Investigation of DNA methylation at the promoter region of the aralkylamine N-acetyltransferase (AANAT) gene in South African children with Autism Spectrum Disorder

Gerrit van Wyk (VWYGER004)

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

Submitted: April 2017

Supervisors: Dr Colleen O'Ryan and Ass Prof Laura Roden

Department of Molecular and Cell Biology

National Research Foundation

Plagiarism Declaration

I, Gerrit van Wyk, know the meaning of plagiarism and declare that all the work in this dissertation, save for that which is properly acknowledged, is my own.

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

Signature

Table of Content

Figures

Tables

List of Abbreviation

AANAT	Aralkylamine N-acetyltransferase	**NCBI**	National Center for Biotechnology Information
ADOS-2	Autism Diagnostic Observation Schedule, Second Edition	**PCR**	Polymerase chain reaction
ASD	Autism Spectrum Disorder	**PPVT**	Peabody Picture Vocabulary Test
ASMT	N-Acetylserotonin O-methyltransferase	**RNA**	Ribonucleic acid
BCL-2	B-cell lymphoma 2	**RNase-A**	Pancreatic ribonuclease A
CpG	Cytosine nucleotide followed by a guanine	**RORA**	RAR-related orphan receptor alpha
CSQ	Children's Sleep Questionnaire	**r_p**	Pearson's r-value
DMSO	Dimethyl sulfoxide	**RRB**	Restricted repetitive behaviour
DNA	Deoxyribonucleic acid	**SA**	Social affect
DNAme	Methylated deoxyribonucleic acid	**SAM**	S-Adenosyl methionine
dNTP	A nucleoside triphosphate	**SAP**	Shrimp alkaline phosphatase
ELISA	Enzyme-linked immunosorbent assay	**SCN**	Suprachiasmatic nucleus
GO	Gene ontology	**SNP**	Single nucleotide polymorphism
IQ	Intelligence Quotient	**TF**	Transcription Factor
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight	**WASO**	Wake after sleep onset
MBD	Methyl-CpG-binding domain protein	**ZIC1**	Zinc finger of the cerebellum
MEIS1	Homeobox protein Meis1	**ZID**	Zinc finger protein with interaction domain
MTNR1A	Melatonin receptor type 1A	**6-OHMS**	6-Hydroxymelatonin Sulphate
MTNR1B	Melatonin receptor type 1B		

Glossary of Terms

ADOS-2: is the Autism Diagnostic Observation Schedule, second edition which is an observational assessment of ASDs and is a semi-structured, standardized assessment of communication, social interaction, play, and restricted and repetitive behaviours. This assessment presents an individual with a number of activities that elicit behaviours directly related to a diagnosis of ASD.

Autism Spectrum Disorder (ASD): Autism Spectrum Disorder is a group of neurodevelopmental disorders that present with deficits in two core symptoms. Firstly, deficits in social communication and social interaction and secondly, restricted, repetitive patterns of behaviour, interests or activities.

Actigraphy: the use of an actigraph to record the movements of an individual, usually to determine the individual's rest-activity cycles. An actigraph is a small unit, worn for a week or more to measure gross motor activity.

Actimetry: the measurement of the rest-activity cycles of an individual.

Circadian rhythm: the endogenous rhythm ensuring the regulation of a roughly 24-hour cycle of biological processes in animals and plants.

Entrainment: the process of aligning a circadian rhythm's period and phase to the period and phase of an external rhythm.

Epigenetic: The modification of a phenotypic trait that is not the results of changes in DNA sequence. It is typically heritable changes in gene expression that results from DNA methylation, histone modification and non coding RNA .

Epoch: a precise and repeating instant of time that is used to measure movements in actimetry.

Hypomethylation: a decrease in the epigenetic methylation of cytosine residues in DNA.

In silico: a computer simulation of an actual event or process.

Monotonic: a monotonic function is a function which is either entirely nonincreasing or nondecreasing. For example, an individual's height will not increase in adulthood even as the individual's age increases.

Parasomnias: a disorder characterized by abnormal or unusual behaviour of the nervous system during sleep that could involve abnormal movements, emotions, perceptions and dreams that

occur while falling asleep, sleeping, between sleep stages, or during arousal from sleep.

Promoter region: a region of DNA that initiates transcription of a particular gene.

Restricted repetitive behaviour: this includes stereotyped or repetitive motor movements; stereotyped speech or repetitive speech patterns and stereotyped use of objects.

Sleep-wake behaviour: the length, timing, and/or rigidity of the sleep-wake cycle relative to the day-night cycle.

Sleep latency: the length of time required for an individual to accomplish the transition from full wakefulness to sleep.

Social affect: this a summary category of the ADOS-2 algorithm. If consists of a number of communication scores (such as conversation and the use of conversational / informative gesture) and reciprocal social interaction scores (such as unusual eye contact, directed facial expressions, shared enjoyment and reciprocal social communication).

Symptomatology: the collective symptoms of a patient or a disease.

Typical developmental profile: a description of a wide range of skills children acquire similar to the majority of children their same age within their same culture.

Waking after sleep onset (WASO): the amount of time an individual has spent awake after initially falling sleep and before they awaken for good.

Acknowledgements

I would like to thank my supervisors, Dr Colleen O'Ryan and Dr Laura Roden, for the opportunity to pursue my studies under their guidance, and both of them have contributed financially and intellectually to the completion of this dissertation. My lab partners from MARSA, Michaella Hulley, Marita de Waal and Bianca Abrahams helped immensely in trouble shooting so many troubles or just keeping morale high. Dr Aron Abera, at Inqaba Biotec, deserves mention for the very supportive role he played in setting up the EpiTyper assay and allowing us to participate in completing one assay run. Dr Reshma Kassanjee, Dr Anneli Hardy and Dr Miguel Lacerda from the University of Cape Town's Department of Statistics assisted in testing assumptions and suggesting alternative ways to analyse data. Acknowledgement also goes to Dr Ramona Hurdayal and Dr Sofia Stathopoulos for taking the time to read drafts and making suggestions. Finally, I have to thank the National Research Foundation (NRF) for awarding me an NRF Scarce Skills scholarship, which made it possible to support myself and to complete my studies.

Abstract

Sleep problems and suppressed melatonin production commonly presents with core Autism Spectrum Disorder (ASD) traits. Aralkylamine N-acetyltransferase (AANAT) acts as the penultimate and rate-limiting enzyme in the melatonin biosynthetic pathway, and a study by Hu *et al.* (2009) reported that *AANAT* expression was suppressed in an ASD population with severe language impairments. The mechanism responsible for this suppressed expression is unknown. Therefore, the aim of the study was to investigate the genetic and epigenetic features of *AANAT* in a cohort of South African children with ASD versus children with typical development in combination with a melatonin production study to explore melatonin's contribution to ASD symptomatology. It was expected that meeting this aim would reveal DNA methylation (DNAme) modifications were statistically significant different between case and control participants. Alternatively, that DNAme features would correlate with distinct ASD traits or sleep problems and/or altered melatonin production in case participants.

Biological samples and phenotypic data were collected from boys, aged between 6 and 14 years old who were assessed with the Autism Diagnostic Observation Schedule (ADOS-2). The promoter region and gene body of *AANAT* was sequenced (case n=26, control n=26) and DNAme analysis was performed with the Epityper massARRAY system (case n=19, control n=20). Urinary 6-hydroxymelatonin sulphate (6-OHMS) was quantified with an enzyme-linked immunosorbent assay (case n=4, control n=4). The 6-OHMS investigation was complemented with actigraphy data and a description of sleep behaviour as determined by an abbreviated version of the Children's Sleep Questionnaire.

Sequence analysis found no novel single nucleotide polymorphisms and no significant differences between case and control participants. In contrast, a difference (p=0.014) in DNAme at the third CpG site in the promoter region (CpG 3) was identified in case participants assessed with ADOS-2 Module 1 in comparison to case participants assessed with ADOS-2 Modules 2 -and 3. In particular, hypomethylation was more common in participants assessed with Module 1 which is the module

used to assess participants with little or no speech abilities. The transcription factor (TF) binding motifs for ZID (zinc finger protein with interaction domain), MEIS1 (Meis homeobox 1) and ZIC1 (zinc finger protein of the cerebellum 1) were identified at or near to CpG 3. These three TFs have known gene ontology terms that relate to neurodevelopment. The age of participants did not correlate with DNAme, and no further statistical significant differences were identified between the DNAme features of case and control participants, nor the correlation analysis of DNAme and ASD traits in case participants. No Module 1 participants volunteered for the 6-OHMS study, and it was therefore not possible to confirm whether DNAme features at CpG 3 correlated with altered melatonin production.

The data from the study suggest that hypomethylation at the promoter region of *AANAT* may be related to speech impairment in ASD, and that epigenetic investigations can uncover molecular underpinnings that correlate to ASD symptomatology. Furthermore, the current study addresses the paucity of molecular information on ASD in Sub-Saharan Africa and thereby contributes to a comprehensive understanding of disease biology. It remains unknown if hypomethylation at *AANAT* also correlates with suppressed melatonin synthesis in ASD individuals with speech impairments and this need further investigation.

1. Introduction

1.1 Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterised by two core traits. The first is persistent deficits in social communication and social interaction across multiple contexts. Secondly, affected individuals show restricted and repetitive patterns of behaviours, activities and interests (American Psychiatric Association 2013). The combination of these two traits present in various permutations and in varying degrees of severity that manifest in a spectrum of symptoms. While some individuals with ASD can lead fulfilling and independent lives, the condition can impact severely on the quality of life for others (Happé 1999; Farley *et al.* 2009).

ASD is commonly diagnosed with observational assessments of behavioural traits combined with a full developmental history. Typically, a range of assessments is used to characterise, and determine the severity, of the ASD in an individual. The Autism Diagnostic Observation Schedule (ADOS-2) is an example of one such standardised assessment and uses structured, and semi-structured tasks that require social interaction between a test participant and an examiner (Gotham *et al.* 2009; Duda *et al.* 2014; Wiggins *et al.* 2015). The examiner observes an individual's behaviour and assigns score values to the individual's responses. These scores are processed with an algorithm that produces scores that reflect the deficits in the core ASD domains; the first being the Social Affect (SA) score which indicates the severity of deficits in social communication and interaction, and the second being the Restrictive and Repetitive Behaviour (RRB) score (Lord *et al.* 2000; Gotham *et al.* 2007). The ADOS-2 can be administered to individuals with various developmental profiles, with different modules suited to the differing verbal abilities or age of an individual. Module 1 is used to assess individuals that have no or very little phrase speech abilities. Module 2 is used to assess individuals that use limited phrase speech, Module 3 is used to assess young children that are verbally fluent and Module 4 is used for verbally fluent adolescent to adults (Lord *et al.* 2000).

While various tools are available to diagnose or describe ASD symptoms accurately, the aetiology of ASD remains unknown. However, it has been reported that a heterogeneous biology most likely

underpins the equally heterogeneous traits of ASD (Newschaffer *et al.* 2007; Lauvin *et al.* 2012; Gu *et al.* 2013). Furthermore, ASD appears four times as often in males than females, and the underpinning biology for this bias is also unknown (American Psychiatric Association 2013). This bias has resulted in the presentation and biology of ASD being better defined and studied in males than it is in females. It has also been proposed that the aetiology for female individuals with ASD might be distinct to that of male individuals (Baron-Cohen 2004; Sarachana & Hu 2013; Madden & Zup 2014; Schaafsma & Pfaff 2014).

The lack of reliable data on the prevalence of ASD in developing countries have prompted author to highlight the urgent need to increase research and to better characterise ASD in these parts of the world (Bakare & Munir 2011; Elsabbagh *et al.* 2012). Indeed, most genetic research focuses on populations of the Northern hemisphere, while populations from regions like Sub-Saharan Africa remain under-represented (Popejoy & Fullerton 2016; Bustamante *et al.* 2011). This includes South Africa, where only few studies have investigated the molecular dimension of ASD (Esau *et al.* 2008; Arieff *et al.* 2010). The bias towards populations in the Northern hemisphere has implications for a comprehensive understanding of disease biology and entrenches the benefit of genomic medicine to a privileged minority (Popejoy & Fullerton 2016; Bustamante *et al.* 2011).

1.2 Sleep-wake behaviour and melatonin in ASD

Sleep problems commonly present as a secondary symptom of ASD (Wiggs & Stores 2004; Limoges *et al.* 2005; Tordjman *et al.* 2005; Allik *et al.* 2006; Malow *et al.* 2006; Richdale & Schreck 2009; Giannotti *et al.* 2011; Kotagal & Broomall 2012; American Psychiatric Association 2013). It has been reported that 44% to 83% of children with ASD show irregular sleep behaviour, a frequency that is much higher than seen in children with other developmental disorders (Singh & Zimmerman 2015). Specifically, children with ASD will have difficulty falling asleep and/or staying asleep, daytime sleepiness, and even parasomnias (Wiggs & Stores 2004; Limoges *et al.* 2005; Richdale & Schreck 2009; Singh & Zimmerman 2015; Yang *et al.* 2016).

Poor sleep-wake behaviour can compound certain ASD symptoms, such as anxiety, repetitive behaviour, depression and reduced attention span. Furthermore, children with ASD who have poor sleep behaviour place additional strain on parent-parent and parent-child relationships (Wiggs & Stores 2004; Malow *et al.* 2006; American Psychiatric Association 2013; Singh & Zimmerman 2015). Parent-child relationships are the primary support system upon which children with ASD depend, and these relationships can ill-afford extended stressors (Bekhet *et al.* 2012). As such, investigating sleep-wake behaviour and its underlying biology in children with ASD is an important field of exploration.

Rhythmic secretion of melatonin contributes to establishing sleep-wake behaviour, as this hormone is best known for its role in the entrainment of circadian rhythms to the light-dark cycles of the day (Serón-Ferré *et al.* 2001; Ackermann *et al.* 2006; Rossignol & Frye 2011; Jenwitheesuk *et al.* 2014; Ruggeri *et al.* 2014). More precisely, the suprachiasmatic nucleus (SCN) in the hypothalamus plays a major role in entraining circadian rhythms to the light-dark cycles of the day. First, neural signals from the retina lead to a synchronisation of the circadian clocks of the SCN to external light-dark cycles. This central synchronisation is then communicated to the rest of the body via neural signalling to other hypothalamic nuclei and the regulation of melatonin secretion from the pineal gland (Serón-Ferré *et al.* 2001; Korkmaz *et al.* 2009). Of note, melatonin only transduces the onset of darkness to tissues that lack photoreceptors, and this melatonin signal then elicits a corresponding change in cell and tissue physiology.

Melatonin secretion undergoes daily fluctuation, with the concentration surging from basal levels at dusk, peaking in the hours after midnight and decreasing to basal conditions at dawn. Light stimuli via the eye to SCN then lead to the maintenance of melatonin concentration at basal levels throughout the day (or however long the eye perceives light signals)(Pandi-Perumal *et al.* 2006; Korkmaz *et al.* 2009). Ultimately, this fluctuation in melatonin helps to time and maintain sleep-wake cycles. However, factors like disruptions in sleep behaviour and exposure to light at night can alter melatonin production and requires consideration when investigating melatonin's daily

fluctuations. The Children's Sleep Questionnaire (CSQ) and actigraphy are two tools often used to describe and assess children's sleep behaviour. While the CSQ relies on parents' reporting on their children's sleep behaviour, actigraphy directly measures the physical activity of a test participant. The recorded actigraphy allows for an objective description of a participant's sleep behaviour, making it well suited to measuring the circadian rhythms of children (Acebo *et al.* 1999; Hering *et al.* 1999; Littner *et al.* 2003; Ancoli-Israel *et al.* 2003). Actigraphy is also the least disruptive form of sourcing sleep behavioural data from research participants, as the actimetry device is small enough to be worn on the wrist. On the other hand, the CSQ is used as a standardised tool to screen for the most common paediatric sleep disorders (Owens *et al.* 2000) and provide descriptive information of each participant's sleep behaviour. In addition, the CSQ can establish whether actimetry data collected from a participant was indicative of typical sleep behaviour, as transient experiences could have altered a participant's typical sleep behaviour during the collection of actigraphy data. For example, factors like illness, exams, or stressful events at home may alter typical sleep behaviour while actigraphy was recorded. Moreover, a number of studies that have investigated the sleep-wake behaviour in ASD have relied on these two tools (Hering *et al.* 1999; Wiggs & Stores 2004; Goodlin-Jones *et al.* 2008; Krakowiak *et al.* 2008).

Apart from regulating sleep-wake behaviour, melatonin also contributes to neurodevelopmental processes. For example, melatonin impacts on the development and functioning of the cerebellum (Rosenstein & Cardinali 1986; Esparza *et al.* 2005; Imbesi *et al.* 2008). Concurrently, studies have explored the role of the cerebellum in a range of population groups (Kassubek *et al.* 2004; Lidzba *et al.* 2008; Stoodley & Schmahmann 2009), including the ASD population (Hodge *et al.* 2010). These studies report that the cerebellum plays an important role in language production, while Tordjman *et al.* (2012) report that irregular melatonin levels often coincide with individuals with ASD that have speech impairments. Therefore, irregular melatonin synthesis or secretion can contribute to establishing aberrant neurological development besides sleep problems.

A number of studies have investigated melatonin regulation and biosynthesis in ASD populations

(Tordjman *et al.* 2005; Melke *et al.* 2008; Jonsson *et al.* 2010; Tordjman *et al.* 2012; Pagan *et al.* 2014). These studies focussed on determining suppressed or elevated concentrations, or arrhythmic fluctuations of melatonin production (Leu *et al.* 2011; Tordjman *et al.* 2012), the administration of supplemental melatonin (Galli-Carminati *et al.* 2009; Rossignol & Frye 2014), or the sequence identity of genes involved in melatonin biosynthesis in individuals with ASD (Melke *et al.* 2008; Chaste *et al.* 2010). Rare sequence variants at the gene acetylserotonin O-methyltransferase (*ASMT*) in particular have been implicated in suppressed melatonin synthesis in individuals with ASD (Toma *et al.* 2007; Melke *et al.* 2008; Pagan *et al.* 2014). To date, only Jonsson *et al.* (2010) have investigated multiple genes in the melatonin biosynthesis and signalling pathway in a single study. This included the genes aralkylamine N-acetyltransferase (*AANAT*) and *ASMT* in the melatonin synthesis pathway, and melatonin receptor type 1A (*MTNR1A*) and melatonin receptor type 1B (*MTNR1B*) in the melatonin signalling pathway. The DNA sequence of these genes were analysed in 109 ASD patients, and rare sequence variants were identified at a splice sites of ASMT, protein coding region of *MTNR1B* and promoter regions of *ASMT*, *MTNR1A*, and *MTNR1B*. These rare sequence variants were present at frequencies that were statistically significantly different to those of the general population, while the no statistically significant difference in the sequence variants for *AANAT* were found. However, the Jonsson *et al.* (2010) study was limited to a Swedish population, and importantly, did not correlate melatonin production to the rare sequence variants identified in *ASMT*, *MTNR1A*, and *MTNR1B*. Failing to investigate the relationship between sequence variants and melatonin production also meant that it was not determined if other molecular mechanisms, besides sequence variants, could have contributed to altered melatonin production. In particular, epigenetic modification also have the ability to alter melatonin production, and all epigenetic modifications require distinct experimental analysis. The aims of the current study, presented later, outlines the first attempt to address the limitations presented above.

1.3 Epigenetics

Lipsky (2013) compares the impact of epigenetic modifications to a process of programming and reprogramming versions of the genome, as epigenetic modifications suppress or enhance gene

expression. Epigenetic modifications also explain how environmental stimuli alter some phenotypic outcomes, and thus provide a basis for the plasticity and the diversity seen in many organisms (Goldberg *et al.* 2007; Berger *et al.* 2009; Gibney & Nolan 2010; Felsenfeld 2014). In addition, endogenous processes also evoke epigenetic alterations that impact on typical development and daily biological regulation (Lubin *et al.* 2005; Saito *et al.* 2012; Namihira & Nakashima 2013). Currently, research has identified three types of epigenetic modifications: DNAme, chromatin remodeling and non-coding RNA associated gene silencing. Each modification provides broad opportunities for investigation, although DNAme is the best studied modification (Mehler 2008; Grafodatskaya *et al.* 2010; LaSalle *et al.* 2013; Jensen 2014).

DNAme occurs when DNA methyltransferases transfer a methyl group from S-adenosyl methionine (SAM) to the fifth carbon of a cytosine residue. However, DNAme is restricted to cytosine residues that precede a guanine residue, known as a CpG (Mehler 2008; Laird 2010). CpG sites in the genome are unusually rare, and these sites are often concentrated in regions referred to as CpG islands. The promoter regions in 72% of genes overlap with CpG islands, and these islands also serve as a feature in identifying promoter regions (Antequera & Bird 1993; Saxonov *et al.* 2006). Therefore, DNAme studies have focussed on investigating CpG sites in promoter regions of candidate genes (Nagae *et al.* 2011; Lin *et al.* 2012; LaSalle *et al.* 2013; Lokk *et al.* 2014; Hranilovic *et al.* 2016) as DNAme can serve as a foundation for altering transcriptional regulation.

Once the cytosine residue at a CpG is methylated, it acts as a binding signal to a range of proteins, but predominantly to the poorly characterised Methyl-CpG-binding domain (MBD) protein family (Baymaz *et al.* 2014). The MBD proteins are responsible for altering DNA transcription by recruiting additional proteins that block or stimulate transcription factors, or stimulate or inhibit chromatin packaging (Hendrich & Bird 1998; Hendrich & Tweedie 2003). Therefore, DNAme does not change typical DNA transcription *per se*, but forms the basis for subsequent processes that alter the regulation of gene transcription.

Given that the process of methylating DNA is initiated by cell signalling, the pattern of methylated DNA in one cell may differ from the methylated pattern of neighbouring cells (Nagae *et al.* 2011; Gordon *et al.* 2012). That is, the reach of cellular signals that evoke the processes that methylate DNA will not always cover every cell in the targeted tissue. Therefore, the MBD proteins that bind to methylated cytosine residues could contribute to suppressed or enhanced gene expression in cells in some regions of tissue while the remainder of cells in the tissue remain unaffected. To reflect this difference in DNAme status, the DNAme status of tissue is expressed as a percentage (Nagae *et al.* 2011; Gordon *et al.* 2012). For example, 20% DNAme indicates that only 20% of cells in a tissue are methylated at the CpG site under investigation. This relationship between cell signalling and the creation of different DNAme patterns explain why various tissue types have distinct DNAme patterns in comparison to other tissue types. It has been reported that the DNAme patterns of buccal cells more closely resemble the DNAme patterns of neural tissue than blood (Iwamoto *et al.* 2011; Fernandez *et al.* 2012; Lokk *et al.* 2014). It is hypothesised that the similarity in buccal and neural DNAme patterns stems from the fact that both tissue types develop from the same germ layer - the ectoderm (Lowe *et al.* 2013; Lokk *et al.* 2014). Therefore, the DNAme patterns of buccal cells are thought to be a good proxy for the DNAme patterns of neural tissue from living individuals, as access to post-mortem brain tissue is very limited and these brain samples will not always meet the required demographic or phenotypic criteria of a study.

There are numerous techniques available to explore DNAme, and the length of DNA under investigation is one determining factor to selecting an appropriate technique (Laird 2010). To investigate short lengths (≤500 bp) two options are currently the most accessible in South Africa. These are either DNA sequencing based techniques, or the Agena Epityper massARRAY (mass spectrometry) approach. The massARRAY system relies on the treatment of DNA with bisulphite that deaminates cytosine residues to uracil residues, while methylated cytosine residues remain protected from conversion. Single stranded amplicons are then amplified from the bisulphite-converted DNA, and the amplicons cleaved with RNase-A to produce numerous fragments. The molecular weights of the fragments are then determined with mass spectrometry, and the extent of

DNAme can then be calculated from the recorded mass spectra. Conversely, determining DNAme with sequencing techniques rely on establishing the change in residue identity from unconverted DNA to bisulphite converted DNA. The comparison of bisulphite converted DNA to native DNA sequence identity therefore serves as the means to determine the presence, absence, and extent of DNAme.

1.4 Epigenetics and melatonin synthesis

A study by Hu *et al.* (2009) unexpectedly created a focus on *AANAT* as a gene that may contribute to suppressed melatonin production. This RNA microarray study quantified gene expression levels in lymphoblastoid cell lines from individuals with ASD, and partitioned participants according to distinct phenotypic traits. This partitioning strategy reduced the confounding phenotypic heterogeneity in ASD and revealed stronger correlations between gene expression and particular traits. In fact, several ASD studies that have relied on partitioning strategies have reported enhanced clarity in correlating phenotypic traits to genotypic features. These included strategies to partition research participants based on language abilities (Alarcón *et al.* 2002; Chen *et al.* 2006), or repetitive behaviour (Hollander *et al.* 2000). In the study by Hu *et al.* (2009), *AANAT* and 14 other circadian genes either showed suppressed or elevated expression in an ASD population with severe language impairments, and the authors suggested that *AANAT* may be implicated in the hypomelatonemia and sleep problems often reported in ASD cases. The Hu *et al.* (2009) study did not investigate whether the suppressed gene expression was correlated with melatonin synthesis, nor were the mechanisms responsible for the suppressed expression of *AANAT* investigated.

1.5 Research question

The aim of the current study was to characterise genetic and epigentic features of *AANAT* in a South African ASD cohort for comparison to a control group and to determine which molecular features correlated with distinct phenotypic traits in the case participants. A further aim was to compare the melatonin production and sleep-wake behaviour in case and control participants.

The objectives of the study included:

1. A gene sequence analysis of the promoter region and gene body of *AANAT* to confirm that rare occurring SNPs or mutations did not alter the sequence identity of CpG sites TF binding motifs, splice sites and protein-coding regions.

2. Quantifying DNAme at the CpG island that overlaps with the promoter region of *AANAT* to determine the DNAme profile of each DNA sample collected.

3. Determining the concentration of urinary excreted 6-hydroxymelatonin sulphate (6-OHMS) for comparison between study and control groups, as 6-OHMS serves as a reliable proxy of melatonin production (Nowak *et al.* 1987; Griefahn *et al.* 2003).

4. Determining and comparing the sleep-wake behaviour of study and control participants.

To reduce the confounding factor of ASD heterogeneity, case participants were also partitioned into phenotypic subgroups for additional comparisons. These partitioning strategies relied on ADOS-2 modules used to assess participants and individual SA and RBB scores. Molecular features that showed significant differences were then intended for comparison of melatonin concentrations and sleep-wake behaviour.

The study hypothesised that analysis of DNAme profiles would either reveal molecular differences between ASD and control participants, or DNAme features that correlate with distinct ASD trait in the case participants. In addition, DNAme features could also correlate with sleep problems and/or the altered melatonin production often seen in children with ASD.

Determining the DNAme features that correlate to phenotypic traits will broaden the understanding of melatonin's role in ASD symptomatology and demonstrate that investigations into underrepresented populations can contribute to a global understanding of ASD.

2. Materials and methods

2.1 Cohort

Buccal swabs and urine samples were collected from children with ASD and children with a typical developmental profile for the control group. Participants for the ASD (case) group had a prior, independent ASD diagnosis and were recruited from two specialist autism schools, while participants for the control group were recruited from sports clubs and one mainstream school. All participants were from the Cape Town metropolitan area. Relevant ethics approval was obtained prior to recruitment of participants (Western Cape Government, Ref: 2014002-37506; University of Cape Town Ethics, FSREC 076-2014), and heads of schools provided consent to visit their schools before recruitment occurred at respective schools. Posters, letters and oral presentations were used to recruit and explain the proposed research to parents and teachers. The parents of all children who participated in the study signed a consent form (Appendix A) and participants' personal identifiers were coded to ensure anonymity. The principal investigator, Dr Colleen O'Ryan and postdoctoral fellow, Dr Sofia Stathopoulos (University of Cape Town) completed the recruitment of study participants. Both Dr O'Ryan and Dr Stathopoulos are accredited ADOS-2 administrators to Research Reliability level and completed the ASD phenotyping using ADOS-2 assessments.

All participants were male and between the age of 6 and 14 years old and had no known genetic, or physiological comorbidities that would confound the interpretation of data. For example, two participants with Fragile X syndrome were excluded from the study. The ADOS-2 was administered in English as all the participants spoke and were instructed at school in English. Where siblings were recruited, the data from only one sibling was used in analysis.

Participants who met the above criteria were assessed with the ADOS-2 to verify their suitability as participants in the case or the control group. Case group participants all scored in the SA and RRB domains of the ADOS-2, and the combined SA and RRB scores allowed an overall classification status of "Autism Spectrum Disorder Mild" or "Autism". No participants in the control group scored

in SA or RRB categories of the ADOS-2.

2.2 Buccal swab collection and extraction of DNA

DNA was extracted from buccal cells collected with buccal swabs (Epicentre Catch-All™ Sample Collection Swabs; Madison, Wisconsin, USA). The inside of each cheek was swabbed for a minimum of 30 s, avoiding contact with children's teeth. The swab tips were immediately submerged in lysis buffer, stored at 4 °C and DNA was extracted within four days of swabbing. Children who refused to provide a buccal cell sample were excluded from the remainder of the study.

DNA was extracted from buccal cells using a standard protocol (Aljanabi & Martinez 1997), with the following modifications: the final sodium dodecyl sulphate concentration of the sterile salt homogenising buffer was reduced to 1%, and the proteinase K incubation period was extended to 3 h. For the salt precipitation step, the final concentration of the NaCl was reduced to 6 mM, and the subsequent centrifugation time reduced to 10 minutes, but the relative centrifugal force increased to 16 000 x*g*, and this step was repeated once. After adding ice cold isopropanol, the samples were left at -20 °C overnight rather than 1 h, and the subsequent centrifugation time was reduced to 10 min, but the relative centrifugal force increased to 16 000 x*g*. Finally, DNA was re-suspended in 40 µl distilled H_2O, rather than 300 µl-500 µl.

2.3 Sequencing *AANAT*

A region of 1003 bp upstream from the start of *AANAT* exon 1 to 134 bp downstream from the end of *AANAT* exon 4 (Figure 1) was sequenced for all participants, which produced a total sequence length of 4 kb. The 4 kb DNA length was divided into six regions, and each region overlapped by a minimum of 100 bp. The primer pairs for all six regions are indicated as A to F below (Table 1), and all primers were designed with PerlPRimer Version 1.1.21 (http://perlprimer.sourceforge.net/: June

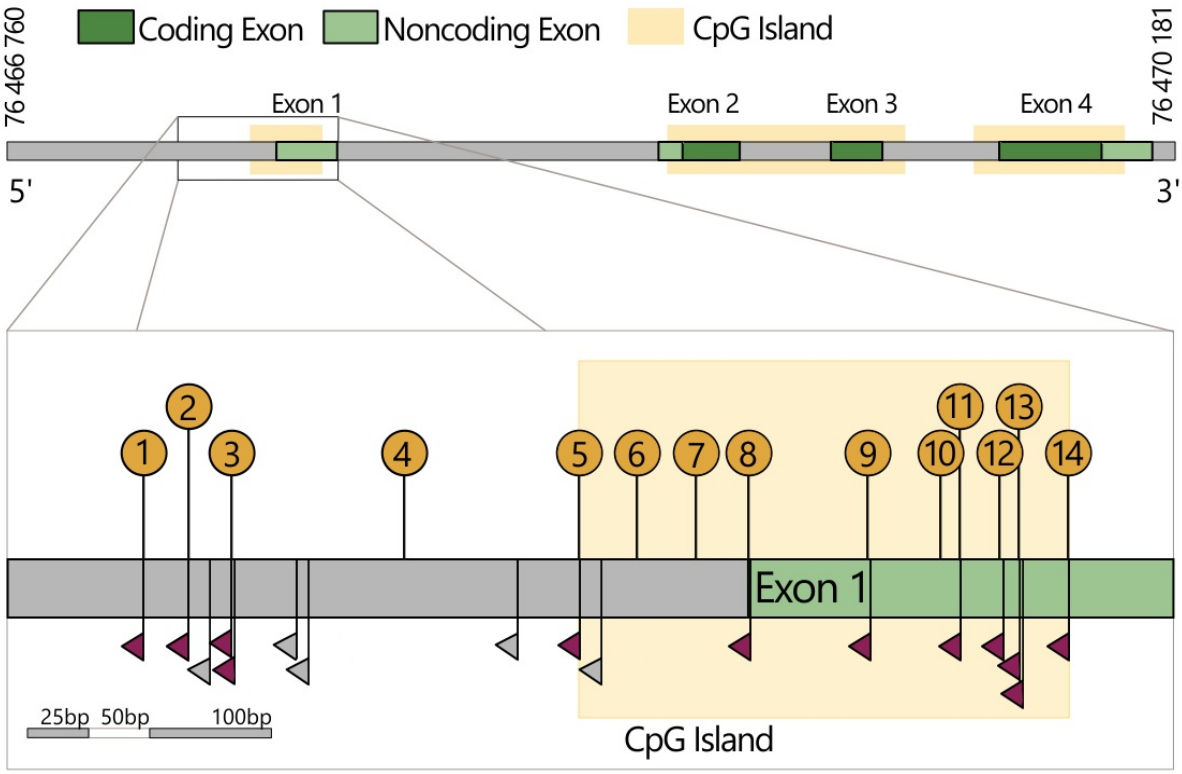


Figure 1. Overview of *Homo sapiens AANAT*, on chromosome 17. Magnified area indicates the 14 CpG sites investigated in the current study (each site numbered in dark yellow circles). Known and rare SNPs that overlap the CpG sites also shown (all inverted flags), while inverted burgundy flags indicate SNPs that overlap CpG sites. Drawing to scale, adapted from NCBI databank (http://www.ncbi.nlm.nih.gov/gene/15: February 2015).

2014). A set of primers were also designed to specifically amplify the DNAme region of interest, and is indicated as primer pair CGI in Table 1.

Table 1. Primers to amplify *AANAT*

Reg[a]	Forward Primer	Reverse Primer	Amp[b]
A	5'-CGAAAGAGTGCACCTGTGAG-3'	5'-GTCCTGGAGGTTGGATTTCAC-3'	426
B	5'-AGGACAGGAAGGACCCTCAG-3'	5'-GGCTGTAGTAGCAGATGTTCCC-3'	801
C	5'-CAGTCAGCCTCTGGTAATCCC-3'	5'-GAGATGGTGGCAGGGATGAG-3'	953
D	5'-AGGATCTTGCACCCAAAGGAC-3'	5'-ACACATGGTCTGATACTCTGGG-3'	857
E	5'-AGGAGGACACTTCCAAAGCTG-3'	5'-GAAGGTCTTGCTCTGGGAGG-3'	784
F	5'-GAGACTCATGCAGGTGAGGA-3'	5'-GAGGAAAGGAGCCAGGAACC-3'	936
CGI	5'-GTAGTCAGAGCCAACAGG-3'	5'-GCTCCAAGTAGAGAAGCTCCT-3'	745

[a] Region, [b] Amplicon size (bp)

The PCR cycling conditions for all primer sets in Table 1 followed an initial denaturing cycle at 98 °C for 1 min, a cycle denature step at 98 °C for 20 s, a final extension step at 72 °C for 10 min, and a final hold cycle at 22 °C for 3 min. The cycle annealing temperature, cycle extension time, as well as the number of cycle repeats were as follow: Primer Set A: 62 °C, 20 s; 72 °C, 13 s with 35 repeats. Primer Set B: 66 °C, 20 s; 72 °C, 24 s with 35 repeats. Primer Set C: 65 °C, 20 s; 72 °C, 25 s with 40 repeats. Primer Set D: 62 °C, 20 s; 72 °C, 25 s with 35 repeats. Primer Set E: 66 °C, 20 s; 72 °C, 23 s with 35 repeats. Primer Set F: 65 °C, 20 s; 72 °C, 24 s with 40 repeats. Primer Set CGI: 62 °C, 20 s; 72 °C, 20 s with 35 repeats.

The amplifications were carried out in 20 µl reaction volumes according to ThermoFisher Scientific's Phusion High-Fidelity DNA Polymerase kit (Waltham, Massachusetts, USA). The final concentrations for reagents were 1x GC buffer, 0.5 mM forward primer, 0.5 mM reverse primers, 3% Dimethyl sulfoxide (DMSO), 200 µM deoxynucleotide (dNTP) mix, 0.02 $U.\mu l^{-1}$ Phusion DNA polymerase, and 0.2 $ng.\mu l^{-1}$ DNA. Amplification of Region C resulted in low yields because of sequence features that created secondary structures, therefore the volume for these reactions were doubled, while reagent

concentrations remained unchanged.

PCR amplicons were loaded onto 1% (w/v) agarose gels, electrophoresed at 100 V for 1 h in Tris base, acetic acid and EDTA (TAE) buffer and DNA was visualised after ethidium bromide intercalation under longwave UV light. The correct sized amplicons were excised and DNA was purified with the Zymoclean™Gel DNA Recovery Kit (Irvine, California, USA). Purified DNA was resuspended in 37 µl distilled H_2O and 20 µl of the DNA was dispatched for Big Dye sequencing at The Centre for Analytical Services (CAF) at the University of Stellenbosch.

CAF sequenced amplicons with the Big Dye Terminator V3.1 sequencing kit (Applied Biosystems) with slight modifications to the manufacturer's protocol. Amplicons were run with the ABI 3730x, while data collection and sequence analysis were performed with the AB Foundation DATA Collection v3.0 and Sequence Analysis v5.3, respectively. Results were supplied electronically in text format and as chromatograms. The text format of amplicon sequences were aligned with Mega6 (http://www.megasoftware.net/: October 2015) to the *AANAT* reference sequence NM_001088.2 (http://www.ncbi.nlm.nih.gov/gene/15: February 2015).

Potential SNPs in the DNA sequences were scored as heterozygous only if two peaks overlapped at a particular position on the chromatogram. In addition, the overlapping peaks had to appear on the chromatograms generated from both the forward and reverse primers. Alternatively, potential SNPs were scored as homozygous when a residue in the sequences generated from both forward and reverse primers did not match the reference sequence. The identity of potential SNPs at *AANAT* were determined by comparing base pair positions of potential SNPs to the position of described SNPs in the National Centre for Biotechnology Information (NCBI) SNP Data Bank.

2.4 DNA methylation profile determination by sequencing of bisulphite converted DNA

The first attempt at describing DNAme patterns was by sequencing bisulphite converted DNA. A 657 bp region that included *AANAT* exon 1 and the promoter region were amplified, followed by cloning and subsequent sequencing. First, 1000 ng DNA was bisulphite converted with Zymo's EZ DNA Methylation™ Kit (Irvine, California, USA), according to manufacturer's instructions. The 657 bp region was amplified with the following primers: Forward primer 5'-GGAATGTTAGTATTAAGTGTTAAGGTT-3'; reverse primer 5'-AAATCCCACCTACCTAAACTATAAA-3'. The primers were designed as described in section 2.3. The cycling conditions were: one initial denaturing cycle at 95 °C for 15 min, cycle denaturing step at 95 °C for 30 s, annealing step at 58 °C for 45 s, and extension step at 72 °C for 50 s. The cycle denaturing, annealing and extension steps were repeated 40 times. This was followed by a final extension step at 72 °C for 15 min, and final hold cycle at 22 °C for 3 min. Amplicons were generated with KAPA DNA polymerase according to manufacturer's instructions to a final volume of 25 µl. The final concentrations for reagents were 1x Taq buffer, 0.4 mM forward primer, 0.4 mM reverse primers, 0.4 mM dNTP, 0.02 $U.\mu l^{-1}$ KAPA Taq DNA polymerase, and 1.5 $ng.\mu l^{-1}$ bisulphite converted DNA. Amplicons were electrophoresed and retrieved from agarose gel by the same procedure described in section 2.3.

Amplicons generated from bisulphite converted DNA were cloned using the ThermoFisher Scientific CloneJET PCR Cloning Kit (Waltham, Massachusetts, USA), according to manufacturer's instructions. However, the volumes of all reagents for ligating amplicons into pJET1.2 blunt vector were halved. The *E.coli* strain used for transformation was DH5α. As the original PCR amplicons were thymine rich, the transformed *E.coli* were grown on Luria broth agar plates supplemented with carbenicillin to a final concentration of 100 $\mu g.ml^{-1}$, at 30 °C for 24 hours. Subsequent colony PCR was performed with primers that were included in the CloneJET PCR Cloning Kit: forward primer 5'-CGACTCACTATAGGGAGAGCGGC-3'; reverse primer 5'-AAGAACATCGATTTTCCATGGCAG- 3', and to cycling conditions specified by the manufacturer. Amplicons generated from five different DH5α colonies were sequenced with Sanger Big Dye sequencing as described in section 2.3.

2.5 Determining DNA methylation profile with EpiTyper massARRAY analysis

2.5.1 *In silico* preparation for massARRAY analysis

All potential CpG islands at *AANAT* were identified with MSP Primer, (http://www.mspprimer.org/cgi-bin/design.cgi: May 2015) (Figure 1), with the sequence entry NM_001088.2 for *AANAT* (http://www.ncbi.nlm.nih.gov/gene/15: February 2015). The range of CpG sites identified for investigation were limited to the CpG island at *AANAT* exon 1 and the four CpG sites immediately upstream from this CpG island (Figure 1).

Inqaba Biotec performed the Agena Epityper massARRAY analysis and designed primers to flank the DNAme region of interest with Sequenom EpiDesigner BETA. The design of the reverse primer included a T7-promoter tag, which is shown in lowercase below, and the forward primer included a 10-mer balancing tag sequence also shown in lowercase:

Forward Primer: 5'-aggaagagagATAAGAGGTGGGTTTGTTTAAGATT-3'; reverse primer: 5'-cagtaatacgactcactatagggagaaggctATTCCTAAAAACTAAAAACCAAAAA-3'.

The primers indicated above produced a 516 bp amplicon that included the 14 CpG sites in the DNAme region of interest. Inqaba Biotec's initial analysis with Agena EpiTyper software indicated that all 14 CpG sites were located on fragments that would produce detectable mass signals after enzymatic digestion. An independent additional analysis was also performed with an *in silico* prediction of expected mass spectra with the amplicon analysis function in the MassArray package of R (Suchiman *et al.* 2015). This *in silico* prediction revealed that only eight CpG sites were located on fragments that would have unique molecular weights (Figure 2). It also showed that the specific sequence at CpG 12 and CpG 13 would result in the production of a single fragment after enzymatic digestion that would contained both CpG sites (Figure 2). Therefore, massARRAY would record only one mass signal for CpG 12 and CpG 13, and this data point was represented as CpG 12-13 in all subsequent figures and tables. The remaining six CpG sites were located on fragments

with molecular weights identical to those of random fragments. The seven CpG sites with unique molecular weights were CpG 3, CpG 4, CpG 6, CpG 7, CpG 9, CpG 12-13 and CpG 14. Therefore, only the massARRAY data collected from these seven CpG sites were used in subsequent DNAme analysis.

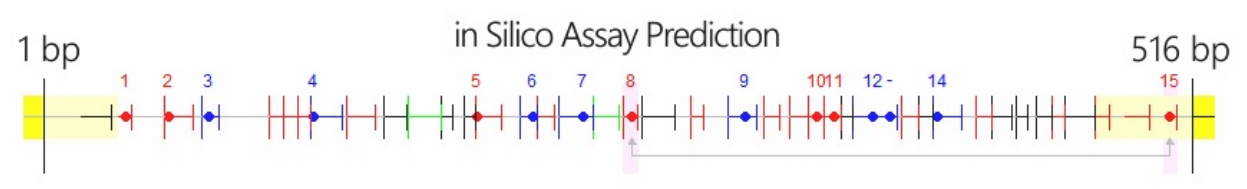


Figure 2. *In Silico* prediction of fragments produced from the PCR amplicon after enzymatic digestion with numbers indicating each CpG site. Numbers correspond to CpG positions indicated in Figure 1. Blue numbers indicate the CpG sites located on fragments that would produce unique mass spectra. Red numbers indicate which fragments would produce mass signals identical to other, random fragments. Fragments coloured in green indicate fragments that could serve as bisulphite conversion controls. Regions shaded in light yellow indicate the sequence of the primers proposed to amplify the region under investigation. Note that CpG 15 was an artefact created by the sequence of the primer.

Finally, *in silico* prediction indicated that enzymatic digestion would produce amplicon fragments that could serve as DNA bisulphite conversion controls. That is, some fragments would contain all the attributes required to calculate if the DNA from which the original amplicons were generated was also 100% bisulphite converted (Figure 2).

2.5.2 EpiTyper massARRAY assay

Inqaba Biotec was provided with 1000 ng DNA from each participant for bisulphite conversion with Zymo's EZ DNA Methylation™ Kit (Irvine, California, USA). In addition, Inqaba included a 1000 ng of 100% methylated DNA, and 1000 ng of 0% methylated DNA in the bisulphite conversion to serve as DNAme assay standards. These DNAme assay standards were used to create standard curves for each CpG site in order to normalise all recorded data generated with the massARRAY system. T7-RNA polymerase was used to amplify the DNAme region of interest from the bisulphite converted

DNA samples. RNA amplicon reactions were treated with shrimp alkaline phosphatase (SAP) to remove unincorporated dNTPs. A T-cleavage solution was added to the SAP treated samples, incubated and then treated with a cleaning resin to eliminate salts that may interfere with mass spectrometry measurements. A robotic system dispensed the RNase-A treated amplicons onto a silicon matrix preloaded SpectroCHIP, and a MassARRAY Compact MALDI-TOF (Agena Bioscience, California, USA) collected the mass spectra. Finally, the EpiTYPER software v1.2 (Agena Bioscience, California, USA) generated the spectra's methylation ratios and represented these in csv-format on Excel (Microsoft, USA) worksheets.

2.6 Investigation of urinary excreted 6-OHMS

A subset of research participants volunteered for the 6-OHMS investigation, and parents signed an additional consent form for this part of the study (Appendix B). Participants for the 6-OHMS investigation adhered to additional basic inclusion criteria. First, participants were not using medication that altered their sleep behaviour, or altered melatonin secretion (such as beta blockers). In addition, participants were not taking melatonin supplements and slept in complete darkness as exposure to light at night suppresses the production of melatonin (Zeitzer *et al.* 2000). The collection procedure was verbally explained to parents and they received a written summary of the procedure (Appendix C).

2.6.1 Collection and processing of urine samples

Each participant voided their bladders in a dedicated 1 L container between 18h00 and the participant's bedtime. A second dedicated 1 L container was used to capture total bladder content after waking up the next day, or any passing of urine after 03h00. Parents recorded the time of urine collection by marking-off the time and date on a label that was fixed onto the sampling bottles. This process was repeated for a second evening-morning period, with parents determining

the days to collect samples. Some parents also requested reminders to perform the collections and therefore received reminders via a WhatsApp message. Samples were stored at 4 ºC at participants' homes after urine collection, and parents were provided with polystyrene cooler boxes and ice packs to transport samples to collection points. From the collection points, samples were taken directly to the laboratory and processed immediately, or stored at 4 ºC for processing within 24 hours. All urine samples were collected over a time period of one week and the case group participants' samples were collected from March 2016 to July 2016. The samples from one control group participant were collected in February 2016 and the remaining control group participants' data were collected between August 2016 and October 2016.

Once samples were collected, the date, time of collection and volume of urine samples for each participant were documented, and 15 ml of urine was removed from the 1 L containers. The 15 ml samples were centrifuged for 20 min at 4 ºC at 2540 x*g*, from which a 1.5 ml aliquot was removed for analysis with an enzyme-linked immunosorbent assay (ELISA). Both the 1.5 ml aliquot, and the remainder of the 15 ml samples were then stored at -20 ºC. Only the 1.5 ml aliquot was later removed and thawed to room temperature to perform the ELISA assay. Urinary 6-OHMS was measured using the melatonin-sulphate urine ELISA kit (IBL International, Hamburg, Germany), according to manufacturer's instructions.

2.6.2 Actigraphy and CSQ

Actigraphy was recorded with the Respironics Actiwatch 2 (Philips, Amsterdam, Netherlands), actimeter which included a light sensor to record ambient light levels. The recorded illumination data was used to confirm that participants slept in darkness on the evenings that urine samples were collected. To ensure the recorded actigraphy was representative of the wearer's sleep behaviour, actiwatches were worn for at least five days (Acebo *et al.* 1999) and preferably for one week. The actiwatches were set to measure data over a seven day period and collected data in 15 s

epochs, and all recorded actigraphy coincided with the urine sample collection period. Instructions on how to use the Actiwatch were provided to parents verbally and in a written summary (Appendix C).

The author and one other investigator independently analysed the collected actigraphy with the Respironics Actiware version 5.7 software (Philips, Amsterdam, Netherlands). Prior to analysis, each investigator examined the data to ensure the behavioural period identified by the software reflected plausible events. For example, participants were asked to remove the Actiwatch when playing sport, and the software would often interpret this absence of activity as an indication that the wearer was asleep. Therefore, the investigator manually adjusted the onset time point the software assigned to a behavioural period. The adjustments made by the investigators were compared to confirm that adjustments did not differ by more than five minutes, and differences greater than five minutes were re-examined together to reach a consensus. Once there was agreement on all the time points, the Respironics Actiware software analysis was performed and the results were then used in statistical analysis.

A parent of each participant completed the CSQ prior to collecting actigraphy to ensure parents did not adjust their responses to match the behaviour during the actigraphy recording period.

2.7 Statistical analyses

Statistical analysis was performed with R (version 3.2.5) and RStudio (version 0.99.893), (https://www.r-project.org/ and https://www.rstudio.com/: 12 April 2016). The p-values were considered significant when $p < 0.05$ and nearing statistical significance when the p-value was ≥ 0.05 and ≤ 0.1.

Differences in SNP frequencies were analysed with Fisher's exact test. Only data recorded for the eight CpG sites with unique mass signals were included in analysis of DNAme. In addition, DNAme samples had to meet several criteria for inclusion in statistical analysis: only samples with technical repeats and with corresponding values within a 10% range of each other were included in analysis. Samples were also included only if more than 70% of CpG sites in each sample produced mass signals. Finally, the bisulphite analysis function in the MassArray package in R (with supplementing script as developed by Suchiman *et al.* (2015)) had to confirm 100% bisulphite conversion of the DNA used to generate amplicons.

The DNAme data (Appendix Figure 1) was treated as nonparametric given that it was not possible to determine a distribution pattern with the small sample sizes (Pallant & Manual 2007; Yazici & Yolacan 2007; Razali & Wah 2011). All box plots indicated the median and interquartile range of 25^{th} and 75^{th} percentile values. Statistically significant differences between groups were determined with the Mann-Whitney U test. Given that the Mann-Whitney U test detects significant differences in either median values, or distribution patterns of data points (Hart 2001; Zimmerman 2004), equality of variance tests were also performed to distinguish between median and distribution patterns. Equality of variance was determined with Levene's test (Lim & Loh 1996) and a p-value greater than 0.05 was interpreted as two groups that had a similar distribution patterns. Conversely, a p-value less than 0.05 was interpreted as statistically significant differences in the distribution of data points between two groups.

Scatter plots were used in the DNAme analysis to investigate possible linear or monotonic relationships between two variables. The strength of linear correlations were determined with Pearson's r-value (r_p) and Spearman's rho-value (ρ) for potential monotonic relationships. Correlation analyses can be sensitive to outlier data points (Aguinis *et al.* 2013), however no outliers were removed in the analysis because there was no *a priori* biological justification to exclude

outliers.

2.8 Additional software used for analysis

Where analysis indicated a statistically significant difference at a CpG site, the sequence identity at the relevant CpG site was further explored to determine the presence of transcription factor (TF) binding motifs near to, or overlapping the CpG site. Transcription binding motifs were identified with m2match platform (www.gene-regulation.com: 15 August 2016) and also to graphically represent the TF binding motifs relative to the DNA sequence at *AANAT* in sequence logo format. Finally, the gene ontology terms associated with TFs were investigated with UniProt (http://www.uniprot.org: 16 August 2016).

The MassArray package in R was used to perform mass spectrometry analysis, and the ggplot2 package in R was used to generate visual representation.

3. Results and analysis

3.1 DNA sequence analysis

DNA samples from 26 case participants and 26 control participants were included in the sequence analysis to determine the sequence variance at *AANAT* (Table 2). Complete sequences of the 4 kb region of interest were generated from all 52 samples, and all sequences were aligned to the *AANAT* reference sequence NM_001088.2 (NCBI, 2016) to identify potential SNPs and mutations.

Table 2. Summary of the age, ancestry of, and ADOS-2 Module used to assess participants that were included in the DNA sequence analysis at *AANAT*

	Age[a]	**Ancestry**			**ADOS-2 Module**		
		African	**European**	**Mixed**	**Mod 1**[b]	**Mod 2**[c]	**Mod 3**[d]
Case n=26	8.35 ±2.34	3	8	15	15	8	3
Control n=26	8.20 ±1.35	2	6	18	n/a	n/a	26

[a] Average age in years with standard deviation

[b] Module 1, little to no phrase speech abilities

[c] Module 2, limited phrase speech abilities

[d] Module 3, fluent speech

A total of 14 SNPs were identified from 22 and 19 participants in the case and control groups respectively. All 14 SNPs were previously described in the NCBI SNP database, and only rs61739395, represented by two and three case and control participant respectively, resulted in sequence identity change in the protein coding region (Threonine to Methionine). All the SNPs were present at frequencies below 0.1 except for rs3760138 and rs4238989, which were both present in the 5'UTR

(Appendix Table 1). The NCBI SNP data bank indicated that rs3760138 and rs4238989 both present at distinct frequencies in different population groups. Demography and ancestry can influence SNP allele frequencies and this has been reported for *AANAT* (Sekine *et al.* 2001; Blomeke *et al.* 2008; Koike *et al.* 2013). Although this study was not designed as a population level SNP study, participants were partitioned according to ancestry to explore trends in SNP frequencies. An obvious caveat to this is the small sizes for each population group (Appendix Table 1). Examining the two informative SNP's (rs3760138 and rs4238989) in the mixed ancestry group, there was no statistically significant difference (Appendix Table 2) between case and control participants.

3.2 DNA methylation analysis

DNA sequencing of the bisulphite converted *AANAT* promoter region was performed, subsequent to PCR amplification and cloning, to determine this region's DNAme profile. However, after successful bisulphite conversion and DNA sequencing, chromatograms consistently showed signal ambiguity in comparison to chromatograms generated from native DNA (Appendix Figure 2). It is not clear why sequencing bisulphite converted DNA with the Sanger Big Dye platform consistently produced irregular results, but potentially depurination and non-specific degradation of the DNA followed by inaccurate DNA strand repair during PCR may have produced irregular sequence results. Given the failure of the sequencing, the Epityper massARRAY analysis was used to determine the DNAme profile of research participants.

Given that earlier analysis depleted DNA samples for some participants, DNA samples for 27 case and 21 control participants were prepared for massARRAY analysis (of which 11 and 15 case and control group samples were also included in the SNP analysis). One control group sample failed to produce any data, and eight case group samples failed to produce mass signals for more than 70% of CpG sites. As such, 19 case group and 20 control group samples produced reliable data for DNAme analysis (Table 3). Native DNA sequence analysis of these 39 reliable samples showed that

all samples shared an identical sequence identity at the DNAme region of interest. There were no SNPs present at any of the 14 CpG sites analysed. Bisulphite conversion analysis with the MassArray package in R also showed 100% bisulphite conversion for all DNA samples.

Table 3. Summary of the age, ancestry of, and ADOS-2 module used to assess participants in the DNAme study

	Age[a]	**Ancestry**			**ADOS-2 Module**		
		African	**European**	**Mixed**	**Mod 1**[b]	**Mod 2**[c]	**Mod 3**[d]
Case n=19	8.66 ±1.91	3	5	11	10	4	5
Control n=20	8.65 ±1.06	0	1	19	n/a	n/a	20

[a] Average age in years with standard deviation

[b] Module 1, little to no phrase speech abilities

[c] Module 2, limited phrase speech abilities

[d] Module 3, fluent speech

Standard curves were produced from the assay standards (0% and 100% methylated DNA) to normalise the data recorded from the research participants in spite of technical repeats not being performed for the DNAme assay standards. (Appendix Figure 3). Statistical analysis was performed with the normalised and unnormalised data concurrently, and the results produced from the normalised data were identical to unnormalised data to three decimal points.

3.2.1 DNA methylation profile at *AANAT*

Only the seven CpG sites with unique mass spectra were included in analysis, and recorded data showed that the median DNAme values were different at each CpG site (Figure 3). Although the

DNAme pattern across CpG sites were variable, both case and control groups shared the same variable DNAme pattern. Indeed, only comparison of case and control median values for CpG 14 showed a statistically significant difference ($p=0.014$). However, the recorded range of data for CpG 14 overlapped with the range of data for all the 100% DNAme assay standards. For example, the range of DNAme for the 100% DNAme assay standards ranged from 89% to 98%, while that of CpG 14 ranged from 90% to 96% (Appendix Table 3, Appendix Figure 3). Given that the massARRAY system has a ±5% margin of error (Ehrich *et al.* 2005), CpG 14 may have been 100% methylated in a number of case and control samples. Thus, this site was therefore excluded from further statistical analysis.

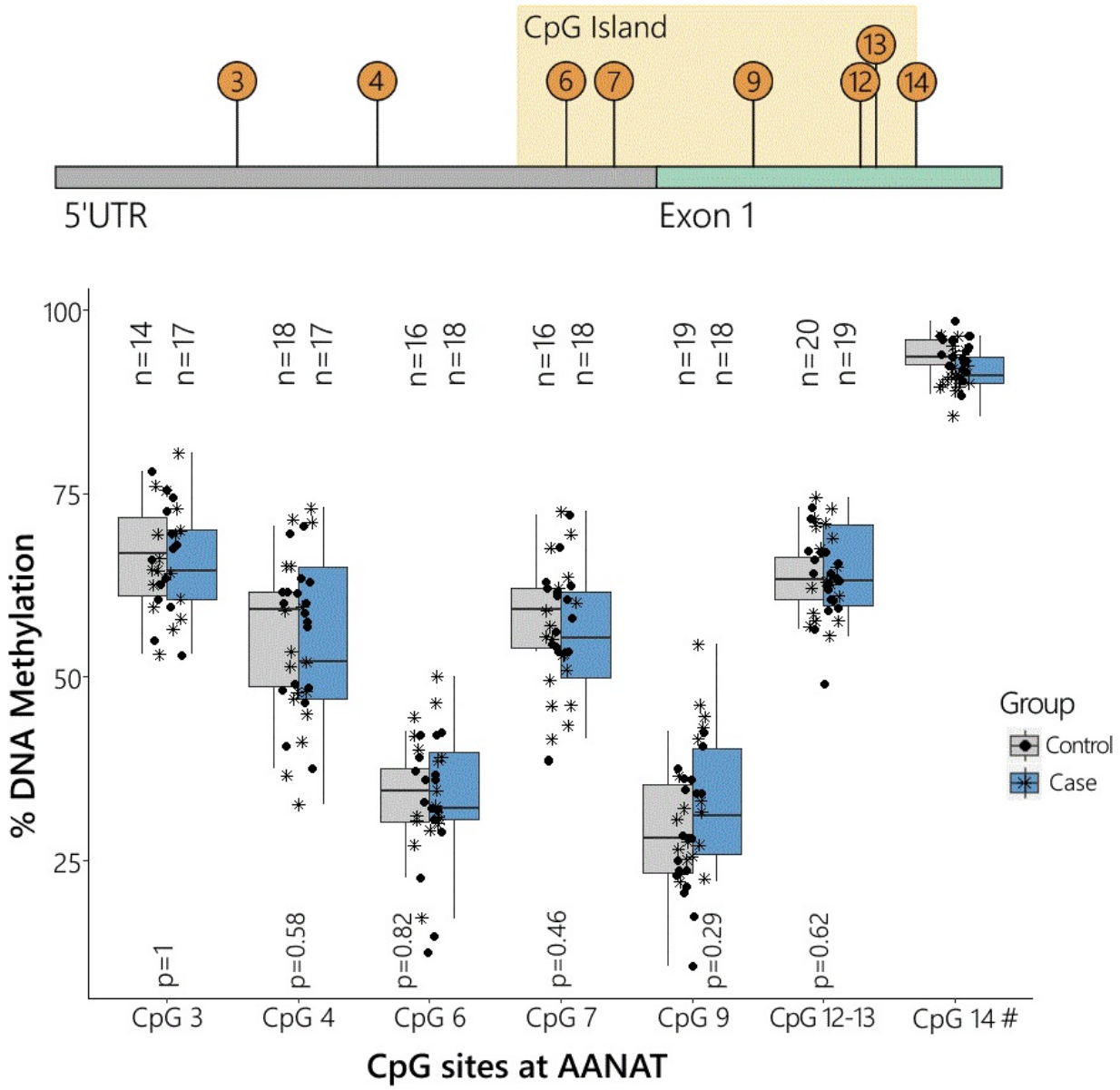


Figure 3. Box plot summary of median, 25th and 75th percentile values for DNAme at eight CpG sites with unique mass spectra at the promoter region of *AANAT* for case and control group. Key at top indicates the position of the eight CpG sites relative to each other. Grey boxes and solid dots indicates control group, while blue boxes and stars indicate the case group. P-values perpendicular to the x-axis were calculated with Mann-Whitney U. Hash indicates CpG site excluded from further statistical analysis.

Finally, there was no correlation between DNAme and the age of participants, which implies that age was not a confounding factor to interpreting data (Appendix Figure 4). This was applicable to

both case and control participants.

3.2.2 Differences in DNA methylation patterns and phenotypic traits within the case group

The analysis of massARRAY data presented in Figure 3 suggested that no statistically significant differences existed between the DNAme profile of the case and control groups. However, ASD is not a homogenous phenotype, and partitioning an ASD population into phenotypic subgroups can enhance the likelihood of finding a link between the underpinning aetiology and a corresponding phenotypic trait. The first strategy was to partition case participants according to the ADOS-2 module used for assessment and then performing statistical comparisons between modules (see Table 3 for composition of participants assessed with each module). All permutations of module pairings were used to perform data comparisons with Mann-Whitney U and Levene's tests. Comparing data from Module 1 with the combined data of Module 2 and Module 3 indicated a statistical significant difference in the medians for CpG 3 (Figure 4). Though Module 2 and Module 3 are distinct categories, combining these two categories represented participants with fluent- to limited-speech as opposed to participants with no, or non-communicative speech abilities. All other permutations of module comparisons indicated no statistically significant differences (Appendix Table 4).

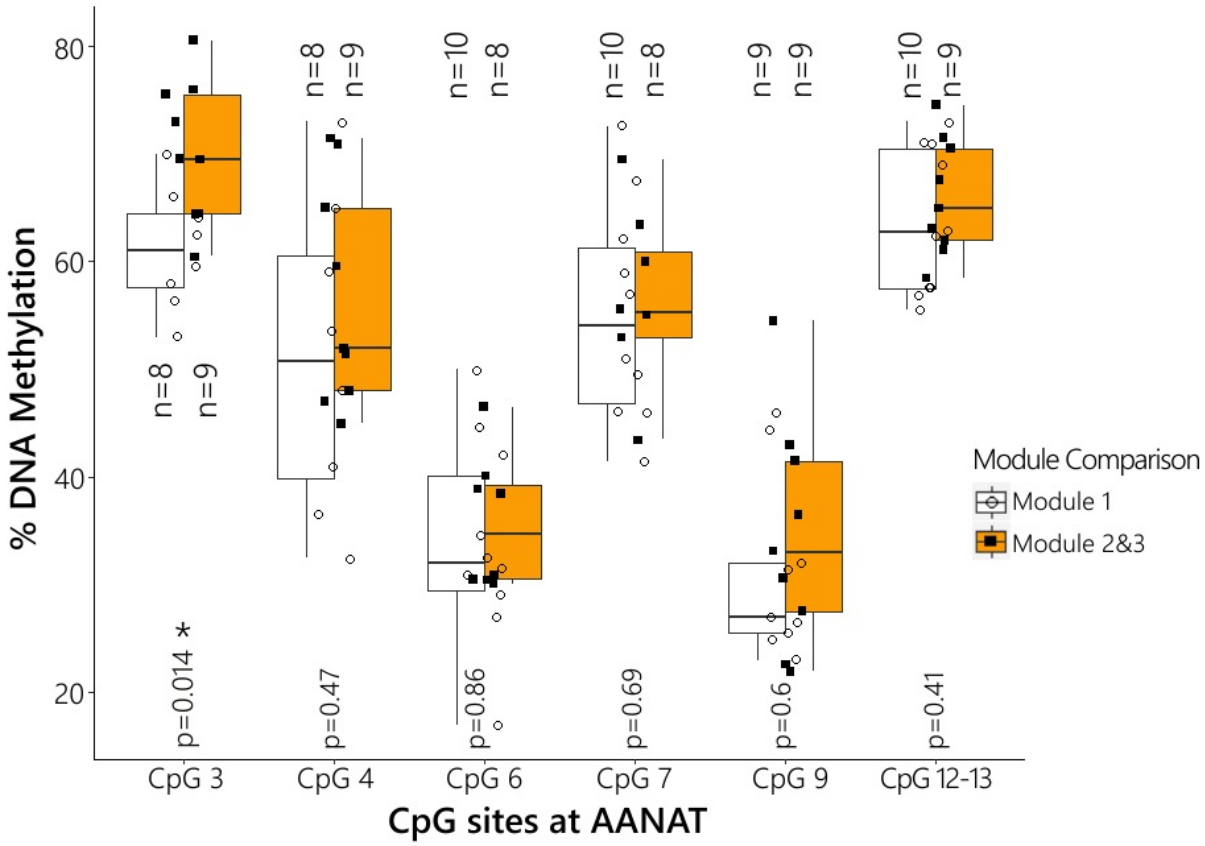


Figure 4. Box plot comparison of median, 25th and 75th percentile values at each CpG site with white and clear circles indicating ADOS-2 Module 1 subgroup while orange and solid squares indicate ADOS-2 Module 2 combined with ADOS-2 Module 3 subgroup. P-values calculated with Mann-Whitney U, and asterisk indicates statistically significant p-value.

For the second partitioning strategy, case group participants were grouped according to sub-scores of the ADOS-2 module, and massARRAY data was then correlated with SA and RRB scores, respectively. However, correlation analysis indicated no statistically significant correlation trends (Appendix Table 5).

A number of additional partitioning strategies were also explored (Appendix Table 4) but analysis was constrained by small sample sizes. None of these comparisons indicated statistically significant differences potentially due to the great differences between the n-values of each comparison.

Therefore, the calculated p-values should be interpreted with caution.

3.2.3 Transcription factor motifs at CpG 3 showed links to neurodevelopment

The analyses of Figure 4 suggested that hypomethylation of CpG 3 correlated with reduced language abilities. A reduction in DNAme may be implicated in lowered MBD proteins activity at CpG 3. In turn, altered MBD proteins activity is thought to alter TF binding activity at, or near CpG sites (Hendrich & Bird 1998; Hendrich & Tweedie 2003). Therefore, the TFs that could bind near to, or at CpG 3 were identified.

First, putative binding motifs for TFs near to, or at CpG 3 were identified with the TF motif prediction algorithm at m2match.com, which identified the binding motifs for transcription factors ZID (zinc finger protein with interaction domain), MEIS1 (Meis homeobox 1) and ZIC1 (zinc finger protein of the cerebellum 1). The relationship between TF motifs and the DNA sequences at CpG 3 are shown in sequence logo format (Figure 5). Note that the large font letters in the sequence logo indicate strong affinity between the TF binding domain and the DNA residue, and smaller font letters indicate poorer affinities (Figure 5). In addition, m2match.com calculated a p-value of the likelihood a TF would bind at the proposed site. Only TFs with p-values below 0.1 were selected for further investigation.

After determining which TF might bind near to or at CpG 3, gene ontology (GO) terms for these TFs were investigated with UniProt annotation. Only the GO terms that described neurodevelopment and neural functioning were selected for presentation (Table 4). All the TFs predicted to bind near to, or at CpG 3 have known implications for neural development or functioning.

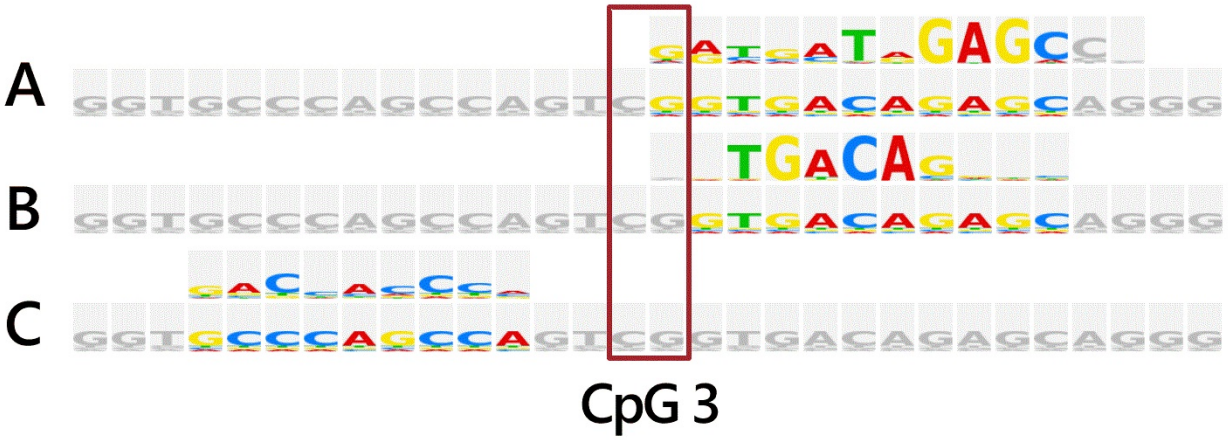


Figure 5. TF motifs matching sequences near to CpG 3. The red frame indicates the position of CpG 3. Figures in grey indicate no match between TF motif and sequence near to CpG 3. TF motifs are shown in top rows of each line and DNA sequence in bottom rows. A: Match for ZID binding motif and sequence overlapping CpG 3, p=0.032. B: Match for MEIS1 binding motif and sequence near to CpG 3, p=0.041. C: Match for ZIC1 binding motif and sequence near to CpG 3, p=0.051 (m2match,com, 2016).

Table 4. Summary of binding motifs for TFs that overlap with, or appear near to CpG 3 and UniProt annotation associated with these TFs

Transcription Factors	**UniProt annotation**
ZID	Regulation of transcription, widely expressed with highest levels in brain
MEIS1	RNA polymerase II core promoter proximal region sequence-specific DNA binding, negative regulation of neuron differentiation, involved in restless leg syndrome, expressed at high levels in the cerebellum, activation of anterior HOX genes in hindbrain development during early embryogenesis
ZIC	Involved in spinal cord development, cerebellar maturation, inner ear morphogenesis, spatial distribution of mossy fibre, high level of expression in the cerebellum

3.2.4 DNA methylation showed linear relationships between CpG sites in case and control group

Analysis of DNAme indicated a trend that had indirect relevance to substantiating or refuting the hypotheses of the current study. However, given that no past study investigated DNAme of the promoter region of *AANAT*, the trend of linear relationship between DNAme of CpG sites relative to each other warrants brief mention. This linear trend was apparent in both case and control participants (Figure 6). To explore the strength of this correlation, Pearson's r-values were calculated for each correlation (Table 5). A Pearson's r-value (r_p) equal to 1 indicates a perfectly positive linear relationship, while a value of -1 indicates a perfectly negative linear relationship. Values close to, or equal to zero indicate the absence of any correlation. Difference in r_p of case and control groups were explored with Fisher's Z-transformation and a subsequent Z-test to identify statistical significant differences in r_p. The Z-test results indicated that three correlations showed statistically significant differences (Appendix Table 6), and these three comparisons are highlighted in Table 5.

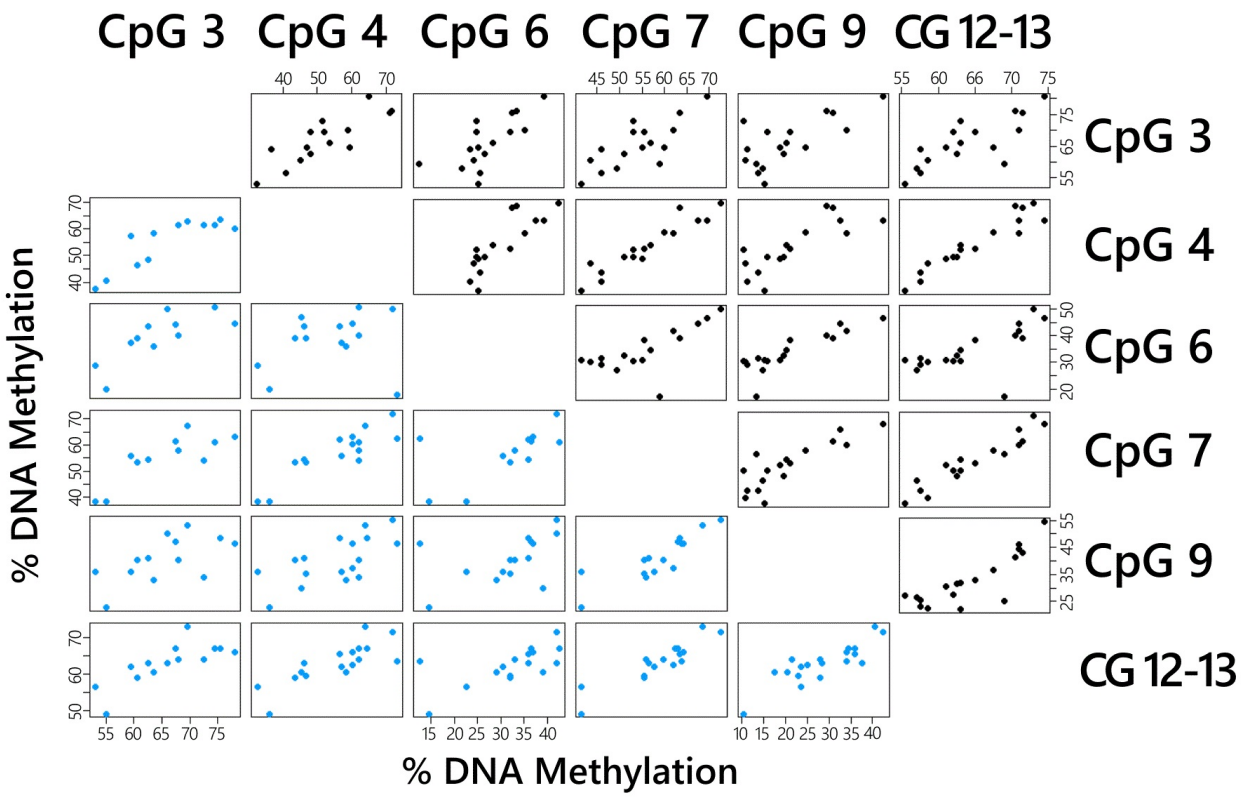


Figure 6. Correlation plot of DNAme at CpG site relative to each other. Blue dots represent individuals in the control group, black dots represent individuals in the case group.

Table 5. Summary of Pearson's r-value for CpG sites relative to each other, with blue numbers representing the control group, and black numbers the case group. Shaded values in the control group showed statistically significant differences to corresponding shaded values in case group

CpG 3	CpG 4	CpG 6	CpG 7	CpG 9	CpG 12-13	
•	0.836	0.710	0.778	0.702	0.779	**CpG 3**
0.850	•	0.845	0.947	0.810	0.951	**CpG 4**
0.774	0.233	•	0.701	0.863	0.660	**CpG 6**
0.805	0.859	0.604	•	0.847	0.954	**CpG 7**
0.612	0.665	0.487	0.867	•	0.842	**CpG 9**
0.768	0.791	0.663	0.916	0.852	•	**CpG 12-13**

3.3 Investigation of urinary excreted 6-OHMS

A subset of individuals (case group n=4 and control group n= 4) were recruited for a pilot investigation of urinary excreted 6-OHMS. Participants were matched for ancestry and age. However, no parents of participants who were assessed with the Module 1 version of the ADOS-2 agreed to participate (Table 6). A parent of each participant completed the CSQ, and all the participants successfully collected the urine samples and actigraphy data required for analysis. None of the participants had a cumulative CSQ score greater than 41, suggesting that these participants did not have subjective sleep problems (Owens *et al.* 2000). The ambient illumination data obtained from the actigraphy also indicated that all participants slept in complete darkness during the evenings urine samples were collected. The description of sleep behaviour captured with the CSQ corresponded to descriptions of sleep behaviour gathered from actigraphy, except that most parents in the case and control groups underestimated the time their children fell asleep by half an hour.

Table 6. Summary of age, ancestry and ADOS-2 module used to assess participants in 6-OHMS study

	Age[a]	**Ancestry**		**ADOS-2 Module**	
		European	**Mixed**	**Module 2**	**Module 3**
Case n=4	9.75 ±1.77	1	3	2	2
Control n=4	8.75 ±1.30	1	3	n/a	4

[a] Average age in years with standard deviation

The data for case and control subjects were compared with Mann-Whitney U and Levene's test for each category (Table 7). Great interpersonal variation for 6-OHMS was detected in both case and control participants, with no statistically significant difference between the groups (Table 7). However, the small sample size and absence of Module 1 participants also requires consideration.

In addition, case and control samples were collected during different season and therefore exposed to different periods of daylight, which might also have contributed to the variation seen in the 6-OHMS concentrations. However, the impact of seasonal difference on statistical analysis would have been much more limited if a greater number of participants volunteered for the 6-OHMS study. The time it took case and control participants to fall asleep (sleep latency) and the Waking at Night category of the CSQ neared statistical significance (Table 7). There were statistically significant differences in the variances of the case and control group's sleep efficiency and total time awake after initial sleep onset (WASO) (Table 7). Finally, according to information collected with the CSQ, none of the parents of children in the control group indicated that their children resist going to bed, while one and two parents of children in the case group indicated their children usually or sometimes resist going to bed respectively.

Table 7. Summary of the comparison of 17 categories in the 6-OHMS investigation.

	Control[a] n=4	Case[a] n=4	Mann-Whitney U p-value	Levene's Test p-value	Method
Bedtime (PM)	08:34 PM	09:03 PM	0.34	0.50	Actigraphy
Time Awake (AM)	06:11 AM	06:35 AM	0.11	0.49	Actigraphy
Time in Bed (h)	9:25	9:32	0.68	0.81	Actigraphy
Time asleep (h)	8:21	8:15	1	0.59	Actigraphy
Sleep latency (min)	5.41	18.41	0.057	0.46	Actigraphy
Sleep Efficiency (%)	88.69	87.66	1	0.01	Actigraphy
WASO[b] (min)	58.81	45.52	0.69	0.04	Actigraphy
No. of awakenings	58.81	47.37	0.69	0.77	Actigraphy
Age	9.5	8	0.35	0.59	CSQ
Overall CSQ Score	29	34.5	0.48	0.95	CSQ
Bedtime Behaviour	10	14	0.31	0.75	CSQ
Sleep Behaviour	9,5	8.5	0.23	0.76	CSQ
Waking at Night	1.5	4	0.079	0.67	CSQ
Wake-up behaviour	6	7	0.76	0.27	CSQ
PM 6-OHMS ($ng.h^{-1}$)	614.75	453.45	0.88	0.67	ELISA
AM 6-OHMS ($ng.h^{-1}$)	2954.18	2339.76	0.88	0.81	ELISA
Fold Δ[c] 6-OHMS	6.25	8.66	1	0.73	ELISA

[a] Median values shown in each column

[b] Time participant was awake after initial sleep onset

[c] Fold change in concentration of 6-OHMS from PM to AM

4. Discussion

The association of suppressed gene expression of *AANAT* with ASD was first mooted by Hu *et al.* (2009). This unexpected association was uncovered when the authors partitioned the ASD study group into specific subgroups and suppressed *AANAT* expression was reported in ASD population with language impairments (Hu *et al.* 2009). Gene sequence studies focusing on melatonin synthesis in an ASD population however failed to identify *AANAT* as candidate gene implicated in suppressed or elevated melatonin production (Melke *et al.* 2008; Jonsson *et al.* 2010; Pagan *et al.* 2014). Two factors may explain this apparent contradiction, with the first being Hu *et al.* (2009) employed a phenotypic partitioning strategy to reduce the confounding factor of ASD's heterogeneity. Secondly, epigenetic modifications may be the basis for altered gene expression rather than sequence variations. In order to reconcile these differing reports, the current study examined the genetic and epigenetic features at *AANAT* in combination with an investigation into melatonin production to explore melatonin's contribution to ASD symptomatology. This study tested the hypothesised that DNAme would be different between case and control participants. Alternatively, DNAme would correlate with distinct ASD traits, sleep problems and/or altered melatonin production.

In order to test these hypotheses, a number of challenges had to be addressed, and strategies adopted to enhance the accuracy of data included in analyses. Whereas most genetic studies focus on populations of the Northern hemisphere (Popejoy & Fullerton 2016; Bustamante *et al.* 2011), few molecular studies have been performed on ASD in South Africa (Esau *et al.* 2008; Arieff *et al.* 2010). Therefore, an ASD cohort had to be established to provide the molecular data and phenotypic detail required for analysis. Given the heterogeneity of ASD, a strategy was adopted to ensure that ASD participants were phenotyped using a research, standard assessment (ADOS-2) and that typically developing children (control participants) were screened for the absence of autistic traits. In addition, the epigenetic dimension of the study meant restricting participants to a narrow age group as it has been reported that epigenetic modifications can change with age (Maegawa *et al.*

2010; Teschendorff *et al.* 2010; Hernandez *et al.* 2011). The current study also opted to source DNA samples from buccal cells as these are considered a better proxy than blood for neural DNAme patterns in living individuals (Iwamoto *et al.* 2011; Fernandez *et al.* 2012; Lokk *et al.* 2014). However, the total DNA yields from buccal extractions are low and the molecular quality of DNA is poorer in comparison to DNA extracted from blood (Hansen *et al.* 2007). Finally, the technology required to investigate DNAme was new in the South African context, and the service providers that performed the massARRAY assay had to optimise protocols concurrent to executing experimental assays, given that we were their first client.

4.1 DNA sequence variation at *AANAT*

The current study hypothesised that DNAme would form the molecular basis for statistically significant differences at *AANAT* in the case and control participants. Therefore, the potential contribution of rare sequence variants to establishing significant differences between case and control participants had to be ruled out. Indeed, only two sequence variants were present at frequencies consistent with being SNP's (rs3760138 and rs4238989). Furthermore, partitioning research participants according to ancestry for rs3760138 and rs4238989 showed no differences between case and control. It was not possible to do any meaningful comparison among demographic groups because of small and unequal sample sizes. A number of studies have reported that genetic variation at *AANAT* also appears to be related to ethnic differences (Sekine *et al.* 2001; Soria & Martinez- 2010; Koike *et al.* 2013). However, to confirm the distinct SNP identity and frequencies of any South African population group will require a separate large scale investigation. This will present with challenges given South Africa's complex history of forced and free migration of population groups that included the Khoisan, Bantu, Dutch, British, East African and Asian populations. The allelic frequencies of all these populations thus contribute to the genetic profile of South Africa's population of mixed ancestry (de Wit *et al.* 2010; Warnich *et al.* 2011).

4.2 Differences in DNA methylation patterns and phenotypic traits within the case group

No statistically significant difference were found between the DNAme of case and control participants. However, partitioning case and control participants according to phenotypic subgroups indicated that hypomethylation of CpG 3 was more common in participants assessed with the Module 1 version of the ADOS-2 (Figure 4). This implied that language impairment correlated with DNAme at CpG 3. This is in line with the hypothesis that analysis of DNAme profiles would reveal molecular features that correlate with distinct phenotypic traits in the case participants.

Past studies have found statistically significant differences between the DNAme patterns of individuals with ASD and control participants, and these differences were often limited to specific *loci* (Nguyen *et al.* 2010; Ladd-Acosta *et al.* 2014; Nardone *et al.* 2014). For example, Ladd-Acosta *et al.* (2014) reported hypo- and hypermethylation differences in DNA samples extracted from post-mortem brain tissue of individuals that had ASD (n=20) and control samples (n=21). These differences in DNAme were detected at DNase hypersensitive sites, alternative transcript finish sites, promoter regions, gene bodies and alternative splice sites. The differences in DNAme were also common to the DNA collected from ASD sample tissue and thus not a rare occurrence. Nardone *et al.* (2014) reported that DNA samples extracted from post-mortem neural cortex tissue of individuals with ASD (n=13) and controls (n=12) also showed hypo- and hypermethylation differences in various DNA regions. More precisely, distinct differences in the DNAme patterns appeared between DNA samples collected from the Brodmann 10 and Brodmann 24 regions of control samples, while the DNAme patterns in ASD samples showed fewer differences between these two cortical regions. In addition, Nardone *et al.* (2014) demonstrated with real-time PCR that the observed differences in DNAme also resulted in overexpression in hypomethylated regions. Finally, Nguyen *et al.* (2010) investigated differences in the DNAme profile of lymphoblastoid cell lines of three monozygotic twin pairs where only one sibling in the pair had ASD. A number of DNA regions between twins showed differences in DNAme, and the *loci* that overlapped with the

genes retinoic acid-related orphan receptor alpha (*RORA*) and B-cell lymphoma 2 (*BCL-2*), were further investigated with immunohistochemical analysis. This analysis was performed on post-mortem tissue from the frontal cortex and cerebellum of case (n=5) and control samples (n=5), which showed decreased expression of RORA and BCL-2 proteins in case samples. Collectively the Nguyen *et al.* 2010 and Nardone *et al.* 2014 reports confirm that differences in DNAme do appear between ASD and control samples, and that these differences in DNAme can alter gene expression. The current study did not have the opportunity to investigate gene expression, however the difference in DNAme at CpG 3 in Module 1 case and control participants of the current study imply that there may be reason to investigate gene expression in future studies with CpG 3 as a focus *loci*. More specifically, functional analysis of DNAme at *AANAT* with pinealocyte samples (responsible for melatonin production) will have to confirm the impact that hypomethylation at the promoter region has on gene transcription. Even though reduced methylation of CpG 3 might lower MBD activity at this site, it is not clear if this will lead to an increase or decrease in TF activity (Hendrich & Bird 1998; Hendrich & Tweedie 2003). In addition, the MBD proteins that can target the CpG 3 region need to be characterised.

It is also interesting that the current study found that language impairment was the sub phenotype of ASD that correlated with hypomethylation of CpG 3. Tordjman *et al.* (2012) report that irregular melatonin levels often coincide with individuals with ASD that have speech impairments, while Hu *et al.* (2009) reported that *AANAT* expression was suppressed in an ASD population with severe language impairments. Furthermore, the current study identified that the binding motifs for the transcription factors MEIS1 and ZIC1 overlapped with, or appeared close to CpG 3, which have documented roles in cerebellar development and functioning (Table 4). This relationship with the cerebellum may prove significant as it has been reported that melatonin impacts on the development and functioning of the cerebellum (Rosenstein & Cardinali 1986; Esparza *et al.* 2005; Imbesi *et al.* 2008). It has also been reported that the cerebellum plays a role in language production in the general population, particularly in tasks like word stem completion, oral naming

speed and verbal fluency (Lidzba *et al.* 2008; Stoodley & Schmahmann 2009). Meanwhile, studies on cerebellar functioning in individuals with ASD have reported distinct differences of neural architecture of verbal and nonverbal boys with ASD (Hodge *et al.* 2010).

Nonetheless, the current study did not investigate the language ability of any of the research participants, except to use expressive language to determine which ADOS module to use for an assessment. Thus, the best surrogate to describing language abilities was the module of the ADOS-2 used to assess participants. Language is a complex phenotype (Bishop *et al.* 2006; Oliver & Plomin 2007; Newbury & Monaco 2010) which requires its own assessment criteria independent to the assessment of ASD phenotypes. Indeed, Hu and Steinberg (2009) selected participants with a Peabody Picture Vocabulary Test (PPVT) score <80 to confirm severe language deficits before partitioning research participants into the language deficit subgroup. This selection procedure preceded the gene expression study that indicated suppressed *AANAT* expression (Hu *et al.* 2009). Therefore, future studies can expand the phenotypic description of language abilities with the PPVT or other language tests to better characterise the relationship between melatonin production and speech ability.

Finally, there was great interpersonal variation in the DNAme status at each CpG site (Figure 3, Appendix Table 3). This was in contrast to the DNA sequence that spans CpG 1 to CpG 14, as all participants shared an identical sequence for this region of DNA. Therefore, despite sequence homogeneity, the DNAme profile established molecular diversity in this region for both case and control group participants. It also provides provisional support to the proposal that epigenetic modifications may contribute to the biological complexity of diseases, and that epigenetic investigations may yield insight where classic genetic studies have failed (Grafodatskaya *et al.* 2010; Hsieh & Eisch 2010; Heim & Binder 2012; Mohan & Chaillet 2012; Hu 2013; Jensen 2014; Keverne 2014; LaSalle *et al.* 2013; Valor & Guiretti 2014).

4.3 General comments regarding DNA methylation of *AANAT*

Correlation analysis unexpectedly indicated a linear relationship between CpG sites which is not straightforward to interpretation (Figure 6). This demonstrates that much is still unclear about DNAme patterns and associations and that epigenetics is still a relatively young field.

In particular, the same variable DNAme profile at exon 1 and the promoter region of *AANAT* was present in both case and control participants (Figure 3). The common occurrence of the profile implies that a similar DNAme process established the variable DNAme profile, while the variability suggests that a complex DNA methylation process established the profile. This complexity and similarity of the DNAme profile implies that the variable DNAme profile was essential to fulfilling a biological purpose. In addition, a positive correlation between the degree of DNAme for most CpG sites relative to each other was found (Table 5). When considering the high degree of positive correlation of DNAme with the variability of the DNAme profile, it is possible to make two proposals. First, the variable profile was established in a site-specific fashion. Second, once the overall DNAme profile was established, the overall profile was maintained and enhanced through a general DNA methylation process. Future studies will have to investigate DNA methylation mechanisms and cell signalling processes to allow clearer interpretation of DNAme features present at *AANAT*.

4.4 6-Hydroxymelatonin sulphate

The DNAme features that had the clearest correlation with a phenotypic trait was hypomethylation at CpG 3 in case participants assessed with Module 1, however, none of the participants in the 6-OHMS study were assessed with this module. Therefore, it was not possible to confirm the hypothesis that particular DNAme features would correlate with sleep problems and/or altered melatonin production often seen in children with ASD.

Although the 6-OHMS study was statistically underpowered, two similarities appeared between the results from the current study and past reported results. Firstly, analysis of actigraphy data indicated the difference in the sleep latency of case and control participants neared a statistical significant difference, with a median value of 5.41 min and 18.41 min for control and case participants respectively (Table 7). This delay in sleep onset in the case group is in line with past reports that individuals with ASD struggle to fall asleep (Polimeni *et al.* 2005; Malow *et al.* 2006; Krakowiak *et al.* 2008; Sheldon *et al.* 2014). Next, analysis of the scores captured from the CSQ indicated a p-value of 0.079 for the difference in the night-wake activity of case and control groups, confirming previous reports that children with ASD wake up more frequently at night time than typically developing children (Polimeni *et al.* 2005; Malow *et al.* 2006; Krakowiak *et al.* 2008; Sheldon *et al.* 2014). Finally, parents with children that have ASD have reported that their children often resist going to bed at specified times (Wiggs & Stores 2004; Couturier *et al.* 2005; Polimeni *et al.* 2005), and sometimes this was due to the child fixating on daytime events that prevented them from going to bed at a stipulated time (Polimeni *et al.* 2005). The majority of parents with children with ASD indicated on the CSQ that their children resist going to bed to some extent, though a more descriptive analysis is required of this behaviour.

Analysis with Levene's tests indicated statistically significant differences in the variance between the WASO and sleep efficiency (with the latter determined by WASO) of case and control participants (Table 7). However, it has been reported that WASO calculated by actigraphy may not be as accurate as WASO calculated by polysomnography in specific individual cases (McCall & McCall 2012). In addition, WASO calculated by actigraphy can either overestimate, or underestimate WASO depending on the magnitude of the total WASO (Marino *et al.* 2013). Given these limitations, actigraphy might not be best suited to calculating WASO in studies with small sample sizes, as the calculated WASO of a single participant may have a disproportionate impact on the overall result.

To improve the investigation of the melatonin production component of the current study, it is recommended to expand the number of participants and to ensure participants assessed with Module 1 are included. However, any larger scale studies that investigate sleep-wake behaviour need to remain cognisant of the idiosyncrasies that accompany children's sleep behaviour. Richdale and Schreck (2009) reported that sample size, IQ and the age of participants are some of the factors which inform the prevalence of sleep problems. For example, the incidence of sleep problems increased up to 80% in children with IQs below 70. In addition, up to 25% of all typically developing children will develop a sleep problem at some stage of childhood (Krakowiak *et al.* 2008; Sheldon *et al.* 2014), and up to 50% of typically developing children under the age of six experience sleep problems (Sheldon *et al.* 2014). Therefore, studies with larger sample sizes could reduce the range of confounding factors by recruiting older children and selecting participants with IQs above 70.

4.5 Conclusion

In small scale investigations, rare sequence variants provide limited insight into a potential relationship between molecular features and phenotypic traits of ASD (Ladd-Acosta *et al.* 2014). The results from DNA sequence analysis in the current study mirrors this observation as no SNPs in the promoter region or genebody of *AANAT* showed any statistically significant difference between case and control participants. In contrast, small scale studies that have investigated the DNAme in ASD and control participants have found particular DNAme features appear more commonly in ASD participants than control participants. (Nguyen *et al.* 2010; Ladd-Acosta *et al.* 2014; Nardone *et al.* 2014). The current study also found that hypomethylation of CpG 3 was more common in case participants assessed with Module 1 of the ADOS-2 than case participants assessed with other modules. As such, DNAme patterns might also differ within the ASD population, and partitioning strategies are therefore as applicable to epigenetic studies as studies that investigate rare DNA sequence variants. This was in line with the hypothesis that in contrast to DNA sequence features, DNAme features would reveal molecular differences between case and control participants, or

match up with distinct ASD traits.

Hypomethylation of CpG 3 in participants assessed with Module 1 of the ADOS-2 further implied that language impairment correlated with DNAme at CpG 3, and was also in line with the hypothesis that analysis of DNAme profiles would revealed molecular features that correlate with distinct phenotypical traits in case participants. Indeed, suppressed melatonin production and suppressed *AANAT* gene expression have been associated with ASD individuals with language impairments (Hu *et al.* 2009; Tordjman *et al.* 2012). In addition, the current study identified that the binding motifs for the transcription factors MEIS1 and ZIC1 overlapped with, or appeared close to CpG 3, which have documented roles in cerebellar development and functioning (Table 4). This relationship with the cerebellum may prove significant as it has been reported that melatonin impacts on the development and functioning of the cerebellum (Rosenstein & Cardinali 1986; Esparza *et al.* 2005; Imbesi *et al.* 2008).

Correlation analysis indicated a linear relationship between the methylation status of several CpG sites relative to each other which is not straightforward to interpretation. This will require separate investigation, but may relate to the process of establishing DNAme profiles early in life (Figure 6). It was not possible to confirm the hypothesis that sleep problems in children with ASD are related to altered melatonin production. The production of melatonin in the South African ASD population needs further exploration with sampling taking place within a season and participants matched for age. It will then be very interesting to investigate any correlations in melatonin production and CpG methylation within the AANAT promoter region.

Finally, the current study provided a glimpse of opportunities and challenges that remain in ASD investigations in the South African context. However, it also demonstrated the potential reward in

gaining a better understanding of the biology that underpins ASD, which will broaden a global understanding of ASD.

4.6 Limitations and recommendations

A limitation of the current study was that too few participants took part in all of the investigations, and therefore the hypothesis that sleep problems in children with ASD are related to altered melatonin production could not be tested directly. It will be interesting to investigate the extent to which trends seen in one investigation correlate to the trends in other investigations. Of the 14 CpG sites available for investigation, only seven CpG sites were located on fragments with unique mass signals. Therefore, only a partial view of DNAme profile was available for analysis. Future studies will benefit from designing a series of primers to investigate shorter lengths of DNA distributed over the entire region of interest.

In addition, investigating DNAme features in an ASD population with secondary symptoms other than restricted language ability, suppressed or elevated melatonin production, or irregular sleep-wake behaviour could reveal additional DNAme trends. It is not clear to what extent the shortcomings of the standard curves used to normalise data contributed to the trends outlined in analysis, and future studies will benefit from adhering to stringent requirements for DNAme assay standards. Finally, establishing a South African or regional data bank to store phenotypic, molecular and other descriptions will expedite cohort building, allow longitudinal studies and eventually building more comprehensive profiles of study participants.

Appendix Tables and Figures

Table 1. Summary of SNP identities and allele frequencies identified in all research participants, and partitioned along ancestry

	Location on Chr 17	SNP Identity	Alleles		All n=52		African n=5		European n=14		Mixed n=33		Case[c] n=15		Control[c] n=18	
			Anc[a]	Var[b]	Anc	Var	Anc	Var	Anc	Var	Anc	Var	Anc	Var	Anc	Var
5'UTR	76 466 736	rs550115167	C	T	0.99	0.01	1.00	●	1.00	●	0.98	0.02	1.00	●	0.97	0.03
5'UTR	76 466 893	rs57197997	C	T	0.97	0.03	1.00	●	0.96	0.04	0.98	0.02	0.97	0.03	0.97	0.03
5'UTR	76 467 027	rs3760138	G	T	0.62	0.38	0.9	0.1	0.4	0.6	0.66	0.34	0.57	0.43	0.74	0.27
5'UTR	76 467 306	rs4238989	C	G	0.59	0.41	0.9	0.1	0.4	0.6	0.62	0.38	0.57	0.43	0.68	0.32
1	76 467 580	rs72466441	A	G	0.98	0.02	1.00	●	0.96	0.04	0.98	0.02	1.00	●	0.97	0.03
Intron	76 468 257	rs73998903	A	C/G	0.99	0.01	1.00	●	1.00	●	0.98	0.02	1.00	●	0.97	0.03
Intron	76 468 653	rs184073292	C	T	0.98	0.02	1.00	●	1.00	●	0.97	0.03	0.97	0.03	0.97	0.03
2	76 468 754	rs61739395[d]	C	T/A	0.95	0.05	1.00	●	1.00	●	0.92	0.08	0.93	0.07	0.92	0.08
Intron	76 469 366	rs4646261	T	A	0.91	0.09	0.9	0.1	1.00	●	0.89	0.11	0.87	0.13	0.89	0.11
Intron	76 469 393	rs116395846	G	A	0.99	0.01	1.00	●	1.00	●	0.98	0.02	0.97	0.03	1.00	●
Intron	76 469 456	rs58495210	A	G	0.98	0.02	1.00	●	1.00	●	0.98	0.02	0.97	0.03	0.97	0.03
Intron	76 469 516	rs371141066	T	C	0.99	0.01	1.00	●	1.00	●	0.98	0.02	0.97	0.03	1.00	●
4	76 469 889	rs58504104	C	A	0.98	0.02	1.00	●	1.00	●	0.98	0.02	0.97	0.03	0.97	0.03
4	76 470 111	rs72466450	G	A	0.99	0.01	1.00	●	1.00	●	0.98	0.02	0.97	0.03	1.00	●

Intron | Non-coding Exon | Coding Exon

[a] Ancestral allele, [b] Variant allele, [c] Participants of mixed ancestry, [d] SNP that result in change protein code, Dots indicate no variant allele

Table 2. Summary of analysis with Fisher's exact test to determine statistically significant differences in the allele frequencies of case and control participants

SNP	Comparison[a]	Fischer's exact p	Odds ratio	Confidence interval
rs3760138	Case & Control n=15 & n=18	0.294	1.833	0.411 to 8.684
rs4238989	Case & Control n=15 & n=18	0.199	2.075	0.488 to 9.405

[a] Participants of mixed ancestry only

Table 3. Recorded massARRAY values for each CpG site that produced a unique mass signal for each participant included in the DNAme analysis

SampleID	CpG 3	CpG 4	CpG 6	CpG 7	CpG 9	CpG12-13	CpG 14
Case Sample 1	58		27	49,5	26,5	57	91
Case Sample 2	66	53,5	34,5	57	32	63	90
Case Sample 3	56,5	41	31,5	46	25,5	57,5	89,5
Case Sample 4	64,5	59,5		60	36,5	67,5	92,5
Case Sample 5	53	32,5	31	41,5	27	55,5	90
Case Sample 6	62,5	48	32,5	51	31,5	62,5	89
Case Sample 7		65	44,5	67,5	44,5	71	90,5
Case Sample 8	70	59	42	62	46	71	90,5
Case Sample 9	64	36,5	29	46	23	57,5	92
Case Sample 10	80,5	65	46,5	69,5	54,5	74,5	91
Case Sample 11	69,5	52	38,5	55,5	33	65	93
Case Sample 12	64,5	47	31	55	30,5	61	89,5
Case Sample 13	69,5	48	30,5	53	27,5	62	91
Case Sample 14	75,5	71,0	39,0	63,5	43,0	71,5	96,5
Case Sample 15	60,5	45	30	43,5	22,5	58,5	94
Case Sample 16	76,0	71,5	40,0		41,5	70,5	96,5
Case Sample 17		73,0	50,0	72,5		73,0	95,0
Case Sample 18	59,5		17	59	25	69	85,5
Case Sample 19	73	51,5	30,5	53	22	63	94,5
Median	64,5	52	32	55,25	31	63	91
Control Sample 1	60,5	46,5	32	53,5	28	59	91,5
Control Sample 2		49	32	53,5	23	59,5	90,5
Control Sample 3	67,5		36,5	61,5	34,5	67	93
Control Sample 4	62,5	48,5	36	54,5	28,5	63	94
Control Sample 5	74,5	61,5	42,5	61		67	93,5
Control Sample 6		57	36	62	36	65,5	92
Control Sample 7	53	37,5	22,5	38,5	23,5	56,5	92,5
Control Sample 8	78	60	37	63	34	66	93,5
Control Sample 9	59,5	57,5	30,5	56	23,5	62	92,5
Control Sample 10	69,5	63		67,5	40,5	73	93
Control Sample 11	66,0		42,0		37,5	63,0	88,5
Control Sample 12	55,0	40,5	14,5	38,5	10,5	49,0	93,0
Control Sample 13		70,5	12,5	62,5	34,0	63,5	96,5
Control Sample 14	68,0	61,5	33,0	58,0	28,0	64,0	96,0
Control Sample 15	72,5	61,5		54,0	21,5	64,0	94,5
Control Sample 16	75,5	63,5			36,0	67,0	96,5
Control Sample 17		69,5	42,0	72,0	42,5	71,5	96,0
Control Sample 18		48,0	39,0		17,5	60,5	95,0
Control Sample 19		60,0		60,5	25,0	62,5	96,5
Control Sample 20	63,5	58,5	29,0		20,5	60,5	98,5
Median	66,75	59,25	34,5	59,25	28	63,25	93,5

Table 4. Summary of Mann Whitney U test results for comparison of DNAme data for six additional partitioning strategies

		CpG 3	CpG 4	CpG 6	CpG 7	CpG 9	CpG 12-13
Module 1 vs Module 2	p-value	0.05	0.93	0.78	0.62	0.26	0.52
	M1[a] n=	8	8	10	10	9	10
	M2[b] n=	4	4	4	4	4	4
Module 1 vs Module 3	p-value	0.11	0.28	0.81	0.93	0.71	0.44
	M1 n=	8	8	10	10	9	10
	M3[c] n=	5	5	4	4	5	5
Control vs Module 1	p-value	0.14	0.34	1	0.51	0.62	0.91
	Con[d] n=	14	18	16	16	19	20
	M1 n=	4	4	4	4	4	4
Control vs Module 2	p-value	0.26	0.55	0.63	0.81	0.39	0.70
	Con n=	14	18	16	16	19	20
	M2 n=	8	8	10	10	9	10
Control vs Module 3	p-value	0.33	0.58	0.89	0.6	0.36	0.31
	Con n=	14	18	16	16	19	20
	M3 n=	5	5	4	4	5	5
Control vs Module 2 & 3	p-value	0.17	1	0.67	0.6	0.24	0.31
	Con n=	14	18	16	16	19	20
	M2.3[e] n=	9	9	8	8	9	9

[a] Module 1

[b] Module 2

[c] Module 3

[d] Control

[e] Module 2 & 3

Table 5. Spearman's rank correlation p-value for comparisons between DNAme of case participants assessed with Module 1, Module 2 and Module 3 in comparison to ADOS-2 SA and RRB scores

Module 1	**CpG 3 n=8**	**CpG 4 n=8**	**CpG 6 n=10**	**CpG 7 n=9**	**CpG 9 n=10**	**CpG 12-13 n=10**
SA	0.51	0.44	0.41	0.60	0.95	0.57
RRB	0.23	0.63	0.58	0.57	0.74	0.65
Module 2	**CpG 3 n=4**	**CpG 4 n=4**	**CpG 6 n=4**	**CpG 7 n=4**	**CpG 9 n=4**	**CpG 12-13 n=4**
SA	0.5	0.79	0.68	0.68	0.68	0.79
RRB	0.37	0.33	0.08	0.08	0.08	0.33
Module 3	**CpG 3 n=5**	**CpG 4 n=5**	**CpG 6 n=4**	**CpG 7 n=4**	**CpG 9 n=5**	**CpG 12-13 n=5**
SA	0.32	0.80	0.68	1	0.80	0.80
RRB	0.78	0.45	0.75	1	0.52	0.78

Table 6. Summary of Pearson's r-values (p_r) that were converted to Z-transformed values (z_p with Fisher's Z-transformation. z_p of case and control groups were then compared with Z-analysis

Comparison	**Case**		**Control**			
	p_r	**z_p**	**p_r**	**z_p**	**Z-statistic[a]**	**p-value[b]**
CpG 4 & CpG 6	0.845	1.238	0.233	0.237	2.874	0.0017
CpG 4 & CpG 12-13	0.951	1.842	0.791	1.074	2.205	0.012
CpG 6 & CpG 9	0.863	1.305	0.487	0.532	2.219	0.012

[a] Output of Z-analysis when comparing case and control z_p

[b] Corresponding p-value to Z-statistic

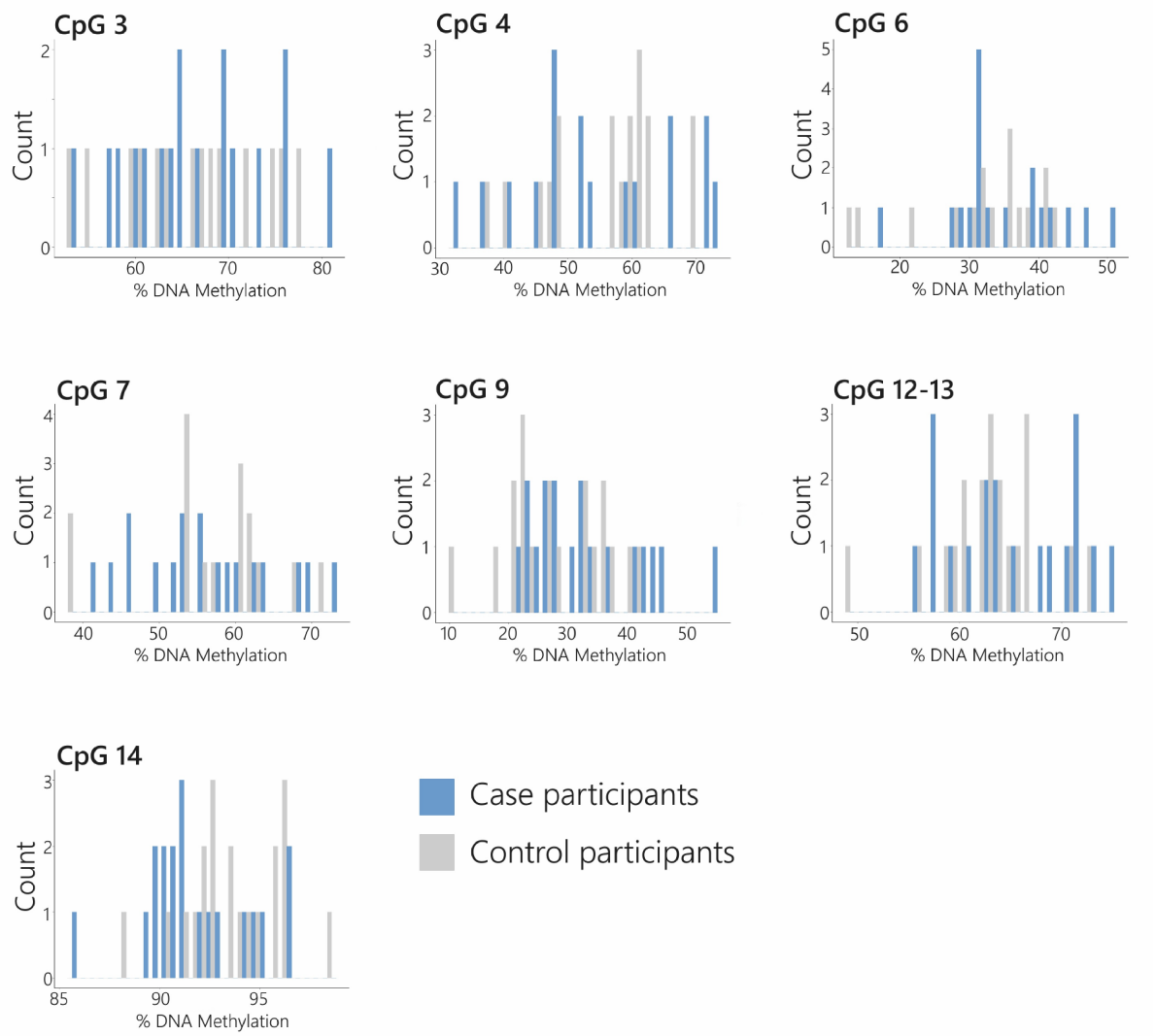


Figure 1. Histogram of distribution of CpG methylation at the eight CpG sites for both case and control groups. Note that no clear distribution pattern appear in any of the histograms indicated.

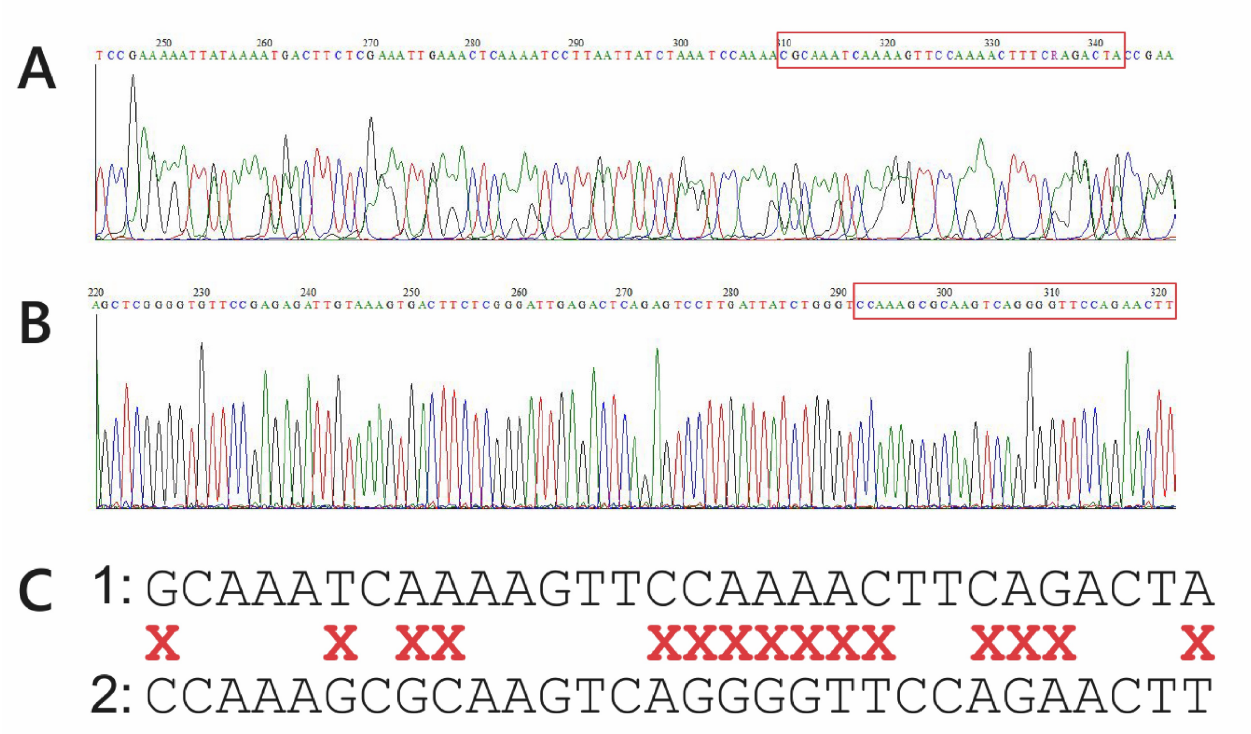


Figure 2. A: Sanger Big Dye chromatogram generated from bisulphite converted DNA. Note the ambiguous quality of chromatogram in A. B: Sanger Big Dye chromatogram for native DNA for the same region as shown in A. Note red frames in A and B highlight the same sequence region. C: Comparison between 1: bisulphite converted and 2: native sequences generated from sequences highlighted in red frames. Red crosses in the middle row indicate unexpected changes in residue identity - only differences in cytosine residues should be apparent.

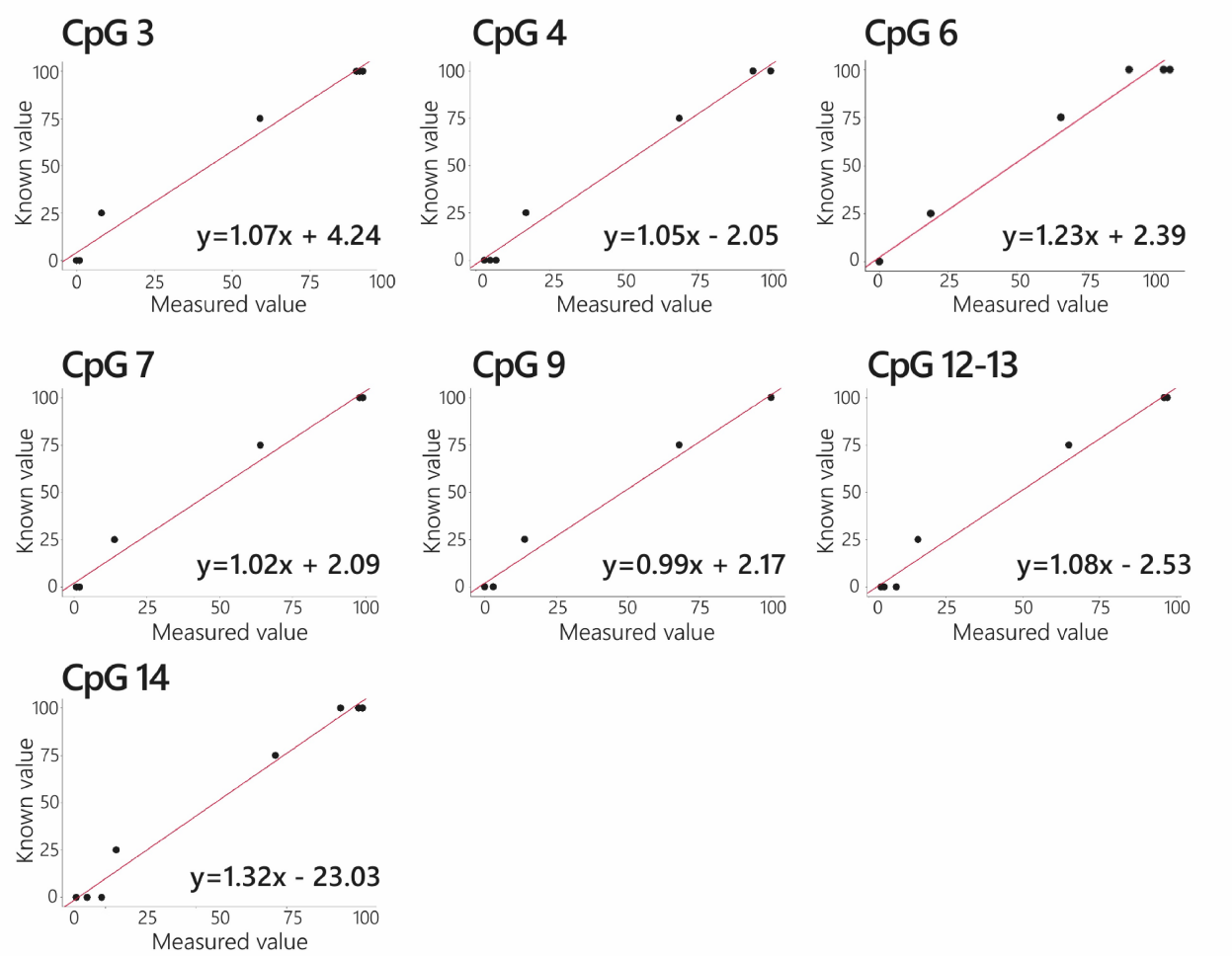


Figure 3. Standard curves for each CpG site used to normalise massARRAY data, with the linear relationship for each site shown on its corresponding graph. Data used to generate curves generated from 0%, 25%, 75% and 100% methylated DNA samples indicated on y-axis and plotted against values recorded from massARRAY system.

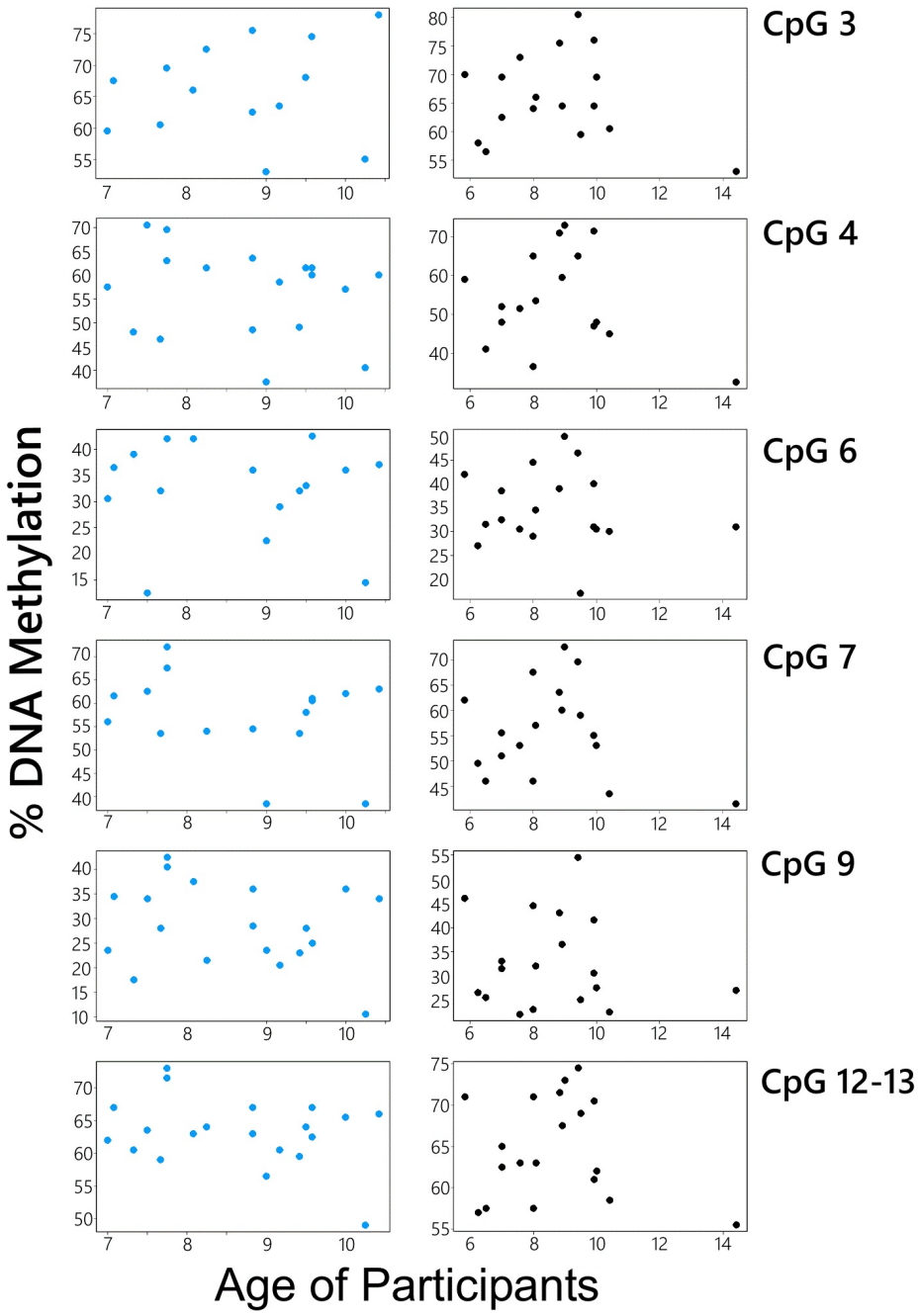


Figure 4. Correlation plots indicating relationship between DNAme and age of research participants for each CpG site. Blue, and black dots indicate control and case group participants respectively. Note that though it was possible to identify a biological outlier in case group (14yr old), removing this outlier would not reveal a linear relationship in the remainder of data points.

Appendix A

CONSENT FORM FOR BIOLOGICAL AND PSYCHOLOGICAL RESEARCH IN ASD PURPOSE

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

PROCEDURES

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

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

RISKS

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

BENEFITS

Although you personally will not receive any direct benefit from this project, individuals (or their family members, and future generations) who might develop, or have the above mentioned diagnosis of ASD may in the future benefit if we can locate genes associated with, or that lead to

such disorders, or locate genes that predict responses to different medications. We do not expect to discover any information of direct clinical relevance to your condition or your treatment during the next few years. In line with the current regulations, we will not make the results of the genetic tests available to subjects or their doctors. However, the results of the play-based scoring sessions, the ADOS-2, may be passed on to your neuro-paediatrician or psychologist on request.

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

CONFIDENTIALITY

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

WITHDRAWAL FROM THE RESEARCH PROJECT

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

CONTACTS

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

PARTICIPATION

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

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

* Contact you for future research Yes /No

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

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

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

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

Appendix B

Consent to Participate in Sleep Study

Thank you for consenting to your child's participation in the sleep study component of our research. His participation will allow us to collect information that will greatly complement the data we obtained from the core part of our molecular study for which you have previously given your consent.

PURPOSE

The sleep study will help us create a profile of your child's typical sleep patterns, and how melatonin levels in the early evening compare to levels in the morning. Melatonin is a hormone that helps regulate sleep patterns, and its levels change across the course of an evening. However, not all people have the same sleep patterns or changes in melatonin levels. Therefore, it is important to measure each person's sleep patterns, and melatonin levels individually. How you child sleep and his melatonin levels are related to the behaviours and genes that we are studying in research.

PROCEDURES

Thank you for previously having completed the Autism Diagnostic Observation Schedule (ADOS-2) assessment and providing the cheek swabs. We will now be measuring your child's activity and rest cycles by actigraphy (the special watch), and collecting urine samples for melatonin measurements (see attached information). We also require you to fill in the brief sleep questionnaire about your child's sleeping behaviour (further below).

PARTICIPATION

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

Wearing an Actiwatch for one week, and collecting four urine samples within the same week my child is wearing the Actiwatch. Yes / No

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

Children's Sleep Habit Questionnaire (Abbreviated)

Preliminary Questions (circle the appropriate option):

1. Does your child take melatonin in any form? Yes / No

2. Does your child sleep with a light on? Yes / No

3. If you answered "Yes" to the above, are you willing to let your child sleep without any light on for the duration of the study? Yes / No

The following statements are about your child's sleep habits and possible difficulties with sleep. Think about the past week in your life when you answer the questions. If last week was unusual for a specific reason, choose the most recent typical week. Unless noted, check "Always" if something occurs every night, "Usually" if it occurs 5-6 times a week, "Sometimes" if it occurs 2-4 times a week, "Rarely" if it occurs once a week, and "Never" if it occurs less than once a week.

Bedtime

Write in your child's usual bedtime: Week nights _____:_____am/pm

Weekends _____ :_____am/pm

	7 Always	5-6 Usually	2-4 Sometimes	1 Rarely	0 Never
1. Child goes to bed at the same time at night.	( )	( )	( )	( )	( )
2. Child falls asleep within 20 minutes after going to bed.	( )	( )	( )	( )	( )
3. Child falls asleep alone in own bed.	( )	( )	( )	( )	( )
4. Child falls asleep in parent's or sibling's bed.	( )	( )	( )	( )	( )
5. Child falls asleep with rocking or rhythmic movements.	( )	( )	( )	( )	( )
6. Child needs special object to fall asleep (doll, special blanket, stuffed animal, etc.).	( )	( )	( )	( )	( )
7. Child needs parent in the room to fall asleep.	( )	( )	( )	( )	( )
8. Child resists going to bed at bedtime.	( )	( )	( )	( )	( )
9. Child is afraid of sleeping in the dark.	( )	( )	( )	( )	( )

Sleep Behaviour

Write in your child's usual amount of sleep each day (combining night time sleep and naps): _____hours and _____minutes

	7 Always	5-6 Usually	2-4 Sometimes	1 Rarely	0 Never
10. Child sleeps about the same amount every day.	( )	( )	( )	( )	( )
11. Child is restless and moves a lot during sleep.	( )	( )	( )	( )	( )
12. Child moves to someone else's bed during the night (parent, sibling, etc.).	( )	( )	( )	( )	( )
13. Child grinds teeth during sleep (your dentist may have told you this).	( )	( )	( )	( )	( )
14. Child snores loudly.	( )	( )	( )	( )	( )
15. Child awakens during the night and is sweating, screaming, inconsolable.	( )	( )	( )	( )	( )
16. Child naps during the day.	( )	( )	( )	( )	( )
Write in the number minutes the nap usually lasts: _____minutes					

Waking During the Night

	7 Always	5-6 Usually	2-4 Sometimes	1 Rarely	0 Never
17. Child wakes up once during the night.	( )	( )	( )	( )	( )
18. Child wakes up more than once during the night.	( )	( )	( )	( )	( )

PLEASE TURN OVER

Morning Wake-Up

Write in the time the child wakes up in the morning: Weekdays_____:_____am/pm

Weekends _____ :_____am/pm

	7 Always	5-6 Usually	2-4 Sometimes	1 Rarely	0 Never
19. Child wakes up by him/herself.	( )	( )	( )	( )	( )
20. Child wakes up very early in the morning (or, earlier than necessary or desired).	( )	( )	( )	( )	( )
21. Child seems tired during the daytime.	( )	( )	( )	( )	( )
22. Child falls asleep while involved in activities.	( )	( )	( )	( )	( )

Appendix C

Using the Actiwatches

Dear Parent/Guardian,

Thank you for consenting to participate in our study. The Actiwatch data will enable us to have a greater understanding of the sleep cycles of our study participants.

The watches are simple to use, and should be minimally intrusive to your lives. Please take note of the following:

1. Please ensure that your child wears the Actiwatch for seven consecutive days.

2. Put the watch on either wrist, and make sure that the watch is fitted securely (as you would a regular wristwatch). Ideally the watch should fit comfortably, but not so loosely that it moves around like a bangle.

3. You will see that the watch has a silver button on its side. When you put your child to bed, and when he rises in the morning, press the button down for three seconds. This acts as a signal in the data for when your child went to sleep and awoke, so this is quite important. It is silent, so don't expect a beep.

Frequently Asked Questions:

What about swimming or bathing? The watch is water resistant for up to 30 minutes at one meter of depth. It is safe to leave the watch on for bathing or showering, but not for swimming. Please do not exceed these limits.

What if the watch irritate my child? If the watch seems to be worrying your child, you can try to put the watch on his other wrist to see if that works better for him. However, if you see your child attempting to remove the watch by whatever means, please remove the watch yourself and return it to us. While we value the contribution you and your child makes to this study, we are ethically bound not to cause your child any undue discomfort.

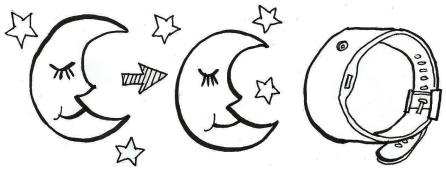
What if the watch gets dirty? If the watch gets dirty, please clean it with a damp cloth, and replace it securely on your child's wrist if you remove it.

What kind of data does the watch collect? The Actiwatch measures the physical activity of your child during the day and night. This will give us an indication of how deeply, or restfully your child sleeps, and if they awaken during the night. The small pink sensor on the watch measures ambient light.

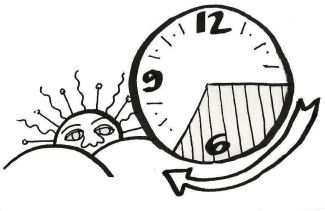
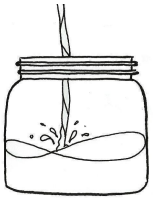
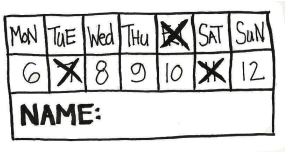
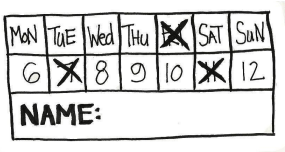
How durable is the watch? The watch is somewhat resilient; however, it is still a sensitive scientific instrument. It will survive average day-to-day activity, but will be damaged if it is dropped, stepped on, or banged against something.

Measuring Melatonin

Our bodies produce melatonin in a cyclical way, with low levels of melatonin present during the day, and high levels late at night. It is possible to measure this fluctuation by performing test on urine samples. To do this, we need you to collect urine from your child on two separate nights,



while your child is wearing the Actiwatch. You have to:

	PM	To Do	AM
1	 Between 6pm in the evening and bedtime…	&	 Between 3am in the morning and including the first wee after waking…
2		….collect all the urine your child passes in the labeled containers provided, secure the lid and..	
3		…mark the day and times the urine was collected. Finally…	
4	4 °C	…store container in the fridge.	4 °C
5	Repeat all of the above for a second night		

References

Acebo, C. *et al.*, 1999. Estimating sleep patterns with activity monitoring in children and adolescents: how many nights are necessary for reliable measures? *Sleep*, 22(1), pp.95–103.

Ackermann, K. *et al.*, 2006. Characterization of human melatonin synthesis using autoptic pineal tissue. *Endocrinology*, 147(7), pp.3235–3242.

Aguinis, H., Gottfredson, R.K. & Joo, H., 2013. Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, 16(2), pp.270-301.

Alarcón, M. *et al.*, 2002. Evidence for a Language Quantitative Trait Locus on Chromosome 7q in Multiplex Autism Families. *American journal of human genetics*, 70(1), pp.60–71.

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.

Allik, H., Larsson, J.-O. & Smedje, H., 2006. Insomnia in school-age children with Asperger syndrome or high-functioning autism. *BMC psychiatry*, 6, p.18.

American Psychiatric Association and others, 2013. Autism Spectrum Disorder. In *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Arlington, Virginia, USA: American Psychiatric Pub, pp. 50–59.

Ancoli-Israel, S. *et al.*, 2003. The role of actigraphy in the study of sleep and circadian rhythms. American Academy of Sleep Medicine Review Paper. *Sleep*, 26(3), pp.342–392.

Antequera, F. & Bird, A., 1993. Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, 90(24), pp.11995–11999.

Arieff, Z. *et al.*, 2010. 5-HTTLPR polymorphism: analysis in South African autistic individuals. *Human biology*, 82(3), pp.291–300.

Bakare, M.O. & Munir, K.M., 2011. Autism spectrum disorders (ASD) in Africa: a perspective. *African journal of psychiatry*, 14(3), pp.208–210.

Baron-Cohen, S., 2004. *Essential difference: Male and female brains and the truth about autism*, Basic Books.

Baymaz, H.I. *et al.*, 2014. MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. *Proteomics*, 14(19), pp.2179–2189.

Bekhet, A.K., Johnson, N.L. & Zauszniewski, J.A., 2012. Resilience in family members of persons with autism spectrum disorder: a review of the literature. *Issues in mental health nursing*, 33(10), pp.650–656.

Berger, S.L. *et al.*, 2009. An operational definition of epigenetics. *Genes & development*, 23(7), pp.781–783.

Bishop, D.V.M., Adams, C.V. & Norbury, C.F., 2006. Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. *Genes, brain, and behavior*, 5(2), pp.158–169.

Blomeke, B. *et al.*, 2008. Arylalkylamine N-acetyltransferase (AANAT) genotype as a personal trait in melatonin synthesis. *Journal of toxicology and environmental health. Part A*, 71(13-14), pp.874–876.

Bustamante, C.D., Burchard, E.G. & De la Vega, F.M., 2011. Genomics for the world. *Nature*, 475(7355), pp.163–165.

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.

Couturier, J.L. *et al.*, 2005. Parental perception of sleep problems in children of normal intelligence with pervasive developmental disorders: prevalence, severity, and pattern. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44(8), pp.815–822.

de Wit, E. *et al.*, 2010. Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Human genetics*, 128(2), pp.145–153.

Duda, M., Kosmicki, J.A. & Wall, D.P., 2014. Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Translational psychiatry*, 4, p.e424.

Ehrich, M. *et al.*, 2005. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44), pp.15785–15790.

Elsabbagh, M. *et al.*, 2012. Global prevalence of autism and other pervasive developmental disorders. *Autism research: official journal of the International Society for Autism Research*, 5(3), pp.160–179.

Esau, L. *et al.*, 2008. The 5-HTTLPR polymorphism in South African healthy populations: a global comparison. Journal of neural transmission, 115(5), pp.755–760.

Esparza, J.L., Gómez, M. & Rosa Nogues, M., 2005. Melatonin reduces oxidative stress and increases gene expression in the cerebral cortex and cerebellum of aluminum exposed rats. *Journal of pineal research*. 39(2), pp.129-136.

Farley, M.A. *et al.*, 2009. Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. *Autism research: official journal of the International Society for Autism Research*, 2(2), pp.109–118.

Felsenfeld, G., 2014. A brief history of epigenetics. *Cold Spring Harbor perspectives in biology*, 6(1), p.a018200.

Fernandez, A.F. *et al.*, 2012. A DNA methylation fingerprint of 1628 human samples. *Genome research*, 22(2), pp.407–419.

Galli-Carminati, G., Deriaz, N. & Bertschy, G., 2009. Melatonin in treatment of chronic sleep disorders in adults with autism: a retrospective study. *Swiss medical weekly*, 139(19-20), pp.293–296.

Giannotti, F. *et al.*, 2011. Sleep in children with autism with and without autistic regression. *Journal of sleep research*, 20(2), pp.338–347.

Gibney, E.R. & Nolan, C.M., 2010. Epigenetics and gene expression. *Heredity*, 105(1), pp.4–13.

Goldberg, A.D., Allis, C.D. & Bernstein, E., 2007. Epigenetics: a landscape takes shape. *Cell*, 128(4), pp.635–638.

Goodlin-Jones, B.L. *et al.*, 2008. Sleep patterns in preschool-age children with autism, developmental delay, and typical development. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(8), pp.930–938.

Gordon, L. *et al.*, 2012. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome research*, 22(8), pp.1395–1406.

Gotham, K. *et al.*, 2007. The Autism Diagnostic Observation Schedule: revised algorithms for improved diagnostic validity. *Journal of autism and developmental disorders*, 37(4), pp.613–627.

Gotham, K., Pickles, A. & Lord, C., 2009. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of autism and developmental disorders*, 39(5), pp.693–705.

Grafodatskaya, D. *et al.*, 2010. Autism spectrum disorders and epigenetics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49(8), pp.794–809.

Griefahn, B. *et al.*, 2003. Melatonin production during childhood and adolescence: a longitudinal study on the excretion of urinary 6-hydroxymelatonin sulfate. *Journal of pineal research*, 34(1), pp.26–31.

Gu, F., Chauhan, V. & Chauhan, A., 2013. Impaired synthesis and antioxidant defense of glutathione

in the cerebellum of autistic subjects: Alterations in the activities and protein expression of glutathione-related enzymes. *Free Radical Biology and Medicine*, 65, pp.488–496.

Hansen, T.V.O. *et al.*, 2007. Collection of blood, saliva, and buccal cell samples in a pilot study on the Danish nurse cohort: comparison of the response rate and quality of genomic DNA. *Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 16(10), pp.2072–2076.

Happé, F., 1999. Understanding assets and deficits in autism-Why success is more interesting than failure. *The Psychologist*, 12(11), pp.540–546.

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

Heim, C. & Binder, E.B., 2012. Current research trends in early life stress and depression: Review of human studies on sensitive periods, gene–environment interactions, and epigenetics. *Experimental neurology*, 233(1), pp.102–111.

Hendrich, B. & Bird, A., 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Molecular and cellular biology*, 18(11), pp.6538–6547.

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

Hering, E. *et al.*, 1999. Sleep patterns in autistic children. *Journal of autism and developmental disorders*, 29(2), pp.143–147.

Hernandez, D.G. *et al.*, 2011. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Human molecular genetics*, 20(6), pp.1164–1172.

Hodge, S.M. *et al.*, 2010. Cerebellum, language, and cognition in autism and specific language impairment. *Journal of autism and developmental disorders*, 40(3), pp.300–316.

Hollander, E. *et al.*, 2000. The relationship between repetitive behaviors and growth hormone response to sumatriptan challenge in adult autistic disorder. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, 22(2), pp.163–167.

Hranilovic, D. *et al.*, 2016. DNA Methylation Analysis of HTR2A Regulatory Region in Leukocytes of Autistic Subjects. *Autism research: official journal of the International Society for Autism Research*, 9(2), pp.204–209.

Hsieh, J. & Eisch, A.J., 2010. Epigenetics, hippocampal neurogenesis, and neuropsychiatric disorders: unraveling the genome to understand the mind. *Neurobiology of disease*, 39(1), pp.73–84.

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., 2013. The expanding genomic landscape of autism: discovering the "forest" beyond the "trees." *Future neurology*, 8(1), pp.29–42.

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.

Imbesi, M., Uz, T. & Manev, H., 2008. Role of melatonin receptors in the effects of melatonin on BDNF and neuroprotection in mouse cerebellar neurons. *Journal of neural transmission*, 115(11), pp.1495–1499.

Iwamoto, K. *et al.*, 2011. Neurons show distinctive DNA methylation profile and higher interindividual variations compared with non-neurons. *Genome research*, 21(5), pp.688–696.

Jensen, P., 2014. Behaviour epigenetics–the connection between environment, stress and welfare. *Applied Animal Behaviour Science*, 157, pp.1–7.

Jenwitheesuk, A. *et al.*, 2014. Melatonin regulates aging and neurodegeneration through energy metabolism, epigenetics, autophagy and circadian rhythm pathways. *International journal of molecular sciences*, 15(9), pp.16848–16884.

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

Kassubek, J., Hickok, G. & Erhard, P., 2004. Involvement of classical anterior and posterior language areas in sign language production, as investigated by 4 T functional magnetic resonance imaging. *Neuroscience letters*, 364(3), pp.168–172.

Keverne, E.B., 2014. Significance of epigenetics for understanding brain development, brain evolution and behaviour. *Neuroscience*, 264, pp.207–217.

Koike, B.D.V. *et al.*, 2013. Screening for polymorphisms in the AANAT gene and their association with extreme diurnal preference. *www. sleepscience. com. br*, p.141.

Korkmaz, A. *et al.*, 2009. Role of melatonin in metabolic regulation. *Reviews in endocrine & metabolic disorders*, 10(4), pp.261–270.

Kotagal, S. & Broomall, E., 2012. Sleep in children with autism spectrum disorder. *Pediatric neurology*, 47(4), pp.242–251.

Krakowiak, P. *et al.*, 2008. Sleep problems in children with autism spectrum disorders, developmental delays, and typical development: a population-based study. *Journal of sleep research*, 17(2), pp.197–206.

Ladd-Acosta, C. *et al.*, 2014. Common DNA methylation alterations in multiple brain regions in autism. *Molecular psychiatry*, 19(8), pp.862–871.

Laird, P.W., 2010. Principles and challenges of genomewide DNA methylation analysis. *Nature reviews. Genetics*, 11(3), pp.191–203.

LaSalle, J.M., Powell, W.T. & Yasui, D.H., 2013. Epigenetic layers and players underlying neurodevelopment. *Trends in neurosciences*, 36(8), pp.460–470.

Lauvin, M.-A. *et al.*, 2012. Functional morphological imaging of autism spectrum disorders: current position and theories proposed. *Diagnostic and interventional imaging*, 93(3), pp.139–147.

Leu, R.M. *et al.*, 2011. Relation of melatonin to sleep architecture in children with autism. *Journal of autism and developmental disorders*, 41(4), pp.427–433.

Lidzba, K. *et al.*, 2008. Reorganization of the cerebro-cerebellar network of language production in patients with congenital left-hemispheric brain lesions. *Brain and language*, 106(3), pp.204–210.

Limoges, E. *et al.*, 2005. Atypical sleep architecture and the autism phenotype. *Brain: a journal of neurology*, 128(Pt 5), pp.1049–1061.

Lim, T.-S. & Loh, W.-Y., 1996. A comparison of tests of equality of variances. *Computational statistics & data analysis*, 22(3), pp.287–301.

Lin, Q. *et al.*, 2012. Promoter methylation analysis of seven clock genes in Parkinson's disease. *Neuroscience letters*, 507(2), pp.147–150.

Lipsky, R.H., 2013. Epigenetic mechanisms regulating learning and long-term memory. *International journal of developmental neuroscience: the official journal of the International Society for Developmental Neuroscience*, 31(6), pp.353–358.

Littner, M. *et al.*, 2003. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep*, 26(3), pp.337–341.

Lokk, K. *et al.*, 2014. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome biology*, 15(4), p.r54.

Lord, C. *et al.*, 2000. The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *Journal of autism and developmental disorders*, 30(3), pp.205–223.

Lowe, R. *et al.*, 2013. Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics: official journal of the DNA Methylation Society*, 8(4), pp.445–454.

Lubin, F.D., Roth, T.L. & David Sweatt, J., 2005. Epigenetic Regulation of bdnf Gene Transcription in the Consolidation of Fear Memory. *The Journal of Neuroscience*. 28(42), pp.10576-10586.

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

Maegawa, S. *et al.*, 2010. Widespread and tissue specific age-related DNA methylation changes in mice. *Genome research*, 20(3), pp.332–340.

Malow, B.A. *et al.*, 2006. Characterizing sleep in children with autism spectrum disorders: a multidimensional approach. *Sleep*, 29(12), pp.1563–1571.

Marino, M. *et al.*, 2013. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, 36(11), pp.1747–1755.

McCall, C. & McCall, W.V., 2012. Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. *Journal of sleep research*, 21(1), pp.122–127.

Mehler, M.F., 2008. Epigenetic principles and mechanisms underlying nervous system functions in health and disease. *Progress in neurobiology*, 86(4), pp.305–341.

Melke, J. *et al.*, 2008. Abnormal melatonin synthesis in autism spectrum disorders. *Molecular psychiatry*, 13(1), pp.90–98.

Mohan, K.N. & Chaillet, J.R., 2012. Chapter 8 - Emerging Role of Epigenetics in Human Neurodevelopmental Disorders A2 - Tollefsbol, Trygve O. In *Epigenetics in Human Disease*. San Diego: Academic Press, pp. 153–173.

Nagae, G. *et al.*, 2011. Tissue-specific demethylation in CpG-poor promoters during cellular differentiation. *Human molecular genetics*, 20(14), pp.2710–2721.

Namihira, M. & Nakashima, K., 2013. Mechanisms of astrocytogenesis in the mammalian brain. *Current opinion in neurobiology*, 23(6), pp.921–927.

Nardone, S. *et al.*, 2014. DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Translational psychiatry*, 4, p.e433.

Newbury, D.F. & Monaco, A.P., 2010. Genetic advances in the study of speech and language disorders. *Neuron*, 68(2), pp.309–320.

Newschaffer, C.J., Croen, L.A. & Daniels, J., 2007. The epidemiology of autism spectrum disorders*. *Annual review of public health*. 28, pp.235-258.

Nguyen, A. *et al.*, 2010. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 24(8), pp.3036–3051.

Nowak, R. *et al.*, 1987. The correlation between serum and salivary melatonin concentrations and urinary 6-hydroxymelatonin sulphate excretion rates: two non-invasive techniques for monitoring human circadian rhythmicity. *Clinical endocrinology*, 27(4), pp.445–452.

Oliver, B.R. & Plomin, R., 2007. Twins' Early Development Study (TEDS): A Multivariate, Longitudinal Genetic Investigation of Language, Cognition and Behavior Problems from Childhood Through Adolescence. *Twin research and human genetics: the official journal of the International Society for Twin Studies*, 10(01), pp.96–105.

Owens, J.A., Spirito, A. & McGuinn, M., 2000. The Children's Sleep Habits Questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. *Sleep*, 23(8), pp.1043–1051.

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

Pallant, J. and Manual, S.S., 2001. A step by step guide to data analysis using SPSS.

Pandi-Perumal, S.R. *et al.*, 2006. Melatonin. *The FEBS journal*, 273(13), pp.2813–2838.

Polimeni, M.A., Richdale, A.L. & Francis, A.J.P., 2005. A survey of sleep problems in autism, Asperger's disorder and typically developing children. *Journal of intellectual disability research: JIDR*, 49(Pt 4), pp.260–268.

Popejoy, A.B. & Fullerton, S.M., 2016. Genomics is failing on diversity. *Nature*, 538(7624), p.161.

Razali, N.M. & Wah, Y.B., 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), pp.21–33.

Richdale, A.L. & Schreck, K.A., 2009. Sleep problems in autism spectrum disorders: prevalence, nature, & possible biopsychosocial aetiologies. *Sleep medicine reviews*, 13(6), pp.403–411.

Rosenstein, R.E. & Cardinali, D.P., 1986. Melatonin increases in vivo GABA accumulation in rat hypothalamus, cerebellum, cerebral cortex and pineal gland. *Brain research*, 398(2), pp.403–406.

Rossignol, D.A. & Frye, R.E., 2014. Melatonin in autism spectrum disorders. *Current clinical*

pharmacology, 9(4), pp.326–334.

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.783–792.

Ruggeri, B. *et al.*, 2014. Biomarkers in autism spectrum disorder: the old and the new. *Psychopharmacology*, 231(6), pp.1201–1216.

Saito, A. *et al.*, 2012. Unfolded protein response, activated by OASIS family transcription factors, promotes astrocyte differentiation. *Nature communications*, 3, p.967.

Sarachana, T. & Hu, V.W., 2013. Differential recruitment of coregulators to the RORA promoter adds another layer of complexity to gene (dys) regulation by sex hormones in autism. *Molecular autism*, 4(1), p.39.

Saxonov, S., Berg, P. & Brutlag, D.L., 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 103(5), pp.1412–1417.

Schaafsma, S.M. & Pfaff, D.W., 2014. Etiologies underlying sex differences in Autism Spectrum Disorders. *Frontiers in neuroendocrinology*, 35(3), pp.255–271.

Sekine, A. *et al.*, 2001. Identification of single-nucleotide polymorphisms (SNPs) of human N-acetyltransferase genes NAT1, NAT2, AANAT, ARD1, and L1CAM in the Japanese population. *Journal of human genetics*, 46(6), pp.314–319.

Serón-Ferré, M. *et al.*, 2001. The development of circadian rhythms in the fetus and neonate. *Seminars in perinatology*, 25(6), pp.363–370.

Sheldon, S.H. *et al.*, 2014. *Principles and Practice of Pediatric Sleep Medicine*, Elsevier Health Sciences.

Singh, K. & Zimmerman, A.W., 2015. Sleep in Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder. *Seminars in pediatric neurology*, 22(2), pp.113–125.

Soria, V. & Martinez-, E., 2010. Resequencing and association analysis of arylalkylamine N-acetyltransferase (AANAT) gene and its contribution to major depression susceptibility. *Journal of pineal research*, 49, pp.35–44.

Stoodley, C.J. & Schmahmann, J.D., 2009. The cerebellum and language: evidence from patients with cerebellar degeneration. *Brain and language*, 110(3), pp.149–153.

Suchiman, H.E.D. *et al.*, 2015. Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. *Frontiers in genetics*, 6,

p.287.

Teschendorff, A.E. *et al.*, 2010. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome research*, 20(4), pp.440–446.

Toma, C. *et al.*, 2007. Is ASMT a susceptibility gene for autism spectrum disorders? A replication study in European populations. *Molecular psychiatry*, 12(11), pp.977–979.

Tordjman, S. *et al.*, 2012. Day and nighttime excretion of 6-sulphatoxymelatonin in adolescents and young adults with autistic disorder. *Psychoneuroendocrinology*, 37(12), pp.1990–1997.

Tordjman, S. *et al.*, 2005. Nocturnal excretion of 6-sulphatoxymelatonin in children and adolescents with autistic disorder. *Biological psychiatry*, 57(2), pp.134–138.

Valor, L.M. & Guiretti, D., 2014. What's wrong with epigenetics in Huntington's disease? *Neuropharmacology*, 80, pp.103–114.

Warnich, L. *et al.*, 2011. Pharmacogenomic Research in South Africa: Lessons Learned and Future Opportunities in the Rainbow Nation. *Current pharmacogenomics and personalized medicine*, 9(3), pp.191–207.

Wiggins, L.D. *et al.*, 2015. Using standardized diagnostic instruments to classify children with autism in the study to explore early development. *Journal of autism and developmental disorders*, 45(5), pp.1271–1280.

Wiggs, L. & Stores, G., 2004. Sleep patterns and sleep disorders in children with autistic spectrum disorders: insights using parent report and actigraphy. *Developmental medicine and child neurology*, 46(6), pp.372–380.

Yang, Z. *et al.*, 2016. Circadian-relevant genes are highly polymorphic in autism spectrum disorder patients. *Brain & development*, 38(1), pp.91–99.

Yazici, B. & Yolacan, S., 2007. A comparison of various tests of normality. *Journal of statistical computation and simulation*, 77(2), pp.175–183.

Zeitzer, J.M. *et al.*, 2000. Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. *The Journal of physiology*, 526 Pt 3, pp.695–702.

Zimmerman, D.W., 2004. Inflaction of type I error rates by unequal variances associated with parametric, nonparametric, and rank-transformation tests. *Revista peruana de medicina experimental y salud publica*, 25(1), pp.103–133.