "High"' no comorbidities;
	ASD Moderate (*n*=20)	BA=2 MA=11 WA= 7	ADOS classification "Autism" or "ASD"; ADOS symptom severity "Moderate"; no comorbidities;
	ASD Low (*n*=3)	MA=2 WA=1	ADOS classification "Autism" or "ASD"; ADOS symptom severity "Low": no comorbidities;
	Controls (*n*=6)	WA=6	ADOS classification "Non-spectrum"; ADOS symptom Severity "Minimal/No evidence"; no comorbidities;
ADOS Module	**Module 1** (*n*=12)	BA=2 MA=5 WA=5	ADOS Module 1 (Few to no words), ADOS classification "Autism" or "Non-spectrum": no comorbidities;
	Module 2 (*n*=9)	BA=1 MA=6 WA= 2	ADOS Module 2 (phrase speech), ADOS classification "Autism" or "Non-spectrum"; No comorbidities;
	Module 3 (*n*=10)	MA=2 WA=8*	ADOS Module 3 (fluent speech); ADOS classification "Autism" or "Non-spectrum"; No comorbidities;

BA= Black Ancestry; MA= Mixed Ancestry; WA= White Ancestry * includes the 6 controls

3.3 Putative Single Nucleotide Polymorphisms Associated with Autism

The SNPs identified in this study are indicated in Table 3.2. This study identified two novel SNPs: IVS4-225C>A and IVS5+124G>T. Table 3.3 indicates the population allele frequencies for the SNPs from the 1000 Genomes Project, where this data is available, in addition to the allele frequencies for these SNPs in the three broad population groups in this current study. The SNP allele and genotypic counts were tested for significance by comparing the different endophenotypes using Fisher's Exact Test for count data. Table 3.4 shows the pairwise comparisons that are significant (*p*-value <0.05), as well as comparisons with *p*-values of between 0.05 and 0.1. While the comparisons are statistically underpowered, the higher *p*-values are included for consideration, as they may indicate a trend towards significance. For this study, correction for multiple testing was not used due to the small sample size.

Using the ADOS module to differentiate between ASD phenotypes, SNP rs148520524 differed in frequency in the Module 1 versus Module 2 comparison of the ASD study participants. These participants were between five and fourteen years of age at the time of assessment, and had severely limited verbal development in relation to their chronological age expectations. The SNP rs6588807 differed in frequency between ASD endophenotypes with High ADOS Symptom Severity Scores compared to ASD participants with Moderate ADOS Symptom Severity Scores. There were three other notable SNPs, of which rs5989681 and rs56690322 significantly differ in frequency between participants of Black Ancestry compared to participants of White or Mixed Ancestry. Both rs5989681 and rs56690322 have previously been associated with ASD (Melke et al. 2008). However, this association with ancestry should be treated with caution because the BA participants are clearly not well represented in the study sample (*n*=3). The overall small sample size, combined with a small control group of WA-only participants, further emphasises that the allele frequencies seen in this study may not be a true representation of the genetic variability in South Africa. The last SNP rs186757787 also occurred significantly more frequently in participants of BA than those of MA (p=0.0207) or WA (p=0.0028). This SNP has no prior associations with ASD in the literature.

Table 3.2 Table indicating the genotype frequencies, and counts in brackets, of the control group and the ASD cohort. SNPs indicated with (*) have previously been associated with ASD in the literature. Underlined SNPs are novel SNPs observed in this study cohort.

SNP	Genotype	Controls % (n)	ASD High % (n)	ASD Mod % (n)	ASD Low % (n)	SNP	Genotype	Controls % (n)	ASD High % (n)	ASD Mod % (n)	ASD Low % (n)
rs4446909*	GG	100 (5)	80 (4)	70 (14)	100 (3)	rs186757787	CC	100 (6)	75 (3)	90 (18)	66 (2)
	GA	-	20 (1)	25 (5)	-		CT	-	-	10 (2)	33 (1)
	AA	-	-	5 (1)	-		TT	-	25 (1)	-	-
rs148520524	CC	100 (5)	100 (5)	85 (17)	100 (3)	rs148855515	TT	100 (6)	100 (4)	100 (20)	100 (3)
	TC	-	-	15 (3)	-		CT	*	-	-	-
	TT	-	-	-	-		CC	-	-	-	-
rs5989681*	GG	80 (4)	40 (2)	40 (8)	33 (1)	rs184583293	GG	100 (6)	100 (4)	95 (19)	100 (3)
	GC	20 (1)	40 (2)	55 (11)	66 (2)		GA	-	-	5 (1)	-
	CC	-	20 (1)	5 (1)	-		AA	-	-	-	-
rs56690322*	GG	80 (4)	60 (3)	70 (14)	66 (2)	rs28675287*	TT	83 (5)	100 (4)	70 (14)	66 (2)
	GA	20 (1)	40 (2)	30 (6)	33 (1)		CT	17 (1)	-	30 (6)	33 (1)
	AA	-	-	-	-		TT	-	-	-	-
rs6644635*	CC	40 (2)	40 (2)	25 (5)	33 (1)	rs191866206	GG	100 (6)	100 (4)	95 (19)	100 (3)
	TC	60 (3)	40 (2)	60 (12)	66 (2)		GA	-	-	5 (1)	-
	TT	-	20 (1)	15 (3)	-		AA	-	-	-	-
rs774990880	GG	100 (5)	100 (5)	95 (19)	100 (3)	IVS5+124G>T	GG	83 (5)	100 (4)	100 (20)	100 (3)
	GC	-	-	5 (1)	-		GT	17 (1)	-	-	-
	CC	-	-	-	-		TT	-	-	-	-
IVS4-225C>A	CC	100 (6)	100 (4)	95 (19)	100 (3)	rs776976390	GG	100 (6)	100 (4)	100 (20)	66 (2)
	AC	-	-	5 (1)	-		GA	-	-	-	33 (1)
	AA	-	-	-	-		AA	-	-	-	-
rs6588807	AA	50 (3)	25 (1)	20 (4)	33 (1)						
	AC	50 (3)	50 (2)	70 (14)	33 (1)						
	CC	-	25 (1)	10 (2)	33 (1)						

Table 3.3 Allele frequencies in the five super population groups, from the 1000 Genomes Project (Batch ID: Phase 3_V1), and the three population groups from this current study. The sample size for each population is indicated in brackets as the number of alleles per group (i.e. 2n). The top allele is the major allele for that population, and the bottom allele is the minor allele.

	1000 Genomes Project					Current Cohort		
	Admixed American (694)	**African (1322)**	**East Asian (1008)**	**South Asian (978)**	**European (1006)**	**BA (6)**	**MA (30)**	**WA (30)**
rs4446909	G=0.8285 A=0.1715	G=0.9939 A=0.0061	G=0.7103 A=0.2897	G=0.8303 A=0.1697	G=0.7147 A=0.2853	G=1.0000 A=0.0000	G=0.8333 A=0.1667	G=0.8667 A=0.1333
rs148520524	C=1.0000 T=0.0000	C=0.9682 T=0.0318	C=1.0000 T=0.0000	C=1.0000 T=0.0000	C=1.0000 T=0.0000	C=0.8333 T=0.1667	C=0.9667 T=0.0333	C=0.9667 T=0.0333
rs5989681	G=0.6628 C=0.3372	C=0.6641 G=0.3359	G=0.6012 C=0.3988	G=0.7761 C=0.2239	G=0.6978 C=0.3022	G=0.1667 C=0.8333	G=0.7667 C=0.2333	G=0.7667 C=0.2333
rs56690322	G=0.9236 A=0.0764	G=0.8267 A=0.1732	G=0.9831 A=0.0169	G=0.8599 A=0.1401	G=0.9026 A=0.0974	G=1.0000 A=0.0000	G=0.8000 A=0.2000	G=0.9000 A=0.1000
rs6644635	C=0.6124 T=0.3876	C=0.8124 T=0.1876	C=0.7639 T=0.2361	C=0.6207 T=0.3793	C=0.6531 T=0.3469	C=0.8333 T=0.1667	C=0.5667 T=0.4333	C=0.5667 T=0.4333
rs6588807	A=0.6398 C=0.3602	A=0.5514 C=0.4486	A=0.6290 C=0.3710	A=0.6881 C=0.3119	A=0.6451 C=0.3549	A=0.5000 C=0.5000	A=0.6000 C=0.4000	A=0.6333 C=0.3667
rs186757787	C=0.9899 T=0.0101	C=0.9085 T=0.0915	C=1.0000 T=0.0000	C=1.0000 T=0.0000	C=1.0000 T=0.0000	C=0.5000 T=0.5000	C=0.9667 T=0.0333	C=1.0000 T=0.0000
rs18458329	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.9960 A=0.0040	G=0.9980 A=0.0020	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.9667 A=0.0333	G=1.0000 A=0.0000
rs28675287	T=0.8256 C=0.1744	T=0.9236 C=0.0764	T=0.5784 C=0.4216	T=0.8211 C=0.1789	T=0.8439 C=0.1561	T=0.8333 C=0.1667	T=0.8667 C=0.1333	T=0.9000 C=0.1000
rs191866206	G=0.9986 A=0.0014	G=0.9985 A=0.0015	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=1.0000 A=0.0000	G=0.9667 A=0.0333	G=1.0000 A=0.0000

BA= Black Ancestry, MA= Mixed Ancestry, WA= White Ancestry

Table 3.4 SNPs found to differ in frequency between endophenotypes within the ASD cohort and controls. Significance is set at $p<0.05$, but p-values <0.1 are also shown. ([1]) indicates endophenotypes where the genetic data for cases and controls was pooled, (*) indicates SNPs that have been linked to ASD in previous studies (Melke et al. 2008; Veatch, Pendergast, et al. 2015)

Subgroup	rs148520524 *p-value*	rs5989681* *p-value*	rs56690322* *p-value*	rs6588807* *p-value*	rs186757787 *p-value*
Module:					
1 vs. 2	0.0711	-	-	-	-
Severity:					
High vs. Moderate	-	-	-	0.0681	-
Ancestry[1]:					
BA vs. MA	-	0.0123	-	-	0.0207
BA vs. WA	-	0.0102	-	-	0.0028
WA vs. MA	-	-	0.0440	-	-

3.4 Haplotypes

The Final Study Sample, consisting of all individuals in the dataset that are free of comorbidities, was examined for linkage disequilibrium between any of the SNPs found in this study. SNPs rs5989681, rs56690322 (P1BC) and rs6644635 are in linkage disequilibrium, which partially corresponds to the 'risk' haplotype identified by Melke *et al* (2008). Melke *et al* (2008) found the haplotype GGGC of SNPs rs4446909, rs5989681, rs56690322 (P1BC) and rs6644635 to be significantly associated with ASD. However, in this study, rs4446909 does not seem to be in significant linkage disequilibrium with rs5989681, rs56690322 (P1BC) and rs6644635 (see Figure 3.1a). The SNP rs186757787, however, displays linkage disequilibrium with rs5989681, despite its location 14kb away from the promoter region (Figure 3.1a). Of these SNPs, all but rs6588807 are significantly associated with ancestry (Table 3.4). These four SNPs are indicated in the figure by a haplotype block.

When comparing the count data of rs5989681 (G>C), rs56690322 (G>A), rs6644635 (C>T), and rs186757787 (C>T) together (see Table 3.5) using Fisher's Exact Test, the haplotype GGCC is more frequent (p-value of 0.0566) in the controls than in the ASD cohort. It is worth noting that the minor alleles of these SNPs do not occur significantly more frequently individually in the ASD cohort than in the controls (Supplementary Figure 6.4). It can also be seen in Table 3.4 that the SNP rs5989681 is significantly more common in the BA group than

in the MA (p-value=0.0123) or the WA (p-value=0.0102) groups. Likewise, rs56690322 is significantly more common in the BA group than in the MA (p-value=0.0207) or WA (p-value=0.0028) groups. As stated before, the ancestry data should be treated with caution, however, due to our sample size that is not representative of the whole South African population.

Table 3.5 Table indicating the haplotype counts in the ASD group vs. the controls, along with the p-values derived using Fisher's Exact test. Significance is set at $p<0.05$. The SNPs are rs5989681 (G>C), rs56690322 (G>A), rs6644635 (C>T), and rs186757787 (C>T)

Haplotype	ASD % (n=27)	Controls % (n=5)	*p*-value
CACC	11.1 (3)	-	0.5897
CACT	3.7 (1)	-	0.8438
CGCC	37.0 (10)	20 (1)	0.4280
CGCT	11.1 (3)	-	0.5897
GACC	14.8 (4)	20 (1)	0.5991
GATC	3.7 (1)	-	0.8438
GGCC	14.8 (4)	60 (3)	0.0566
GGTC	3.7 (1)	-	0.8438

ASD= Autism Spectrum Disorder participants

In an attempt to eliminate the possibility of ancestry being a confounding factor, a linkage disequilibrium map was generated with only WA and MA participants from the Final Sample set. The ideal would be to test for LD for each ancestry group individually, but the small sample size negatively impacts the power of the comparison. The results indicated that only rs56699322 and rs4446909 were in linkage disequilibrium, with insufficient data for the remaining loci (see Figure 3.1b). There was no significant association between the haplotype of these two SNPs and ASD.

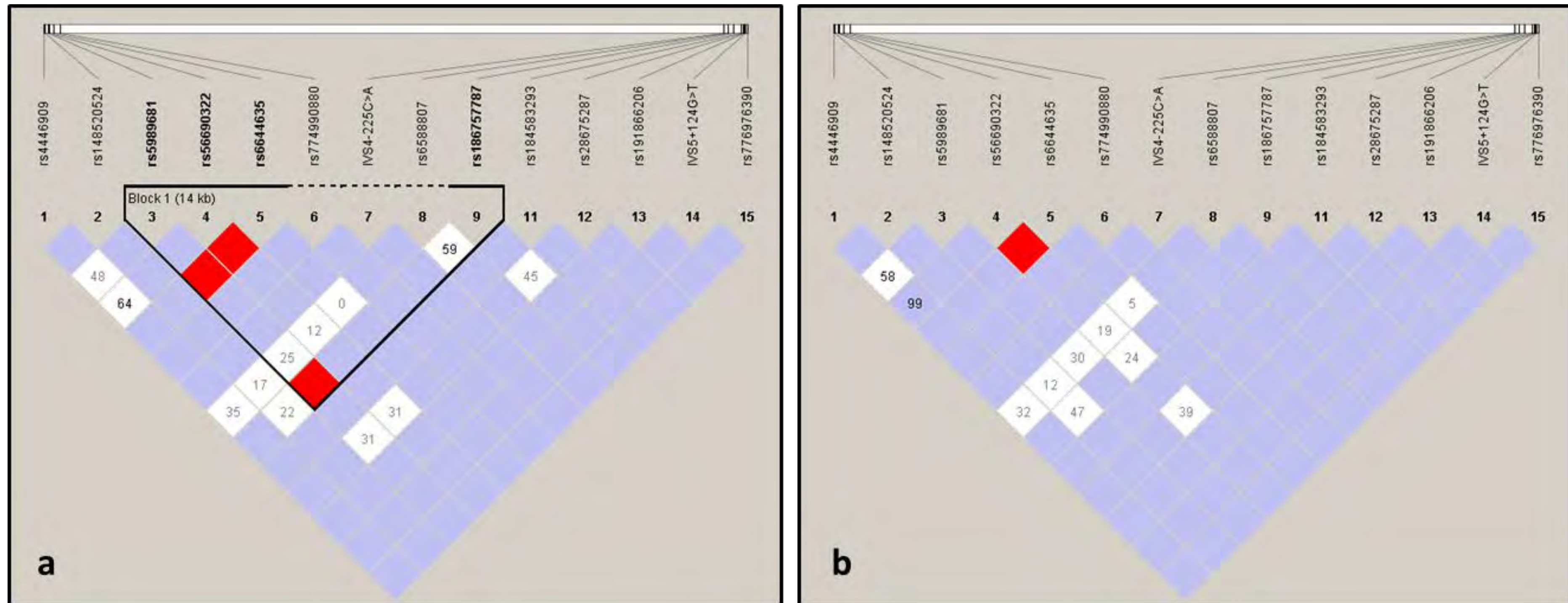

Figure 3.1 Linkage disequilibrium maps generated in Haploview using the genotypic data from participants in the ASD cohort, and the controls. Blue blocks indicate insufficient information for that locus, with the numbered blocks indicating weak linkage disequilibrium, and the red blocks indicating loci that are in strong linkage disequilibrium (Barrett et al. 2005). **(a)** Haplotype structure of ASMT promoter region and target intron/exon boundary, based on genotype data from the final research cohort (Table 2.1). The black haploblock indicated was inserted manually, to indicate the four SNPs that exhibit some LD with each other. **(b)** Haplotype structure of ASMT promoter region and target intron/exon boundary, based on the genotypes of only WA and MA study participants.

The ASD ‘risk’ haplotype frequency, as proposed by Melke *et al* (2008), was also investigated (Table 3.6). The GGGC haplotype was present significantly more frequently (*p*-value= 0.313) in the Control group than in the ASD group. This relationship held true (Supplementary Table 6.5) regardless of whether entire final sample (n= 34) was used, or whether the BA participants were excluded (n= 31).

Table 3.6 Table indicating the haplotype counts in the ASD group vs. the controls, for the ‘risk’ haplotype first identified by Melke *et al* (2008). The p-values derived using Fisher’s Exact test. Significance is set at p<0.05. The SNPs are rs4446909 (G>A), rs5989681 (G>C), rs56690322 (G>A), rs6644635 (T>C).

Haplotype	ASD % (n=27)	Controls % (n=5)	*p*-value
ACAC	7.1 (2)	-	0.7159
GCAC	7.1 (2)	-	0.7159
GGAC	14.3 (4)	20 (1)	0.5859
ACGC	14.3 (4)	-	0.5004
GCGC	32.1 (9)	20 (1)	0.5149
AGGC	3.6 (1)	-	0.8485
GGGC	10.7 (3)	60 (3)	**0.0313**
GGAT	7.1 (2)	-	0.7159
GGGT	3.6 (1)	-	0.8485

ASD= Autism Spectrum Disorder participants

3.5 Deleterious Single Nucleotide Polymorphisms

For the majority of the SNPs observed in this study, the effect of alleles on gene expression has not been characterised, if indeed the polymorphisms have any effect. This is particularly the case for the SNPs in the intronic region, where there is no prior information on the contribution of either allele to gene function or expression. Indeed, *in silico* software (e.g. SIFT and Ensembl Variant Effect Predictor) were unable to predict the effects of these intronic SNPs. However, a missense SNP and a splice donor variant were identified (Table 3.7) in addition to three SNPs in the promoter region. The SNPs, their potential effects and frequencies are summarised in Table 3.7. Analysis using Alibaba identified putative transcription factor binding sites in both regions.

Table 3.7 The position and functional implication of non-synonymous SNPs identified in this cohort. Frequency of SNPs is indicated per endophenotype. SNPs present in individuals who were excluded from the final sample are reported as "Unknown Developmental Deficit". Exclusion was due to presenting ASD with comorbidity, or presenting some undiagnosed developmental deficit without ASD. Functional implication were predicted using dbSNP and Ensembl Variant Effect Predictor on GRCh38.p2. ([1]) indicates putative transcription factor binding sites predicted using Alibaba

SNP ID (Location)	Function	Impact	Frequency per Endophenotype % (n)	
rs4446909 (X: 1614890)	promoter region	Disrupts CCAC box (Jonsson et al. 2010)	ASD Moderate: ASD High: UDD:	30 (6) 20 (1) 50 (2)
rs148520524 (X: 1614929)	promoter region	In AP-1 binding site (Jonsson et al. 2010)	ASD Moderate:	15 (3)
rs5989681 (X: 1614999)	promoter region	In NF-Kappab binding site (Jonsson et al. 2010)	ASD Low: ASD Moderate: ASD High: Controls:	66 (2) 60 (12) 60 (3) 20 (1)
rs56690322 (X: 1615027)	promoter region	Predicted SRF binding site[1]	ASD Low: ASD Moderate: ASD High: Controls: UDD:	33 (1) 30 (6) 40 (2) 20 (1) 75 (3)
rs6644635 (X: 1615101)	promoter region	Predicted MYF6 binding site[1]	ASD Low: ASD Moderate: ASD High: Controls: UDD:	100 (3) 85 (17) 80 (4) 100 (5) 100 (4)
rs774990880 (X: 1615260)	protein-coding	moderate (V [Val] ⇒ L [Leu]) (dbSNP, NCBI)	ASD Moderate:	5 (1)
rs6588807 (X: 1629688)	intron	Predicted CEBPG binding site[1]	ASD Low: ASD Moderate: ASD High: Controls: UDD:	66 (2) 80 (16) 75 (3) 50 (3) 100 (6)
rs148855515 (X: 1629941)	splice donor variant	high (dbSNP, NCBI)	UDD:	10 (1)
rs184583293 (X: 1629967)	intron	Predicted SP1 binding site[1]	ASD Moderate:	5 (1)
rs28675287 (X: 1630017)	intron	Predicted CEBP alpha binding site[1]	ASD Low: ASD Moderate: Controls: UDD:	33 (1) 30 (6) 17 (1) 50 (3)
rs776976390 (X: 1630098)	intron	Predicted CEBP alpha binding site[1]	ASD Low:	33 (1)

ASD= Autism Spectrum Disorders; UDD= Unknown Developmental Deficit.

The SNPs, rs4446909, rs148520524, and rs5989681 are predicted to be within the transcription factor (TF) binding sites of the promoter region of ASMT (Figure3.2) (Jonsson et al. 2010). Analysis using Alibaba identified two additional SNPs, rs56690322 and rs6644635, which are within putative TF binding sites.

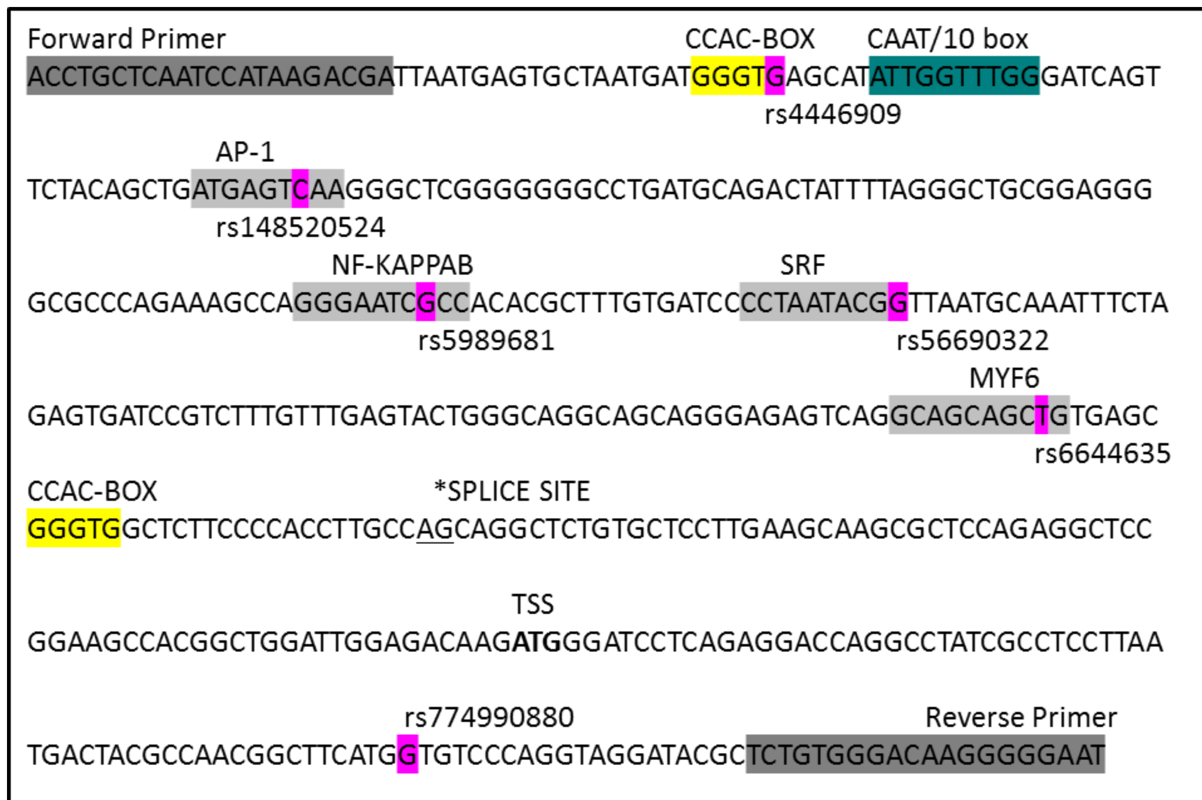

Figure 3.2 Sequence and putative transcription factor binding sites in the ASMT gene promoter region spanning X: 1614846-1615299. The primer sequences are indicated in dark grey. The SNPs found in this study are indicated in dark pink, with the binding sites in light grey. CCAC boxes and CAAT-BOX are indicated in yellow and green respectively. Binding sites were identified using Poly-Phen and SIFT analysis by Jonsson *et al* (2010). The design and layout of this image is taken from Jonsson *et al* 2010, and edited to include the primers and SNPs from this study.

The missense SNP, rs774990880, in within exon 1 of ASMT, which is included in three protein coding transcripts of ASMT (Ensembl transcript IDs: ENST00000381241, ENST00000381233, ENST00000381229). The clinical significance of this amino acid change is unknown. This SNP was found in the genome of VG15 only (see Supplementary Table 6.3) who is a 14-year old pre-verbal male with ASD of WA. This participant was assessed using ADOS Module 1, and scored as “Autism” with Moderate Symptom Severity.

In the region spanning X: 1629573-1630120, the SNPs rs6588807, rs184583293 and rs28675287 occur in putative TF binding sites (Table 3.7, Figure 3.3), and rs148855515 occurs in a splice donor region. Two of the TF binding site SNPs occurred in only one individual each: rs184583293 in CP14, a 7-year old male of MA, who scored as Autism Moderate on ADOS Module 2; and rs776976390 in CP05, a 6-year old male of MA, who scored as ASD Low on ADOS Module 1. Their respective ages at the time of assessment, and the ADOS Modules with which they were assessed, indicates that CP14 has phrase speech, and that CP05 is pre-verbal.

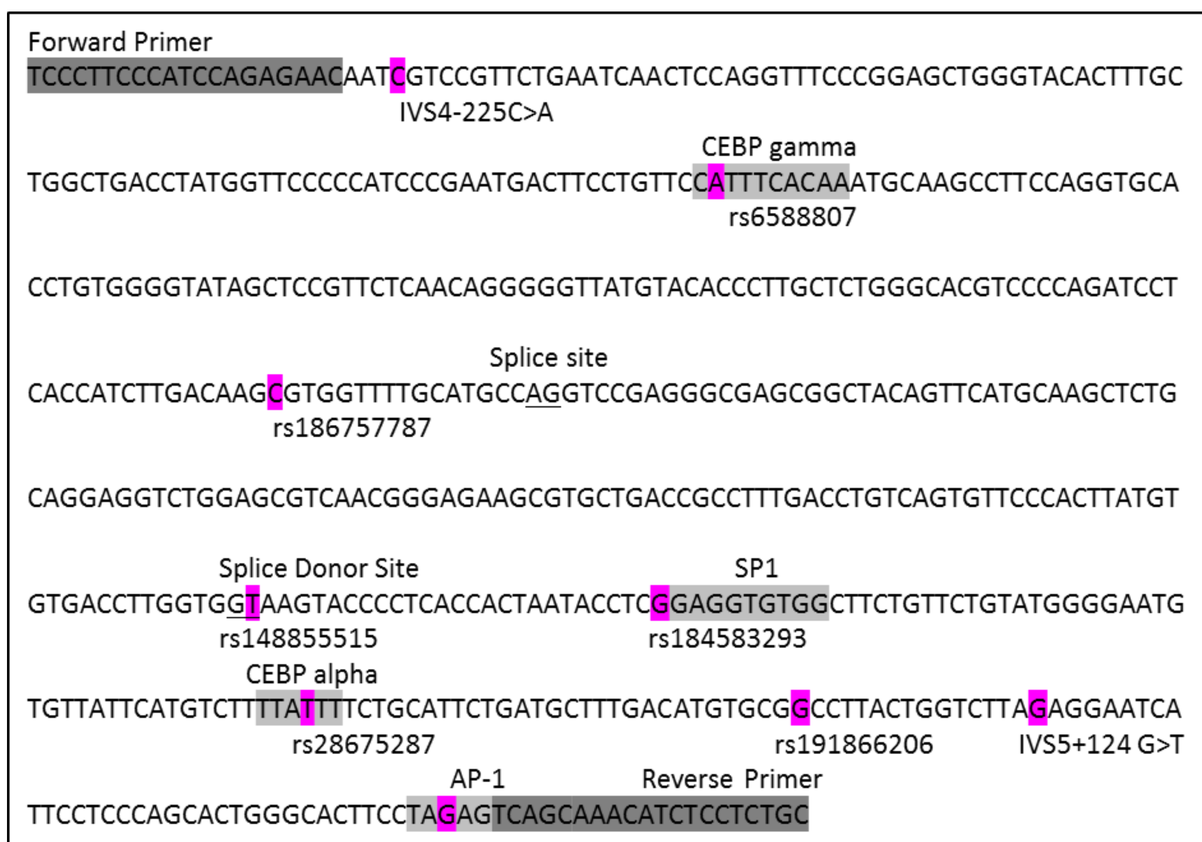

Figure 3.3 Sequence and putative transcription factor binding sites of the section of the ASMT gene intronic region spanning X: 1629573-1630120. The primer sequences are indicated in dark grey The SNPs found in this study are indicated in dark pink, with the putative binding sites in light grey. Binding sites were identified using Alibaba

The splice donor variant, rs148855515 (Table 3.7), occurs in intron 5 of ASMT, two bases from the end of exon 5 (see Figure 3.3). This SNP location corresponds to Hg18 build IVS5+2T>C, which was previously found to be associated with ASD, and ASD with severe intellectual impairment (Pagan et al. 2011). This splice site polymorphism disrupts ASMT activity, which in turn affects melatonin synthesis, as shown in cell culture as well as *in vivo*

(Melke et al. 2008). This SNP was present in the genome of only one participant, CP15, who was excluded from the final study sample due to undiagnosed developmental limitations. This participant is a male of WA, and was 9 years old at the time of assessment. He was assessed using ADOS Module 3, and scored as "Non-spectrum".

Table 3.8 Table indicating the participant data for SNPs found to be associated with a particular endophenotype. SNPs indicated with (*) have been associated with ASD previously. ([1]) indicates TF binding sites predicted using Alibaba.

SNP	Module		Symptom Severity		Participant Ages (years)	Impact	Ancestry		Significance (p=value)
rs148520524 (n=3)	Module 2=	3	Moderate=	3	7-9	In AP-1 binding site (Jonsson et al. 2010)	BA= MA= WA=	1 1 1	Mod1 vs. Mod 2 (0.0711)
rs6588807* (n=24)	Module 1= Module 2= Module 3= Controls=	11 6 3 4	Low= Moderate= High= Controls=	2 16 3 3	6-14	Predicted CEBPG binding site[1]	BA= MA= WA=	3 11 10	High vs. Moderate (0.0681)
rs5989681* (n=18)	Module 1= Module 2= Module 3= Controls=1	8 5 4 1	Low= Moderate= High= Controls=	2 12 3 1	6-14	In NF-Kappab binding site (Jonsson et al. 2010)	BA= MA= WA=	3 8 7	BA vs. MA (0.0123) BA vs. WA (0.0102)
rs56690322 (n=11)	Module 1= Module 2= Module 3= Controls=	6 2 2 1	Low= Moderate= High= Controls=	2 7 1 1	6-12	Predicted SRF binding site[1]	MA= WA=	9 2	WA vs. MA (0.0440)
rs186757787 (n=4)	Module 1= Module 2=	3 1	Low= Moderate= High=	1 2 1	6-12	unknown	BA= MA=	2 2	BA vs. MA (0.0207) BA vs. WA (0.0028)

BA=Black Ancestry; MA= Mixed Ancestry; WA=White Ancestry; ASD= Autism Spectrum Disorder

4 DISCUSSION

4.1 VERBAL FLUENCY AND AUTISM SPECTRUM DISORDERS

Before starting an ADOS assessment, one of the first decisions that the research assessors make, is which ADOS module is most applicable to the verbal fluency and age of the participant, as previously discussed (Table 1.1). Module 1 is used for individuals who are "pre-verbal", with fewer than five words, or for individuals unable to use spontaneous, meaningful phrases. Module 1 is generally intended for younger children, but can be adapted for use in older children or adults who do not have phrase speech or fluent language. Module 2 is intended for individuals with phrase speech; which is defined as using more than three words with simple grammatical markings. Module 3 is aimed at individuals who speak with verbal fluency, and this module was completed by all our control participants (five to ten years old). Here language is comprised of complex sentences are used in a mostly grammatically correct manner, with some complex sentences using clauses. Module 4 is used for adolescents and or adults for whom an interview assessment is more appropriate than a play-style interview (Lord et al. 1989).

Using the developmental marker of spontaneous, expressive language with respect the different ADOS module used, a significant association was found with ASD and SNPs in the promoter region of ASMT. In the Module 1 vs. Module 2 comparison, rs148520524 is only present in ASD participants assessed with Module 2 ($p=0.0711$). DSM-5 and ADOS-2 do not require delays in expressive language development as a criterion for an ASD diagnosis, because language acquisition in ASD is so variable. In fact, individuals who are completely verbally fluent (Module 3) can still have debilitating impairments in social communication or pragmatic language. However, the Module 1 vs. Module 2 differences observed in this study hint at a real ASD endophenotype. It has been reported that individuals with ASD who are pre-verbal (Module 1) tend to experience greater symptom severity, often with accompanying intellectual impairment (Luyster et al. 2008).

The SNP rs148520524 falls within a putative AP-1 transcription factor (TF) binding site (Figure 3.2) in the promoter region of ASMT (Jonsson et al. 2010). AP-1 TFs regulate gene expression in response to a variety of stimuli, such as cytokines or growth factors (Angel &

Michael Karin 1991). AP-1 is involved in a variety of cellular processes, such as apoptosis and cell differentiation (Ameyar et al. 2003). These TFs usually form a complex with other subunits in order to bind to palindromic 5'-TGAG/CTCA-3' sequences, which means that the rs148520524 C>T substitution may disrupt AP-1 binding (Angel et al. 1987). However, specificity of the AP-1 complex may vary depending on dimer composition (Angel & Michael Karin 1991), so the SNP may not have an appreciable effect on ASMT transcription.

If this SNP does have a large enough effect on ASMT transcription to impact melatonin biosynthesis, the contribution this polymorphism would have to the development of language deficits is not immediately apparent. Decreased expression of A-ANAT (Hu et al. 2009; Hu & Steinberg 2009), and low melatonin levels (Tordjman et al. 2013), have (in some cases) been found to be associated with severe language delays in ASD. Despite this putative association, no direct mechanism by which melatonin could influence language has not been characterised. However, potential causes have been suggested.

Firstly, melatonin has been linked to neural development and calcium signalling (De Faria Poloni et al. 2011). Extra-pineal melatonin is synthesised in mammalian placental tissue during the first trimester of pregnancy, and melatonin receptor gene expression is higher in embryos than in adults (Iwasaki et al. 2005; Von Gall et al. 2002). It is thought that it plays a role in embryonic neurodevelopment through melatonin regulating the calcium signalling system, in turn modulating intercellular signalling cascades (Iwasaki et al. 2005)

Secondly, sleep disruption has been linked to language delays in children with developmental disorders (Giannotti et al. 2008; Axelsson et al. 2013). Melatonin deficiency has been associated with disrupted sleep (Kawabe et al. 2014). The deleterious effects of sleep disruption in general has been well-documented in typically developing adults (Samkoff & Jacques 1991), as well as the effect of sleep disruption on the behaviour of typically developing children (Sadeh 2007). These negative health outcomes are compounded in children with developmental disorders, where sleep disruption may form part of the genetic aetiology of their disorders. The effects of sleep disruption may contribute to existing behavioural and cognitive impairments (Axelsson et al. 2013).

The possible association between rs148520524 (p=0.0711) and the phrase-speech endophenotype suggests that ASMT could possibly have a role in structural and pragmatic

language impairments. However, ASD language phenotypes are complex, and have been associated with multiple polymorphisms in the genome (Bradford et al. 2001; Chen et al. 2006; Alarcón et al. 2005). The language phenotype needs to be more clearly defined and stratified before strong claims can be made about possible associations between ASMT and language delay in ASD. In addition, it has been noted that the contactin-associated protein-like 2 gene has been implicated specific language impairment (SLI), as well as an ASD (Grzadzinski et al. 2013). Therefore it is important to have an independent test of language ability in addition to the ADOS-2 assessment. This would allow the discrimination of whether the difference in SNP frequency observed between Module 1 and Module 2 is due to a SLI, an ASD endophenotype or another associated factor.

When considering language endophenotypes, it is necessary to consider factors such as cognitive impairment. Cognitive impairment is more common among individuals with ASD who meet the criteria for a Module 1 ADOS assessment, and it could influence language development (Luyster et al. 2008). There are many language assessment tools (e.g. Peabody Vocabulary Test), which could confirm impairment in spoken language. However, most tools have some limitations. These limitations are the potential for cultural biases in South Africa (Haitana et al. 2010; Rock & Stenner 2005), and lack of applicability across the autism spectrum (e.g. Raven's Progressive Matrices only used for pre-verbal participants).

A further factor to consider, particularly in the South African context, is that it is a multi-lingual country, with most people speaking at least two languages. In this study the ADOS was administered in English, the schools that were approached for this study use English as the language of instruction, and three of the participants in this study did not use English as their home language. Noteworthy is that all three of these boys were preverbal and completed a Module 1 ADOS assessment. There has been a debate about the application of any diagnostic tool in a variety of languages, particularly tools designed to assess verbal interaction, and whether it may negate the standardisation. Non-Western cultural norms may also interfere with tests that were developed for Western children. An example of this bias is as found by Rock & Stenner (2005) when reviewing school readiness tests. These additions will follow as this cohort is grown, as ADOS has not been widely used in assessing ASD in South Africa. This the first study of its kind in South Africa, and the significance of studying a South African population is the genetic diversity as well as the cultural diversity.

Therein lays the need to validate tools that have been developed for use in other contexts. As the cohort grows, potential bias this will be easier to evaluate.

The SNP rs6588807 was found to be tentatively associated (p-value= 0.06812) with ADOS Symptom Severity, occurring more frequently in the participants with “Moderate” than “High” Symptom Severity. The putative CEBPG site containing this SNP does not give a clear indication of its functional significance, as this TF has been linked to bronchial carcinoma, but no other clinical or psychiatric conditions. It could conceivably form part of a downstream enhancer complex for ASMT, as CEBPG TFs do act as enhancers, but there is no published evidence of an enhancer region in ASMT introns. The SNP was identified in and ASD study on European subjects, but no statistical significance was found (Holt et al. 2010).

In addition to potential cultural and linguistic effects, the unique ancestral heritage of the South African population may act as a confounding factor in ASD genotyping. Two SNPs, rs5989681 and rs186757787 occurred significantly more frequently in the BA group than in the MA (p-values of 0.0123 and 0.0207 respectively) or the WA (p-values of 0.0102 and 0.0028 respectively) groups (Table 3.8). As the BA group is clearly underrepresented in the study sample (n=3), it is possible that this may be due to the limited sample size. However, data from the 1000 Genomes Project indicates that the allele frequency for rs5989681 is different in the African super population than in the other four super populations (Table 3.3). In the Admixed American, South and East Asian, and European populations the G allele is the major allele, but in the African super population the C allele is the major allele. By inspection of Table 3.3 it is visible that the disparity in frequency if the G versus the C allele is greater in our BA group than in the African super population, but it is reasonable to assume that this is the product of our small sample size. The minor (T) allele frequency for rs186757787 is greater in the African super population than in the Admixed American super population, and is absent from the other three super populations (Table 3.3). In our three ancestry groups, the T allele appears most frequently in the BA group, appears at a lower frequency in the MA group, and is absent from the WA group. While these observations hint at demographic effects in our study, it would be necessary to increase the number of BA participants in the study, as well as increasing the sample size overall, to establish if this association is due to sample size, or population demography.

The functional effect of rs186757787 is unknown, as it falls in the intronic region, and none of the software used in this study found any putative TF binding sites at this position (Figure 3.3). The SNP rs5989681, on the other hand, has been associated with ASD endophenotypes in previous studies, and falls in a nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) binding site (Jonsson et al. 2010). Previous studies have suggested that NF-κB TFs play a role in synaptic plasticity and memory, which have in turn been linked to language development (Mattson & Camandola 2001; Romano et al. 2006; Kalkman 2012; Hu et al. 2009). Melke *et al.* (2008) found the G allele of rs5989681 to appear more frequently in ASD, and to be associated with decreased expression of ASMT transcript (P= 2×10^{-10}).

The last SNP approaching a significant difference in frequency was rs56690322 (p-value= 0.0440), between participants of MA (n=15) and WA (n=16). These two groups are better represented in the study cohort, which makes the comparison more meaningful than the comparisons between the BA group and the WA or MA group. This SNP has been identified in previous ASD studies (Melke et al. 2008; Pereira et al. 2013; Wang et al. 2013), and forms part of the Melke *et al.* GGGC "risk" haplotype. However, this study found no significant difference in the minor (A) allele frequency when comparing ASD endophenotypes to controls. This minor allele appeared more frequently in the MA ancestry group, suggesting that it may be a more variable region in certain ancestry groups, without contributing to a disease state. The allele frequencies for this SNP from the 1000 Genomes Project (Table 3.3), do suggest that the minor allele occurs at a low frequency in all five of the super populations. If this SNP did contribute to the ASD phenotype, a possible mode of action for this would be through its location in a putative SRF TF binding site. A knockout mouse study found that if this family of TFs is not expressed in the hippocampus, it negatively impacts the neurological process of long-term depression (Etkin et al. 2006). Long-term depression (LTD) is a process that weakens synaptic affinity in response to a stimulus, which contributes to the process of memory formation and learning (Massey & Bashir 2007). SRF TFs perform several functions in different tissues, which also includes neuronal embryonic development in humans (Cooper et al. 2009). It is possible that this SNP does not contribute to the ASD phenotype on its own, but rather as part of a haplotype.

4.2 AUTISM RISK HAPLOTYPES

While several polymorphisms in ASMT have previously been found in ASD patients, there has not been consensus on which SNPs, if any, contribute directly to ASD aetiology (Veatch, Goldman, et al. 2015). The finding of a "risk" haplotype identified by Melke *et al* (2008) has not been reproduced in subsequent studies (Wang et al. 2013; Veatch, Pendergast, et al. 2015). Indeed, in this study, the GGGC haplotype was found to be significantly associated (p=0.0313) with the typically-developing Control group (Table 3.6). As the data from this study hints at ancestry effects (Table 3.4, Figure 3.1b), this comparison was repeated with WA and MA participants only. Unfortunately, attempting the LD comparison using the WA or MA groups individually resulted in a map without enough statistical power to indicate LD at all. However, regardless of whether entire final sample (n= 34) was used, or whether the BA participants were excluded (n= 31), the GGGC haplotype was still significantly associated with the Control group (Supplementary Table 5.7). Furthermore, the ACGC haplotype, which Melke *et al* (2008) found to be significantly associated with their Control group, showed no significant association (p=0.5004) with either the ASD group or the Control group.

When testing for linkage disequilibrium among the SNPs found in this study (Figure 3.1a), three of the SNPs from the Melke (2008) 'risk' haplotype are in linkage disequilibrium (rs5989681, rs56690322, and rs6644635). Additionally, the SNP rs186757787 forms part of this haplotype, despite its position 14kb downstream from the 'risk' haplotype SNPs. These four SNPs are indicated by a black haploblock, inserted manually for clarity (Figure 3.1a). Three of these SNPs (rs5989681, rs56690322, and rs186957787) are associated with ancestry in this study (Table 3.4), with the remaining SNP (rs6644635) showing no significant association with any of the endophenotypes in this study. As can be seen in Table 3.5, the GGCC haplotype shows strong association with the Control group. This haplotype constitutes the major alleles for rs5989681, rs56690322, s186757787, rs6644635 (Table 3.2).

Due to the strong association between three of the SNPs in this haplotype and ancestry, the LD analysis was repeated using only the MA and WA participants from the final sample set (Figure 3.1b). Based on the genotypes of the MA and WA participants, only rs56690322 and rs6644635 are in LD. None of the possible haplotypes of these two SNPs are significantly

associated with either the ASD or Control group. Though not all of the SNPs are in significant LD, the GGCC haplotype indicated in Figure 3.1a is still strongly associated with the Control group (Supplementary Table 6.6).

4.3 Conclusion and Future research

Based on the results of this study, the "risk" haplotype identified by Melke *et al* (2008) appears to not be implicated in the aetiology of endophenotypes in South African individuals with ASD. The homozygous presence of the major alleles of rs4446909 and rs5989681 is strongly associated with the Control group in this study, in contrast to what was found by Melke *et al* (2008). Thus this study found no association between this haplotype and ASD risk, similar to previous studies (Toma et al. 2007; Wang et al. 2013). However, the small sample size of this study necessitates that this finding be treated with caution. It is not known whether the GGGC "risk" haplotype that these authors found to be associated with decreased ASMT expression, and by extension decreased melatonin levels, will have the same effect in this South African population. Instead, this study has found some evidence of possible rare variants in BA individuals with ASD, though further studies on much larger sample sizes would be required before this evidence can be considered conclusive. Larger sample sizes would also allow for corrections for multiple testing, such as the Bonferroni Correction, or the Holm-Bonferroni Method.

This study also identified a SNP not previously associated with ASD, rs148520524, which is associated with individuals with ASD who were tested using ADOS Module 2 (phrase speech), and not in individuals tested with ADOS Module 1 (pre-verbal). This SNP falls in an AP-1 TF binding site (Figure 3.2), which could have an effect on neural development, as previously discussed. This association could potentially be elucidated using language testing tools in addition to the ADOS-2 to establish whether it is due to an ASD endophenotype, or an independent language impairment. Further endophenotyping could also aid in establishing the association between the SNP rs6588807 and ADOS Symptom Severity.

As part of the central premise of the putative role of ASMT in ASD is melatonin synthesis, quantitative testing of blood melatonin levels (or melatonin metabolites in urine) would be required to confirm the association. To confirm whether melatonin or melatonin metabolite

levels are associated with sleep disruption in ASD, sleep diaries and actigraphy data would give an indication of the circadian cycles of the ASD cohort. Additionally, expanding the genetic screening of participants to include other genes involved in the melatonin pathway could better characterise the contribution of melatonin to ASD aetiology. These genes could include the genes for melatonin receptors (MT1 and MT2), the rate limiting enzyme in melatonin biosynthesis (A-ANAT), and the gene for the enzyme responsible for metabolising melatonin (CYP1A2). If little to no variation in these genes is found to be associated with ASD, it would imply possible dysfunction in the central oscillator. If it is established that melatonin deficiency (as a result of a polymorphism) results in sleep disruption in individuals with ASD, and that these same individuals have impairments in language development, it could confirm the role of melatonin in ASD aetiology.

5 BIBLIOGRAPHY

Alarcón, M. et al., 2005. Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs. *Molecular psychiatry*, 10(8), pp.747–757.

Aljanabi, S.M. & Martinez, I., 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic acids research*, 25(22), pp.4692–4693.

American Psychiatric Association. 2000. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders: DSM-V. Arlington, VA: American Psychiatric Association.

Ameyar, M., Wisniewska, M. & Weitzman, J.B., 2003. A role for AP-1 in apoptosis: The case for and against. *Biochimie*, 85(8), pp.747–752.

Angel, P et al., 1987. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. Cell, 49(6), pp.729-739.

Angel, P & Michael Karin. 1991. The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. BBA - Reviews on Cancer. 1072 (2), pp.129-157

Arendt, J., 1998. Melatonin and the pineal gland: influence on mammalian seasonal and circadian physiology. *Reviews of Reproduction*, 3(1), pp.13–22.

Arieff, Z., Kaur, M. & Gameeldien, H., 2010. 5-HTTLPR polymorphism: analysis in South African autistic individuals. *Human biology*, 82(3), pp.291–300.

Axelsson, E.L. et al., 2013. Sleep problems and language development in toddlers with Williams syndrome. *Research in Developmental Disabilities*, 34(11), pp.3988–3996

Barrett, J.C. et al., 2005. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), pp.263–265.

Bartlett, C.W. et al., 2005. Three autism candidate genes: a synthesis of human genetic analysis with other disciplines. *International journal of developmental neuroscience: the official journal of the International Society for Developmental Neuroscience*, 23(2-3), pp.221–34.

Botros, H.G. et al., 2013. Crystal structure and functional mapping of human ASMT, the last enzyme of the melatonin synthesis pathway. *Journal of pineal research*, 54(1), pp.46–57.

Bradford, Y. et al., 2001. Incorporating language phenotypes strengthens evidence of linkage to autism. *American Journal of Medical Genetics - Neuropsychiatric Genetics*, 105(6), pp.539–547.

Brzezinski, A. et al., 2005. Effects of exogenous melatonin on sleep: a meta-analysis. *Sleep Medicine Reviews*, 9(1), pp.41–50.

Buscemi, N. et al., 2006. Efficacy and safety of exogenous melatonin for secondary sleep disorders and sleep disorders accompanying sleep restriction: meta-analysis. *BMJ*, 332(385).

Cajochen, C. et al., 2011. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. *Journal of applied physiology*, 110(5), pp.1432–1438.

Campino, C. et al., 2011. Melatonin exerts direct inhibitory actions on ACTH responses in the human adrenal gland. *Hormone and Metabolic Research*, 43(5), pp.337–342.

Chaste, P. et al., 2010. Identification of pathway-biased and deleterious melatonin receptor mutants in autism spectrum disorders and in the general population. *PloS one*, 5(7), p.e11495.

Chen, G.K. et al., 2006. Quantitative trait locus analysis of nonverbal communication in autism spectrum disorder. *Molecular psychiatry*, 11(2), pp.214–220.

Cooper, S.J. et al., 2009. Serum response factor binding sites differ in three human cell types. *Genome Res.*, 17, pp.136–144.

Cotton, S.M. & Richdale, A.L., 2010. Sleep patterns and behaviour in typically developing children and children with autism, Down syndrome, Prader-Willi syndrome and intellectual disability. *Research in Autism Spectrum Disorders*, 4(3), pp.490–500.

Courchesne, E. et al., 2007. Mapping early brain development in autism. *Neuron*, 56(2), pp.399–413.

Czeisler, C.A. et al., 1999. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science*, 284(5423), pp.2177–2181.

Elsabbagh, M. et al., 2012. Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Research*, 5(3), pp.160–179.

Etain, B. et al., 2012. Genetic and functional abnormalities of the melatonin biosynthesis pathway in patients with bipolar disorder. *Human molecular genetics*, 21(18), pp.4030–7.

Etkin, A. et al., 2006. A Role in Learning for SRF: Deletion in the Adult Forebrain Disrupts LTD and the Formation of an Immediate Memory of a Novel Context. *Neuron*, 50(1), pp.127–143.

De Faria Poloni, J., Feltes, B.C. & Bonatto, D., 2011. Melatonin as a central molecule connecting neural development and calcium signaling. *Functional and Integrative Genomics*, 11(3), pp.383–388.

Gałecki, P. et al., 2010. Single-nucleotide polymorphisms and mRNA expression for melatonin synthesis rate-limiting enzyme in recurrent depressive disorder. *Journal of pineal research*, 48(4), pp.311–7.

Von Gall, C., Stehle, J.H. & Weaver, D.R., 2002. Mammalian melatonin receptors: Molecular biology and signal transduction. *Cell and Tissue Research*, 309(1), pp.151–162.

Geoffroy, P.A. et al., 2015. Genetic association study of circadian genes with seasonal pattern in bipolar disorders. *Scientific reports*, 5, p.10232.

Giannotti, F. et al., 2008. An investigation of sleep characteristics, EEG abnormalities and epilepsy in developmentally regressed and non-regressed children with autism. *Journal of Autism and Developmental Disorders*, 38(10), pp.1888–1897.

Glessner, J.T. et al., 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, 459(7246), pp.569–73.

Goldman, S.E. et al., 2014. Melatonin in Children with Autism Spectrum Disorders: Endogenous and Pharmacokinetic Profiles in Relation to Sleep. *Journal of Autism and Developmental Disorders*, 44(10), pp.2525–2535.

Gringras, P. et al., 2012. Melatonin for sleep problems in children with neurodevelopmental disorders: randomised double masked placebo controlled trial. *Bmj*, 345(nov05 1), pp.e6664–e6664.

Grzadzinski, R., Huerta, M. & Lord, C., 2013. DSM-5 and Autism Spectrum Disorders (ASDs): an opportunity for identifying ASD subtypes. *Molecular Aurism*, 4(12), pp.2–7.

Haitana, T., Pitama, S. & Rucklidge, J.J., 2010. Cultural biases in the Peabody Picture Vocabulary Test-III: Testing tamariki in a New Zealand sample. *New Zealand Journal of Psychology*, 39(3), pp.24–34.

Hardeland, R. et al., 2011. Melatonin-A pleiotropic, orchestrating regulator molecule. *Progress in Neurobiology*, 93(3), pp.350–384.

Hardeland, R., 2012. Neurobiology, pathophysiology, and treatment of melatonin deficiency and dysfunction. *The Scientific World Journal*, 2012, p.640389.

Holt, R. et al., 2010. Linkage and candidate gene studies of autism spectrum disorders in European populations. *European Journal of Human Genetics*, 18(9), pp.1013–1019.

Hu, V.W. et al., 2009. Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. *Autism research : official journal of the International Society for Autism Research*, 2(2), pp.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: official journal of the International Society for Autism Research*, 2(2), pp.67–77.

Iwasaki, S. et al., 2005. Melatonin as a local regulator of human placental function. *Journal of Pineal Research*, 39(3), pp.261–265.

Jan, J. & Freeman, R., 2004. Melatonin therapy for circadian rhythm sleep disorders in children with multiple disabilities: what have we learned in the last decade? *Developmental Medicine & Child Neurology*, 46, pp.776–782.

Jin, X. et al., 1999. A Molecular Mechanism Regulating Rhythmic Output from the Suprachiasmatic Circadian Clock. *Cell*, 96(1), pp.57–68.

Jonsson, L. et al., 2014. Association between ASMT and autistic-like traits in children from a Swedish nationwide cohort. *Psychiatric genetics*, 24(1), pp.21–7.

Jonsson, L. et al., 2010. Mutation screening of melatonin-related genes in patients with autism spectrum disorders. *BMC medical genomics*, 3, p.10.

Kalkman, H.O., 2012. Potential opposite roles of the extracellular signal-regulated kinase (ERK) pathway in autism spectrum and bipolar disorders. *Neuroscience and biobehavioral reviews*, 36(10), pp.2206–13.

Kawabe, K. et al., 2014. The Melatonin Receptor Agonist Ramelteon Effectively Treats Insomnia and Behavioral Symptoms in Autistic Disorder. *Case Reports in Psychiatry*, 2014(561071).

Ko, C.H. & Takahashi, J.S., 2006. Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, 15(SUPPL. 2), pp.271–277.

Kojo, K., Pukkala, E. & Auvinen, a, 2005. Breast cancer risk among Finnish cabin attendants: a nested case-control study. *Occupational and environmental medicine*, 62(7), pp.488–93.

Li, X., Zou, H. & Brown, W.T., 2012. Genes associated with autism spectrum disorder. *Brain research bulletin*, 88(6), pp.543–52.

Lord, C. et al., 1989. Austism diagnostic observation schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, 19(2), pp.185–212.

Lord, C., Rutter, M. & Couteur, A.L., 1994. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), pp.659–685.

Luyster, R.J. et al., 2008. Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(8), pp.1426–1438.

Macchi, M.M. & Bruce, J.N., 2004. Human pineal physiology and functional significance of melatonin. *Frontiers in Neuroendocrinology*, 25(3-4), pp.177–195.

Malow, B. et al., 2012. Melatonin for sleep in children with autism: A controlled trial examining dose, tolerability, and outcomes. *Journal of Autism and Developmental Disorders*, 42(8), pp.1729–1737.

Malow, B.A. et al., 2006. Characterizing sleep in children with autism spectrum disorders: a multidimensional approach. *Sleep*, 29(12), pp.1563–71.

Massey, P. V. & Bashir, Z.I., 2007. Long-term depression: multiple forms and implications for brain function. *Trends in Neurosciences*, 30(4), pp.176–184.

Mattson, M.P. & Camandola, S., 2001. NF- κ B in neuronal plasticity and neurodegenerative disorders. *The Journal of Clinical Investigation*, 107(3), pp.247–254.

Melke, J. et al., 2008. Abnormal melatonin synthesis in autism spectrum disorders. *Molecular psychiatry*, 13(1), pp.90–8.

Nicholas, B. et al., 2007. Association of Per1 and Npas2 with autistic disorder: support for the clock genes/social timing hypothesis. *Molecular psychiatry*, 12(6), pp.581–92.

Pagan, C. et al., 2011. Mutation screening of ASMT, the last enzyme of the melatonin pathway, in a large sample of patients with intellectual disability. *BMC medical genetics*, 12(1), p.17.

Pagan, C. et al., 2014. The serotonin-N-acetylserotonin-melatonin pathway as a biomarker for autism spectrum disorders. *Translational psychiatry*, 4(11), p.e479.

Pandi-Perumal, S.R. et al., 2008. The roles of melatonin and light in the pathophysiology and treatment of circadian rhythm sleep disorders. *Nature clinical practice. Neurology*, 4(8), pp.436–447.

Pereira, D.S. et al., 2013. Screening for polymorphisms in the HIOMT gene and associations with circadian phenotypes. *www. sleepscience. …*, 6(2), pp.66–71.

Pinto, D. et al., 2010. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 466(7304), pp.368–72.

Rock, D.A. & Stenner, A.J., 2005. Assessment issues in the testing of children at school entry. *Future of Children*, 15(1), pp.15–34.

Roenneberg, T., Daan, S. & Merrow, M., 2003. The Art of Entrainment. *Journal of Biological Rhythms*, 18(3), pp.183–194.

Romano, A. et al., 2006. Evolutionarily-conserved role of the NF-KB transcription factor in neural plasticity and memory. *European Journal of Neuroscience*, 24(6), pp.1507–1516.

Rossignol, D. a & Frye, R.E., 2011. Melatonin in autism spectrum disorders: a systematic review and meta-analysis. *Developmental medicine and child neurology*, 53(9), pp.783–92.

Sadeh, A., 2007. Consequences of Sleep Loss or Sleep Disruption in Children. *Sleep Medicine Clinics*, 2(3), pp.513–520.

Samkoff, J.S. & Jacques, C.H., 1991. A review of studies concerning effects of sleep deprivation and fatigue on residents’ performance. *Academic Medicine*, 66(11).

Shibui, K. et al., 2003. Melatonin, cortisol and thyroid-stimulating hormone. *Sleep and Biological Rhythms*, 1, pp.209–214.

Shimozuma, M. et al., 2011. Expression and cellular localizaion of melatonin-synthesizing enzymes in rat and human salivary glands. *Histochemistry and cell biology*, 135(4), pp.389–96.

Smith, A.K. et al., 2015. DNA Extracted From Saliva for Methylation Studies of Psychiatric Traits: Evidence Tissue Specificity and Relatedness to Brain. *Am J Med Genet B Neuropsychiatr Genet*, 0(1), pp.36–44.

Szatmari, P. et al., 2015. Developmental Trajectories of Symptom Severity and Adaptive Functioning in an Inception Cohort of Preschool Children With Autism Spectrum Disorder. *JAMA psychiatry*, 72(3), pp.276–283.

Talarowska, M. et al., 2014. ASMT gene expression correlates with cognitive impairment in patients with recurrent depressive disorder. *Medical science monitor : international medical journal of experimental and clinical research*, 20, pp.905–12.

Toma, C. et al., 2007. Is ASMT a susceptibility gene for autism spectrum disorders? A replication study in European populations. *Molecular psychiatry*, 12, pp.977–979.

Tordjman, S. et al., 2013. Advances in the research of melatonin in autism spectrum disorders: Literature review and new perspectives. *International Journal of Molecular Sciences*, 14(10), pp. 20508-20542

Veatch, O.J., Pendergast, J.S., et al., 2015. Genetic Variation in Melatonin Pathway Enzymes in Children with Autism Spectrum Disorder and Comorbid Sleep Onset Delay. *Journal of Autism and Developmental Disorders*, 45(1), pp.100–110.

Veatch, O.J., Goldman, S.E., et al., 2015. Melatonin in Children with Autism Spectrum Disorders: How Does the Evidence Fit Together? *Journal of nature and science*, 1(7), p.e125.

Wang, L. et al., 2013. Sequencing ASMT identifies rare mutations in Chinese Han patients with autism. *PloS one*, 8(1), p.e53727.

Wimpory, D., Nicholas, B. & Nash, S., 2002. Social timing, clock genes and autism: a new hypothesis. *Journal of intellectual disability research*, 46(4), pp.352–358.

Wittmann, M. et al., 2006. Social jetlag: misalignment of biological and social time. *Chronobiology international*, 23(1-2), pp.497–509.

Wright, B. et al., 2011. Melatonin versus placebo in children with autism spectrum conditions and severe sleep problems not amenable to behaviour management strategies: A randomised controlled crossover trial. *Journal of Autism and Developmental Disorders*, 41(2), pp.175–184.

Yousaf, F. et al., 2010. Efficacy and Safety of Melatonin as an Anxiolytic and Analgesic in the Perioperative Period: A Qualitative Systematic Review of Randomized Trials. *Anesthesiology*, 113, pp.968–976.

6 SUPPLEMENTARY MATERIALS

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:

Supplementary Figure 6.1 Sample of consent form used in this study

Supplementary Table 6.1 Polymerase chain reaction (PCR) cycling conditions

GRCh38.p2 position X:1614846-1615299			**GRCh38.p2 position X:1629573-1630120**		
Time (min)	Temperature (°C)	Cycles	Time (min)	Temperature (°C)	Cycles
0:30	98	x1	0:30	98	x1
0:10	98		0:10	98	
0:20	58	x35	0:20	56	x35
0:14	72		0:17	72	
5:00	72	x1	5:00	72	x1

Supplementary Table 6.2 Polymerase chain reaction (PCR) reagent concentrations using ThermoScientific Phusion High-Fidelity DNA Polymerase #F-530S

Component	20µL reaction (µL)	Final Concentration
H_2O	13.0	-
5x Phusion GC Buffer*	4.0	1x
10mM DNTPs	0.4	200 µM each
Forward Primer	0.2	0.5 µM
Reverse Primer	0.2	0.5 µM
Template DNA	2.0	8ng per reaction
Phusion DNA Polymerase	0.2	0.02 U/µL

*Kit comes with HiFi and GC buffer

Supplementary Table 6.3 Demographic and ADOS data of all study participants

SampleID	Module	ADOS Comparison score	ADOS Classification	Level of Symptoms	Comorbidities	Age	Ethnic group
CP01	3	3	Non-Spectrum	Low	Yes	7	WA
CP03	3	6	Autism	Moderate	No	6	MA
CP05	3	4	ASD	Low	No	8	WA
CP07	1	7	Autism	Moderate	No	6	MA
CP09	2	6	Autism	Moderate	No	6	WA
CP12	3	6	Autism	Moderate	No	7	WA
CP14	2	7	Autism	Moderate	No	7	MA
CP15	3	2	Non-Spectrum	Minimal/No evidence	Yes	9	WA
NT01	3	1	Non-Spectrum	Minimal/No evidence	No	7	WA
NT02	3	1	Non-Spectrum	Minimal/No evidence	No	6	WA
NT03	3	6	Autism	Moderate	Yes	7	WA
NT04	3	1	Non-Spectrum	Minimal/No evidence	No	10	WA
NT07	3	1	Non-Spectrum	Minimal/No evidence	No	6	WA
NT10	3	1	Non-Spectrum	Minimal/No evidence	No	6	WA
NT11	3	1	Non-Spectrum	Minimal/No evidence	No	7	WA
VG01	2	7	Autism	Moderate	No	10	MA
VG02	1	9	Autism	High	No	9	WA
VG03	1	7	Autism	Moderate	No	8	MA
VG05	1	8	Autism	High	No	6	WA
VG06	2	6	Autism	Moderate	No	7	MA
VG07	2	3	Non-Spectrum	Low	Yes	8	MA
VG08	3	8	Autism	High	No	12	MA
VG10	1	7	Autism	Moderate	No	6	MA
VG11	1	6	Autism	Moderate	No	8	BA
VG12	1	4	ASD	Low	No	6	MA
VG13	1	4	ASD	Low	No	7	MA
VG15	1	6	Autism	Moderate	No	14	WA
VG16	1	6	Autism	Moderate	No	14	WA
VG17	1	6	Autism	Moderate	No	8	WA
VG18	1	10	Autism	High	No	8	MA
VG19	1	9	Autism	High	No	8	BA
VG20	1	6	Autism	Moderate	No	5	MA
VG21	2	6	Autism	Moderate	No	10	WA
VG23	3	6	Autism	Moderate	No	10	WA
VG24	2	7	Autism	Moderate	No	9	BA
VG26	2	6	Autism	Moderate	No	7	MA
VG27	2	2	Non-Spectrum	Low	Yes	10	MA
VG28	2	6	Autism	Moderate	No	10	MA
VG29	2	6	Autism	Moderate	No	10	MA

Supplementary Table 6.4 P-values of Fisher's Exact Tests comparing allele frequencies between endophenotypes

	Ancestry			ADOS	ADOS Module			ADOS Symptom Severity					
	BA vs. CA	**BA vs. WA**	**WA vs. CA**	**Non vs. ASD**	**1 vs. 2**	**1 vs. 3**	**2 vs. 3**	**Control vs. High**	**Control vs. Mod**	**Control vs. Low**	**Mod vs. High**	**High vs. Low**	**Mod vs. Low**
rs4446909	0.5701	1	0.7085	0.336	0.6852	0.6731	1	1	0.3193	1	1	1	0.5829
rs148520524	0.2945	0.3095	1	1	0.0711	1	0.2308	1	1	1	1	1	1
rs5989681	0.0123	0.0102	1	0.09365	0.7417	0.5186	1	0.3034	0.2463	0.5588	0.7172	0.638	1
rs56690322	0.3031	1	0.0440	0.6783	0.6852	1	1	1	1	0.5588	0.6527	1	0.6047
rs6644635	0.3713	0.3854	0.8017	0.3377	0.7555	0.5344	0.5145	1	0.488	1	1	0.638	0.4402
rs774990880	1	1	0.4839	1	1	1	1	1	1	1	1	1	1
IVS4-225C>A	1	1	0.4839	1	1	1	1	1	1	1	1	1	1
rs6588807	1	0.6582	0.6127	0.1025	0.5371	0.5434	1	0.3563	0.3184	0.3563	0.06812	1	1
rs186757787	0.0207	0.0028	0.4923	0.5754	0.6133	0.2326	0.4762	0.1474	1	0.4	0.1241	1	0.4288
rs148855515	1	1	1	1	1	1	1	1	1	1	1	1	1
rs184583293	1	1	1	1	0.45	1	0.4762	1	1	1	1	1	1
rs28675287	1	0.5348	1	1	0.6419	1	1	1	1	1	0.571	1	1
rs191866206	1	1	1	1	1	1	1	1	1	1	1	1	1
IVS5+124G>T	1	1	0.4839	0.1818	1	1	1	1	0.2308	1	1	1	1
rs776976390	1	1	0.4839	1	1	1	1	1	1	0.4	1	1	0.1667

Supplementary Table 6.5 Table indicating the haplotype counts in the ASD group vs. the controls (WA and MA only), for the 'risk' haplotype first identified by Melke *et al* (2008). p-values derived using Fisher's Exact test. Significance is set at $p<0.05$. The SNPs are rs4446909 (G>A), rs5989681 (G>C), rs56690322 (G>A), rs6644635 (C>T).

Haplotype	*p*-value	ASD % (n=25)	Controls % (n=5)
ACAC	0.6897	7.1 (2)	-
GCAC	0.6897	7.1 (2)	-
GGAC	0.6272	14.3 (4)	20 (1)
ACGC	0.4616	14.3 (4)	-
GCGC	0.6711	32.1 (9)	20 (1)
AGGC	0.8333	3.6 (1)	-
GGGC	0.0413	10.7 (3)	60 (3)
GGAT	0.6897	7.1 (2)	-
GGGT	0.8333	3.6 (1)	-

ASD= Autism Spectrum Disorder participants

Supplementary Table 6.6 Table indicating the haplotype counts in the ASD group vs. the controls WA and MA participants only, along with the p-values derived using Fisher's Exact test. Significance is set at $p<0.05$. The SNPs are rs5989681 (G>C), rs56690322 (G>A), rs6644635 (C>T), and rs186757787 (C>T)

Haplotype	ASD % (n=24)	Controls % (n=5)	*p*-value
CACC	12.5 (3)	-	0.5539
CACT	4.2 (1)	-	0.8275
CGCC	37.5 (9)	20 (1)	0.4243
CGCT	4.2 (1)	-	0.8275
GACC	16.7 (4)	20 (1)	0.6421
GATC	4.2 (1)	-	0.8275
GGCC	16.7 (4)	60 (3)	0.0747
GGTC	4.2 (1)	-	0.8275

ASD= Autism Spectrum Disorder participants

Supplementary Table 6.7 Table indicating the Hardy-Weinberg test output for the ASD group vs. the controls, with the p-values derived using Hardy-Weinberg Exact test. Significance is set at $p<0.05$. Test performed using GenePop web version 4.2 (http://genepop.curtin.edu.au/genepop_op1.html)

	Locus	*p*-value	S.E.	Fis Estimates	
				W&C	R&H
ASD	**rs4446909**	0.4373	0.0022	0.1429	0.1458
	rs148520524	1	0	-0.0385	-0.0391
	rs5989681	0.4276	0.0027	-0.1773	-0.18
	rs56690322	1	0	-0.0425	-0.0432
	rs6644635	0.7054	0.0017	-0.1309	-0.133
	rs17149149	-	-	-	-
	IVS4-225C>A	1	0	-0.04	-0.0407
	rs6588807	0.7107	0.0017	-0.0986	-0.1003
	rs186757787	1	0	-0.04	-0.0407
	rs148855515	-	-	-	-
	rs184583293	-	-	-	-
	rs28675287	1	0	-0.1304	-0.1326
	rs191866206	-	-	-	-
	IVS5+124 G>T	-	-	-	-
	rs77697639	-	-	-	-
Control	**rs4446909**	-	-	-	-
	rs148520524	-	-	-	-
	rs5989681	-	-	-	-
	rs56690322	-	-	-	-
	rs6644635	1	0	-0.3333	-0.3571
	rs17149149	-	-	-	-
	IVS4-225C>A	-	-	-	-
	rs6588807	1	0	-0.25	-0.2667
	rs186757787	-	-	-	-
	rs148855515	-	-	-	-
	rs184583293	-	-	-	-
	rs28675287	-	-	-	-
	rs191866206	-	-	-	-
	IVS5+124 G>T	-	-	-	-
	rs77697639	-	-	-	-

ASD= Autism Spectrum Disorder participants

Supplementary Table 6.8 Table indicating the Hardy-Weinberg test output for the entire cohort, with the p-values derived using Hardy-Weinberg Exact test. Significance is set at p<0.05. Test performed using GenePop web version 4.2 (http://genepop.curtin.edu.au/genepop_op1.html)

Locus	*p*-value	S.E.	Fis Estimates	
			W&C	R&H
rs4446909	0.3817	0.0023	0.1616	0.1645
rs148520524	1	0	-0.0323	-0.0327
rs5989681	0.68	0.0019	-0.1327	-0.1345
rs56690322	1	0	-0.0383	-0.0389
rs6644635	0.4864	0.0027	-0.1537	-0.1557
rs17149149	-	-	-	-
IVS4-225C>A	1	0	-0.0323	-0.0327
rs6588807	0.7251	0.0017	-0.1013	-0.1027
rs186757787	1	0	-0.0323	-0.0327
rs148855515	-	-	-	-
rs184583293	-	-	-	-
rs28675287	1	0	-0.1228	-0.1245
rs191866206	-	-	-	-
IVS5+124 G>T	-	-	-	-
rs77697639	-	-	-	-