Proteomic and SNP analysis of the Cadherin 10 type-II (*CDH10)* gene, in the South African Autistic population

A thesis submitted in fulfillment of the requirements for the degree of Magister Scientiae in the Department of Biotechnology, University of the Western Cape

Firzana October

Supervisor: Dr. Z. Arieff

Co-supervisor: Prof. B.K. Ndimba

UWC 2013

Key words

Autism, Association study, Single Nucleotide Polymorphism's (SNPs), Taqman® Technology, Cadherin 10 type II (*CDH10)*, Urine Proteomics, Biomarkers, 1D SDS-PAGE, 2D SDS-PAGE, Mass Spectrometry

Declaration

I declare that “Proteomic and SNP analysis of the Cadherin 10 type-II (*CDH10)* gene, in the South African Autistic population” is my own work that has not been submitted for any degree or examination in any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Firzana October December 2013

Signed……

Table of Contents

Acknowledgments

- I would like to make thanks to my Creator, the Almighty (S.W.T) who has fulfilled my every prayer. Completing this project has required strength, patience and hope, and I thank thee for granting me all the patience.
- I thank my supervisor/s (Dr. Z Arieff and Prof B. Ndimba) and their postdoctoral fellows (Dr. J Sharma and Dr. R Ngara) for their support, the opportunity and their unflinching patience with me throughout the years.
- My family, my parents I give shukr for raising me and teaching me to open my mind to opportunities as rewarding as this M.Sc. My siblings, of which I am very fortunate to have a whole soccer team, (M. Reza, Amie Shakeel, Tietie Alijah, Aziza, Gamieda , Abdus-Samad, Shaabirah, Waafiqah, Gaironesa, Najmunisa, Aqeelah, Mahir, Zafirah, Mustapha, Ismaa-eel, Humaam Sulaiman and Zainonesa).
- The principal, teachers, pupils that gave samples for this project (Nolunthando and Lantana primary school).
- To the Human Genetics Research Group lab members, thank you for enduring me through all my phases and for the entertainment you provided.
- The Biotechnology Department staff for their administrative assistance.
- The National Research Foundation (NRF) for my bursary and for funding this project.
- My friends, the coffee breaks, the very late nights on campus, the weekend work and at times the overnight attempts at completing this degree.
- To my beloved husband Taariq, thanks for your understanding and patience.

List of abbreviations

1D	One dimensional
2D	Two dimensional
APS	Ammonium persulphate
BSA	Bovine serum albumin
CHAPS	3-[(3-Cholamidopropyl) dimethylammonio]- 1-propanesulfonate
cm	Centimetre
dL	Decilitre
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
et al	et alia
HWE	Hardy-Weinberg equilibrium
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
kDa	Kilo Daltons
MW	Molecular Weight
M	Molar
m/z	Mass per charge
ml	Millilitre
mm	Millimetre
mM	Millimolar
MOWSE	Molecular weight search
NCBI	National Centre for Biotechnology Information
PCR	Polymerase chain reaction

pI	Isoelectric point
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel Electrophoresis
SNP	Single nucleotide polymorphism
TEMED	N, N, N1,N1- Tetra methylethelene-diamine
µg	Microgram
µg/µl	Microgram per microliter
µl	Microliter
V	Volts
v/v	Volume per volume

List of tables

Chapter 1

Chapter 2

Chapter 3

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 1

Abstract

Autism or autism spectrum disorder (ASD) is a very diverse neurological disorder that manifests specifically in children and infants between the ages of two to three years of age. An individual suffering is deemed as autistic and individuals suffering would be classed under the banner of ASD. It is observed that sufferers have impairment in their social and interactive skills. It has both genetic and environmental factors that contribute to its diversity and although the primary cause of autism is still unclear, scientist are investigating both factors. In this study we aimed to investigate the molecular genetics of autism in the South African (SA) population. This was done in two parts, a genetic association study and a functional genomics (proteomic study). An association study of the 2 single nucleotide polymorphisms (SNPs) of the Cadherin 10 type II gene (*CDH10*) (rs4307059 and rs4327572) was investigated in the SA healthy and autistic population. The proteomic approach was used to determine the differential expression of genes of the healthy population and compared to the autistic population of African descent. In both parts of the project, objectives were achieved. The SNPs were successfully genotyped however no association was determined for autism in the SA population. The urine protein profiles with 1 dimensional (1D) and 2 dimensional (2D) Sodium Dodecyl Sulfate-Poly Acrylamide Gel Electrophoresis (SDS-PAGE) generated in this study has revealed the following proteins, Uromodulin, Vitelline membrane outer layer protein homologue, kinninogen-1, Alpha-1-Antitrypsin, Ig Kappa chain region C, and CD59 glycoprotein that require further investigation. The results indicated that six of the identified proteins were expressed in both groups but were found to be either quantitatively or statistically significant. However, a statistically significant difference was observed in the expression of one protein (Uromodulin) which was observed

to be expressed in the healthy group but absent in the experimental group. However further investigation is required validation of these findings.

Chapter 1- Literature review

Table of Contents

1. Autism

1.1. Background to autism

"Autism is the most severe of the pervasive developmental disorders (PDD) with primary symptoms of impaired social skills, delayed or non-existent language skills, and the presence of stereotypic, repetitive movements" (Association, 1994). It is characterized based on the deficiencies in social interactions, behavior patterns and language abilities of the individual (Sebat *et al.*, 2007).

Wilson (2011) described it as too much chaos internally in the brain and nervous system and this results in the child withdrawing him/herself to reduce stimulation and chaos from outside. Autism Spectrum Disorder (ASD) and Pervasive Developmental disorder (PDD) are such diverse disorders in nature and has different degrees in which they present themselves (Muhle *et al.*, 2004). Individuals have one: erratic behavior, two: tend to be very sensitive and incapable of holding a conversation, and then at the other end of the spectrum the savant individuals. Savants are individuals with mental disabilities but are skilled as well typically limited to areas such as music, art, calendar calculating, mechanical skills and numerical ability (Treffert, 1999).

In the United States of America a group of psychiatric and psychologists developed a set of criteria for assessing individuals with mental disorders named the diagnostic and statistical manual of disorders, manual four (DSM-IV). It is the most widely used diagnostic tool for ASD and for an individual to be considered autistic, he/she should display features described in this manual. Apart from these common features some individuals may also suffer from different types of disorders in communal behavior, such as sensory dysfunctions, sleep and

gastrointestinal disorders, seizures and mental retardation (Bauman, 2010). The manual has been through several revisions over the years with advances made by the scientific community. Charney *et al.*,(2002) criticized the DSM-IV manual because it did not effectively "reflect a neurobiological basis" for most mental disorders (Charney *et al.*, 2002). In the DSM-IV manual autism disorder, Asperger's, Retts Syndrome, Childhood Disintegrative disorder and Pervasive Developmental disorder Not otherwise specified (PDD-NOS) are all categorized and diagnosed under pervasive Developmental Disorder. However in the DSM-V manual, Asperger's and PDD-NOS have been removed. PDD-NOS are classified as a Social Communication Disorder whereas Asperger's has become a subtype under ASD. The revised manual DSM-V includes a grading system of disorders based on the severity of the individual according to the assistance or support required (Grant and Nozyce, 2013).

1.2.Co-morbid disorders

This complex disorder has been associated with various other disorders due its diverse nature and symptoms. It has been recorded back in time as early as 1911 by a psychiatrist, Eugen Bleuler although the term applied to adult schizophrenia. Schizophrenia along with other disorders such as Rett's Syndrome (RS), Fragile X syndrome (FraX), Attention Deficit Hyperactivity Disorder (ADHD), Asperger's, Severe Mental retardations and Pervasive Developmental Disorders-Not Otherwise Specified (PDD-NOS) all share similar symptoms. Most of these disorders fall under the "umbrella" of ASD also known as pervasive developmental disorders (figure 1.1)

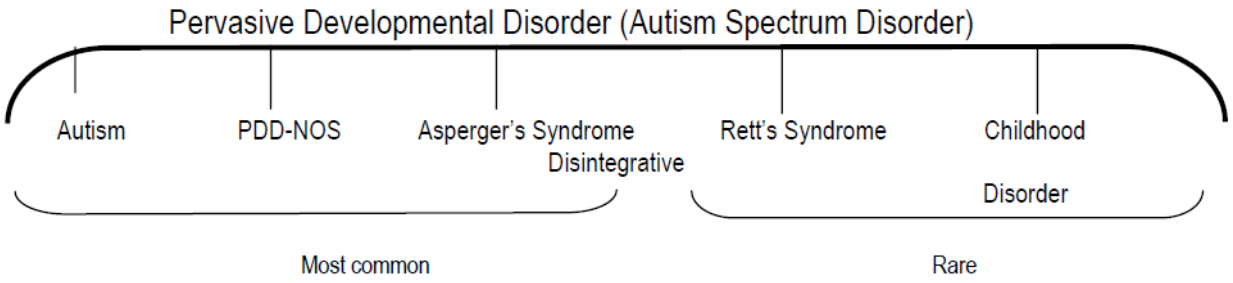


Figure 1.1: ASD umbrella (Connecticut Autism Spectrum Resource Centre, 2012**)**

Rett's syndrome (RS): Rett's syndrome like Autism is a "severe neuro-developmental disorder that contributes significantly to severe intellectual disability" (Weaving *et al.*, 2005). This syndrome was first described in 1983 and unlike autism, only occurred in females. This disorder showed unusually similar in terms of the clinical presentation of loss of motor skills, ataxia, and acquired microcephaly (Weaving *et al.*, 2005). Typical features of RS are early onset at the ages of 6-18 months of age followed by normal development. It is caused by a mutation in the gene, *MECP2* which encodes the methyl-CgP-binding protein and it is an X-linked dominant disorder that affects the mechanisms of synaptic development (Armstrong *et al.*, 1995, Percy, 2001).

Fragile X Syndrome (*FraX*): Even though this disorder does not fall under the spectrum as indicated in figure 1.1, it still shares similarities with autism. *FraX* is presently regarded the most related cause of inherited mental retardation (Kaufmann *et al.*, 2004), and is the expansion of a CGG mutation within the 5' un-translated region of the FMR1 gene, depending on how many repeats are present. The affected alleles are classed as 1: normal of 5-40 repeats; 2: intermediate/gray zone of 45-54 repeats; 3: permutation of 55-200 repeats and 4: full mutation greater than 200 repeats (Maddalena *et al.*, 2001). Individuals that have a

full mutation have cognitive impairment and behavioral abnormalities (Kaufmann *et al.*, 2004).

Schizophrenia: Almost exactly a century ago patients with autism was diagnosed as schizophrenic (Bleuler, 1911). It is suggested to be the most disturbing form of psychiatric illnesses (Luhrmann, 2007). It presents consistently with a combination of different groups of symptoms starting with psychosis, progressing to emotional withdrawal ending with cognitive dysfunction. In recent research a clear distinction has been made between these disorders where autism appears during early childhood development and schizophrenia would present during early adulthood. Furthermore typical symptoms in schizophrenia would include hallucinations and delusions, which are rarely observed in autism (The University of Texas at Dallas News Center, 2012). Other differences separating autism from schizophrenia were familial backgrounds and differential treatment responses (Meyer *et al.*, 2011). Regardless of these differences, overlapping of symptoms are strong between the two disorders on a behavioral level which includes the impairment of social interaction and cognition. Other symptoms observed between both autism and schizophrenia include the "disruption of emotional processing and sensorimotor gating and impairments of executive functions" (Rapin and Tuchman, 2008, Cheung *et al.*, 2010, Meyer *et al.*, 2011).

Asperger's: Asperger's syndrome (AS) were described by a pediatrician Mr. H. Asperger in 1944, who studied children exerting problematic social adaptation and communication skills (Todorov and Arnaoudova, 2012). Autism and Asperger's has been considered as highly related disorders as most symptoms are clinically indifferent making diagnosis between them difficult (Sanders, 2009, Todorov and Arnaoudova, 2012). Both disorders appear with

"distinct verbal styles, motor signs, emotion perception and pragmatic reasoning" (Yu *et al.*, 2011). Distinguishing between the two however can be observed by the lack of clinically significant delay in cognitive development and language and the appearance observed only after the age of 3 years in autism and not in AS (Todorov and Arnaoudova, 2012).

1.3. Prevalence of autism

In the 1900's in the USA it was proposed that metal preservatives used in medicines and vaccines caused autism. In the 1960's and 1970's the incidence of this disorder were reported as no more than 5 in every 10 000 (Gillberg *et al.*, 1991, Newschaffer *et al.*, 2007) and in the 1980's of ~10 in 10 000. In the 1990's it was reported to range from 5 in 10 000 to 72 in 10 000 (Sponheim and Skjeldal, 1998, Kadesjö *et al.*, 1999, Newschaffer *et al.*, 2007). The variation of the prevalence of autism became indicative of the increased awareness or information within communities. Up until 2007, surveys still showed increases of the incidence and variations amongst population. The incidence increased to 110 in 10 000 in a British report in a population of 55 000, giving rise to urgency for investigating autism, (Baird *et al.*, 2006, Newschaffer *et al.*, 2007). Based on studies conducted in different sites of the USA, the current prevalence of ASD is estimated to be one in 150 to one in 91 individuals with a male-to-female ratio of 4:1 (Kogan *et al.*, 2009, Curtin *et al.*, 2010). These statistics are consistent with the prevalent rates reported in other countries such as Europe and Asia, however some of the results for prevalence estimates are higher (Kogan *et al.*, 2009, Boyle *et al.*, 2011, Kim *et al.*, 2011). Based on the evidence reviewed by Elsabbagh *et al.*, (2012) median of worldwide prevalence estimates of ASDs is 62/10 000 (Elsabbagh *et al.*, 2012). In China a prevalence rate of autism of 2.38/10,000 was observed, but this was based on a data set derived from only disabled children affected with the disorder (Li *et al.*, 2011). Many children in SA remain undiagnosed or untreated due to lack of awareness among SA

communities. It has also been revealed that the knowledge for ASD awareness are lacking and that in the last decade only twelve scientific papers have been publish on ASD in Africa (Bakare and Munir, 2011). A total of only two papers shed light on the epidemiology and prevalence of autism in African countries. However these papers only shed light on the prevalence in Egyptian (33.6%) and in Tunisian (11.5%) population for autism (Seif Eldin *et al.*, 2008, Bakare and Munir, 2011).

There are no diagnostic or molecular tests for autism and cases have only been characterized on basis of behavior and social development amongst individuals by a qualified psychologist. It is reported that ASD is more prevalent in males with ratios of 4: 1 (male: female) for classic autism whereas Asperger syndrome observes higher ratios with 11:1 males to female (Baron-Cohen *et al.*, 2011). The reasons for this observation continue to be blurry although it has been reported that the male bias for autism is thought to be caused by a hormonal perturbation during pregnancy with regards to genetic control and may be linked to development in the brain and immune systems. These events may thus overlap between autism and autoimmunity in determining male sex bias and may thus provide a link or evidence of etiology among autism (Becker, 2012).

1.4.Causes of autism

Autism is a neurological disorder which implicates the central nervous system (CNS). The CNS is composed of the brain, a very complex organ that regulates the motor- and neurological aspects of an individual and the spinal cord. The spinal cord harbors the stem of the neuro-systems and transmits messages from the brain to the extending body and as well

as receiving messages from extended body appendages. Since its first case report, the precise cause for this disorder is still elusive; etiology is multifactorial, which suggests a strong genetic foundation as well as an environmental component. Studies of genetic and environmental factors have provided some clues to elucidate the intricate pathogenesis of this disorder (Benvenuto *et al.*, 2013).

1.4.1. Environmental Factors

Environmental factors are defined as those factors external to the individual, chemicals, drugs, light and temperature, food sources both biotic and abiotic. Foll and Giaggotti (2006) showed that a simulation could be used to identify the effects of environmental factors both biotic and abiotic on genetic variability and structuring and the likelihood of these factors having a genetic impact on generations or populations under various conditions. The most common environmental factors implicated in autism include chemical and drug influences.

Chemical influences- Mercury poisoning: From the period of 1890 to 1950, a number of children suffered from a disease named Acrodynia (also known as pink disease) which had a fatality of 20% , (Initiative, 2005). Acrodynia is observed by the symptoms of alterations in temperament and skin, neurologic indicators, tachycardia, and stomatitis (Bjørklund, 1995). Warkany and Hubbard (1953) demonstrated that mercury was a causative agent because 25 out of 28 sufferers had mercury poisoning. It was illustrated that the teething powders used on infants had mercuric chloride as the active agent (Bjørklund, 1995). It was also observed that among the individuals who were exposed to the teething powders showed symptoms similar to that of autism. To date research has shown that mercury is highly neurotoxic, especially during brain development (Costa *et al.*, 2004, Davidson *et al.*, 2004).

A survey done by the National Health and Nutrition Examination Survey has found that females ingesting seafood on a regular basis tend to have high levels of mercury in their blood system, as opposed to females who do not ingest seafood regularly, and that this places the infants carried by them at risk for neurological damage (Mahaffey *et al.*, 2004). Thimerosal is a mercury based preservative used in various vaccines (Miles, 2011). The underlying mechanisms in mercury induced toxicity are still unknown. Some studies have reported that methyl mercury can easily cross blood-brain barriers and accumulate in the brain at high concentrations (Clarkson and Magos, 2006). Methyl mercury has also been reported to have interactions with a very broad range of cellular targets which results in multiple cellular function disruptions (Farina *et al.*, 2011).

It has thus been proposed that the increase of autism might have been caused by the worldwide increased exposure of mercury through seafood's, fish and industrial sources and through increased exposure of thimerosal (Mutter *et al.*, 2005).

Drug influences- Prenatal stage: The prenatal stage for a newborn is the most critical state because brain development is at its most vital stage of the maturation process. The developing brain is at risk during this time in the womb as it can be subjected to damage and underdevelopment when fed with harmful addictive substances such as alcohol, cocaine and other chemicals, as this has been observed with fetal alcohol syndrome, and also in the etiology of Autism (Baird *et al.*, 2003).

Prenatal exposure to ethanol has been demonstrated to be associated and considered a risk factor for the development of behavioral abnormalities in children (Middleton *et al.*, 2012). It has been demonstrated that ethanol exposure affects the amygdala and ventral striatum causing deficits in social behavior in a way that is time, age and gender dependent (Mooney and Miller, 2001). The prenatal exposure to cocaine in rats was demonstrated to have an induced alteration in the brain serotonin system. The drug cocaine is an addictive chemical and causes differential changes in neuroendocrine responses (Carrasco *et al.*, 2006). Serotonin has been implicated in all the different behavioral types and perturbations of this system and can elicit alterations to those behavioral types (Frazer and Hensler, 1999).

The use of antibiotics during pregnancy has also been reported to be a strong risk factor for autism in the developing brain. A study showed twofold increased risk of infantile autism when mothers experienced an immune activation during the pregnancy (Atladóttir *et al.*, 2012). Another chemical or drug that poses risk includes Valproic acid (VPA). VPA is a known teratogen and when exposed during pregnancy has demonstrated to cause a threefold increase in the rate of developmental problem which are often observed with autism (Ornoy, 2009). An animal model induced during prenatal exposure to VPA has thus been proposed to study autism. This study demonstrated numerous behavioral abnormalities in the VPA rats (Bambini-Junior *et al.*, 2011). Another study also showed that VPA exposure rats showed several behavioral abnormalities which are also found in autism (Mychasiuk *et al.*, 2012).

1.4.2. Genetic factors

Genetics factors play a predominant role in autism pathogenesis. Several studies have given convincing results for the involvement of genetic factors in the development of autism

(Folstein and Rutter, 1988, Bailey *et al.*, 1995, Korvatska *et al.*, 2002, Lauritsen *et al.*, 2005, Waltes *et al.*, 2014). It has been reported that gene mutations, deletion, substitutions or additions have resulted in diseases and disorders. A copy-number change by definition is the deletion or duplication of a piece of DNA and can range from a kilo-base pair to mega-base pairs and or could even include the whole chromosome (Feero *et al.*, 2012). Deletions may be heterozygous, in which case one of the usual two copies is missing; homozygous, when both copies are missing; or hemizygous (e.g., X-chromosome deletions in a male patient) (Feero *et al.*, 2012).

Twin studies: Studies done on autistic twins and families have shown that it is heritable and that the recurrent risk of PDD in siblings of children with autism is 2 to 8% (Muhle *et al.*, 2004, Chaste and Leboyer, 2012) and is ~50 to 100 times greater than individuals with no history of the disorders or minor forms of it (Bolton *et al.*, 1994, Philippe *et al.*, 1999). It was demonstrated that twins who share 100% identical genes, (monozygotic (MZ)) in comparison to twins who only share 50% identical genes, (dizygotic (DZ)) ten to have a much higher (about 82%) occurrence for this disorder than the other (about 10%). The first genome-wide scans for genes implicated in the etiology for autism implicated numerous chromosomes which included chromosomes 4, 5, 6, 7,10, 16, 19, 22 (Philippe *et al.*, 1999). Many more chromosomes have been identified since, however the most popular implicated would be chromosome 7. This chromosome has been labeled as a "hotspot" for genes of interest as it harbors a number of candidate genes table 1.1.

Chromosome 5 has also been previously associated with autism (Gillberg, 1998) and is even more significant since two genes Cadherin 9 (*CDH9*) and Cadherin10 (*CDH10*) have been

implicated in the etiology of autism (Wang *et al.*, 2009, Prandini *et al.*, 2012). These genes transcribe adhesion peptides that are very important for cell-cell signaling expressed in the frontal cortex, associated with autism (Flintoft, 2009).

Table1.1: Genes previously associated with ASD

Chromosome	Genes	Gene name	Reference
2	*SLC25A12*	Solute Carrier Family 5 member 12	(Ramoz *et al.*, 2004)
5	*CDH10*	Cadherin gene 10 type II	(Prandini *et al.*, 2012)
6	*GL01*	Glyoxalase	(Junaid *et al.*, 2004)
6	*GRIK2*	Glutamate receptor Inotropic kainite 2	(Kim *et al.*, 2007)
7	*HOXA1*	Homeobox A1	(Ingram *et al.*, 2000)
7	*RELN*	Reelin	(Sharma *et al.*, 2013)
7	*WNT2*	Protein Wnt-2 precursor	(Wassink *et al.*, 2001)
7	*FOXP2*	Forkhead-box Protein P2	(Li *et al.*, 2005)
7	*EN2*	Engrailed homeobox 2	(Gharani *et al.*, 2004)
7	*UBE2H*	Ubiquitin-conjugating enzyme E2H	(Vourc'h *et al.*, 2003)
10	*PTEN*	Phosphatase and tensin homolog	(Butler *et al.*, 2005)
11	*HRAS*	Harvey rat sarcoma viral oncogene homolog	(Comings *et al.*, 1996)
12	*AVPR1A*	Arginine vasopressin receptor 1A	(Wassink *et al.*, 2004)
12	*CACNA1C*	Calcium channel, Voltage-dependant, L type alpha 1C subunit	(Lu *et al.*, 2012)
15	*UBE3A*	Ubiquitin protein ligase E3A	(Glessner *et al.*, 2009)
15	*ATP10C*	ATPase, Class V, type 10C	(Herzing *et al.*, 2001)
15	*GABRB3*	Gamma-aminobutyric acid (GABA) A receptor, subunit beta 3	(Mccauley *et al.*, 2004)
17	*SLC6A4*	Solute carrier family 6 (neurotransmitter transporter, serotonin), member 4	(Sutcliffe *et al.*, 2005)
17	*NF1*	Neurofibromin 1	(Marui *et al.*, 2004)
17	*CACNA1G*	Calcium channel, voltage-dependant, T type, alpha 1G subunit	(Lu *et al.*, 2012)
19	*APOE2*	Apolipoprotein E	(Persico *et al.*, 2004)
20	*ADA*	Adenosine deaminase, RNA-specific	(Bottini *et al.*, 2001)
22	*SHANK3*	Shank 3	(Yasuda *et al.*, 2011)
22	*CACNA1I*	Calcium channel,voltage-dependant, T type, alpha 1I subunit	(Lu *et al.*, 2012)
X	*NLGN3*	Neuroligin 3	(Jamain *et al.*, 2003)

1.5.Diagnosis and Treatment

Thus far, autism has been diagnosed only by behavioral patterns by psychologists or psychiatrists. It is observed in children between the ages of 2-3 years and it was noted that the child does not socially interact easily or at all with peers or family members. Common features that may suggest that a child may be autistic includes a) lack of a social smile and b) appropriate facial expressions c) poor or no attention, d) impaired social interactions e) preference to aloneness and f) ignoring people (Baird *et al.*, 2003). A clinician would screen the child using the DSM-IV manual and other tests to confirm the diagnosis.

There is no known medical cure for these disorder only treatments (Lee and Kong, 2012). Individuals with autism who tend to have erratic behavior are medicated to reduce the risk of harming themselves or others. The medication previously used includes a selection of typical anti-psychotic, atypical anti-psychotics, antidepressants, stimulants, anti-convulsants and others. Respiredone is a commonly used atypical antipsychotic but others include Olanzapine and Clozapine (Broadstock and Doughty, 2003). Commonly used treatments are dietary, modifications, music therapy and many other types of behavioral therapy, which have been used to calm an autistic individual (Grant, 2011). The efficacy however is quite limited from one to another because there is a huge difference in severity and symptoms (Lee and Kong, 2012). Sometimes it becomes frustrating for the families to deal with the behavior changes, and attempt to find solutions for example, a case recently reported the plight of a father who explored the idea of using roundworms. The reason being that roundworm modulates the inflammatory immune responses, which alleviated his son's disruptive behavior (Grant, 2011). Although this may be a breakthrough it still does not become a cure due to the diverse nature of autism.

1.6. Association studies

Genetic association studies or association studies are investigated in order to find the correlation between a disease status and a genetic variation that result in the identification of candidate genes or genome regions responsible for that disease (Lewis and Knight, 2012). These studies are conducted by determining the genetic variants like single nucleotide polymorphisms (SNPs), copy number variants (CNVs) and structural variants present in different groups. The genome-wide association method has been scientifically convincing and has underlined numerous and novel biological pathways in disease processes (Knight, 2009). The present study conducted aimed at incorporating this method to establish and determine the variation of selected markers or SNPs in the autistic SA population.

1.6.1. Single nucleotide polymorphisms (SNPs)

The human has 23 chromosomes that constitute its genome. It is large and vast and has been studied extensively for decades in the hope of understanding how the human species operate or sustain itself on a molecular level. SNPs by definition, are "a source of variance in a genome, a single base mutation in DNA and are the most simple and common source of genetic poly-morphism in the human genome" (Smith, 2002). When cells divide and multiply the genome is copied as well and during this process random mutations occur; deletion, substitution or addition of nucleotides. These mutations can then be transferred from generation to generation that allows for its tracking. Since the mutation is in the direct path of DNA sequence that is related to protein, the change in sequence thus results in change in transcript and can results either in the transcription of a faulty or a less active protein. Research has shown that differences on the DNA level (inherited), adds to phenotypic variations, affecting the anthropometric characteristics of an individual, and the risk of disease (Sachidanandam *et al.*, 2001). The frequency of variations occurring between

members of a species or between paired chromosomes in an individual is >1% and provides geneticists with a powerful tool for the study of human medical genetics (Wang *et al.*, 2008). SNPs are the most frequent type of variation in the human genome and thus yields as a very powerful tool and marker for genetic studies (Wang *et al.*, 1998).

1.6.2. SNPs studies on autism

Due to the high frequency of SNPs in the human genome it has created an opportunity for researchers to track and establish genetic variations on a large scale in cohort's studies. Genome wide association studies (GWAS) have investigated thousands of SNPs in thousands of individuals and have thus been able to implicate common genetic variants in more than 80 human diseases (Hindorff *et al.*, 2009). These results have generated large amounts of data extracted from assayed SNPs that rendered very significant insights with respect to associations made between loci and common diseases that may not have previously been shown to share etiological pathways. Other benefits were the associations made in chromosomal regions or spaces that were considered to be deprived of functional genes or genetic loci, (Manolio *et al.*, 2008, Hindorff *et al.*, 2009). These studies have been very useful by categorizing key characteristics of observed associations and complex human disorders (Hindorff *et al.*, 2009). This has allowed the assessment of autism, and a few genes have been successfully implicated by scanning of the genome. Some of these genes of interest include Engrailed Homeobox 2 (*EN2*), Serotonin (*SERT)*, Reelin (*RELN*), and Cadherin 10 type II (*CDH10*) (Benayed *et al.*, 2009, Wang *et al.*, 2009, Gadow *et al.*, 2013, Sharma *et al.*, 2013).

Engrailed Homeobox 2 gene (*EN2):* The Engrailed Homeobox 2 transcription factor came to light when certain similarities were observed with regards to the development of the cerebellum in mouse brains. The *EN2* gene is located on the human chromosome 7q36 and has been linked to ASD (Benayed *et al.,* 2005). This location has been under investigation as it seemed to house other interesting genes that have important roles in language and speech development, regulation of brain cell formations and primary development of bodily organs and particularly the nervous system. *EN2* gene is also responsible for the development of the cerebellum in humans when comparative studies were done between mice and an autistic human sample. A study done by Gharani (2004) and coworkers focus on four SNPs for the *EN2* gene. The results for that study showed that two intronic SNPs, rs1861972 (*P=0.0018)* and rs1861973 (*P* = 0.0003) had strong signal for autism with a haplotype p value of 0.000005 (Gharani *et al.,* 2004). In addition, Benayed and coworkers (2005) reinforced the significance of the SNPS and demonstrated in a population-attributable risk test, that the risk allele contributed to 40% of ASD cases in the general population.

Serotonin gene (*SERT):* The *SERT* gene (5-HTT) encodes for a serotonin transporter, which is "involved in the pre-synaptic reuptake of serotonin to terminate and modulate the serotonergic neurotransmission" (Nakamura *et al.,* 2000). The dysfunction of this protein have been implicated in other psychiatric disorders which includes anxiety, depression and more over is mainly focused on mood change and behavior which are co-morbid symptoms of ASD (Nakamura *et al.,*2000). In recent studies association of this gene to autism is strengthening in SNP studies. Gadow and coworkers (2013) reported that the SNP rs25531 for the SERT gene may modulate severity of ADHD and ASD symptoms for subjects with ASD.

Reelin gene *(RELN):* The *RELN* gene located on the distal long arm of chromosome 7 has also been associated with ASD primarily because of its role in brain development. It has been reported that the *RELN* protein plays a pivotal role in the migration of several neuronal cell types and in the development of neural connections (Skaar *et al.*, 2004). It has been observed that *RELN* mRNA and protein levels are significantly reduced in multiple brain areas of patients with schizophrenia and bipolar disorder with psychosis (Impagnatiello *et al.*, 1998, Guidotti *et al.*, 2000, Skaar *et al.*, 2004). It is for this reason that it was considered as a candidate gene in the pathogenesis of ASD as these other psychiatric disorders are all co-morbid with ASD and their symptoms all fall under the same spectrum of complex disorders. Sharma and coworkers (2013) considered the potential implications of this gene to autism and conducted an association study in SA. A significant association of the intronic SNP (rs736707) with autism was observed in the SA population.

Cadherin 10 type II gene *(CDH10):* Similarly like the other candidate genes, *CDH10* is a gene that plays important roles in brain development, in the early stages of fetal growth. It has been under investigations in a GWAS that resulted in the detection of a very significant SNP (rs4307059) found between the inter-genic regions of genes *CDH10* on chromosome 5 (Pourcain *et al.*, 2010). This gene encodes a calcium dependent cell surface glycoprotein that is involved in cell adhesions and cell signaling during early stages of development. The gene *CDH10* form part of a subgroup of the family known as classic cadherins type II that mediate strong cell-cell homophilic interactions and are localized to the nervous systems and functions for regionalization of the brain, formation, maintenance and the plasticity of neural circuits (Redies and Takeichi, 1996, Hirano *et al.*, 2003, Takeichi, 2007, Hulpiau and Van

Roy, 2009, Etzrodt *et al.*, 2009). Two individual association studies investigated this gene for associations in different population settings. In a European study, the cohorts were genotyped using an Illumina HumanHap550 BeadChip which contained up to 550,000 SNP markers. Only six SNPS showed strong signals for association of which the most significant SNP was SNP rs4307059 with a *P* value of 3.4 $X10^{-8}$ (Wang *et al.*, 2009). This same result was observed in an family cohorts specific for Italian ancestry. The SNPS rs4307059 displayed strong association once again with P values of 0.0017 (Prandini *et al.*, 2012).

1.7. Application of proteomics in autism research

Proteomics is the comprehensive investigation and characterization of proteins, through biochemical methods (Pandey and Mann, 2000). This method functions on the concept of proteins being the down-stream products of gene expression which are the active agents in cells and as a results is useful in the drug development arena and therein diagnostics (Pandey and Mann, 2000).

Proteomics is defined as the study of a proteome, the total protein content of one cell or all cells of an organism where cells are from the bodily fluids e.g. blood/urine/sweat or tissues or organs and bodily parts (Dihazi and Müller, 2007). This term was originally established by Dr. Marc Wilkins of the Macquarie University, a PhD student at the time, denoting proteomics as the protein quota for a known genome (Verrills, 2006).

It is a study that investigates isolated proteins and is aimed at identifying proteins that may have or cause un-welcomed effects within a system. Figure 1.2 demonstrates the relationship between DNA and proteins. Proteins are the downstream results of genes and their expression

and thus have very important roles within its system (Genetics Home reference, 2011). Since a large amount of DNA sequences have been investigated and collectively accumulated in databases, it has been observed by researchers that these enormous loads of data on complete sequences does not suffice in understanding basic biological operations, and therefore directed their attention towards proteins as there are no strict linear link or path between genes and their resulting products-proteins (Pandey and Mann, 2000). Functional genomics thus focuses on the characterizations of sub-cellular interactions of proteins and macromolecular complexity of cellular organisms (Monteoliva and Albar, 2004). The human genome has successfully been sequenced in 2003 with 600 000 DNA base pairs, about 20 000-25 000 genes all located on 24 chromosomes (The human genome project, 2003).

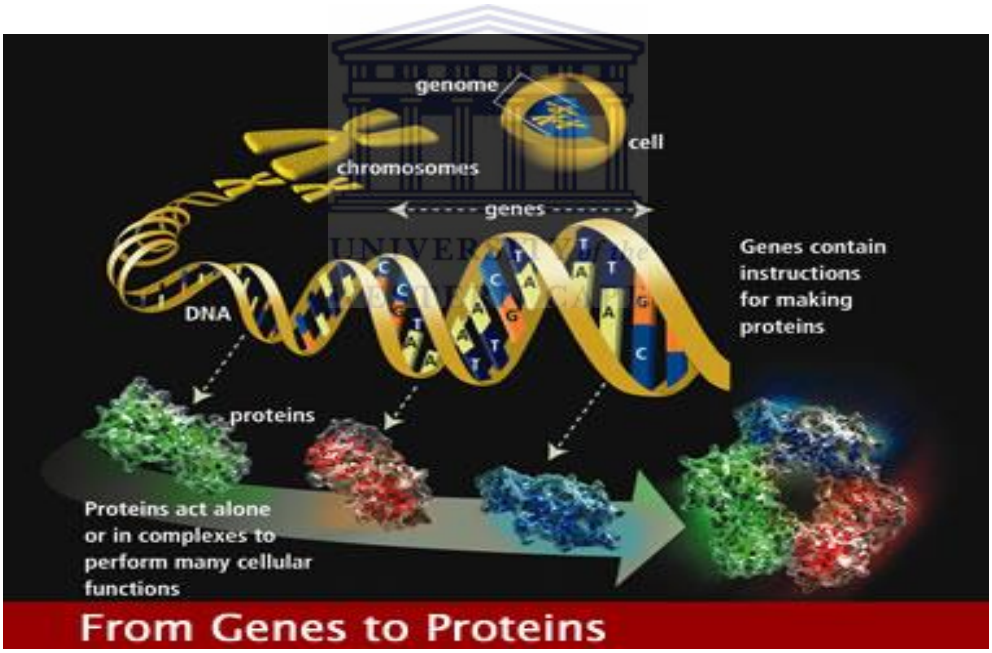


Figure 1.2: Genes to proteins (The human genome project, 2003).

The proteome of an organism is very unique as the protein profile tends to change in response to different stimuli, both extra and intra cellular stresses including, cell cycles, temperature etc., because of this reason, more researchers gaze towards a proteomic approach in molecular science (Dihazi and Müller, 2007). This approach has become the leading

technology in understanding various conditions and physiological states of different tissues in both prokaryotes and eukaryotes (Oh *et al.*, 2005). There are three main areas in proteomics, one: Protein micro-characterizations two: differential displays and three: Protein-protein interactions which focus on the identification for proteins and their post translational modifications, comparisons of protein levels and their interactions respectively (Dihazi and Müller, 2007). There are three areas which one can pursue the investigation of proteins viz expressional, structural and functional proteomics (figure 1.3). Each method harboring different ways of isolating, identifying and analyzing a sample.

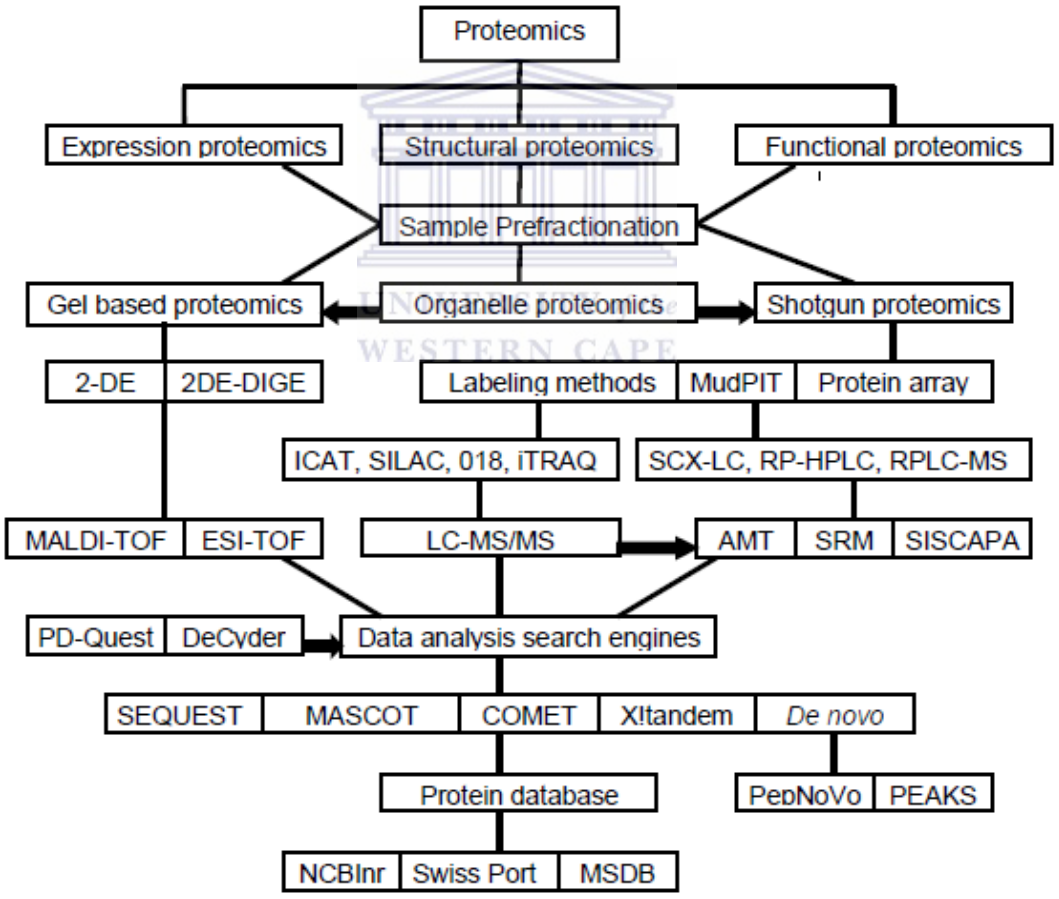


Figure 1.3: Proteomic studies and approaches (Adapted from Chandramouli and Qian, 2009).

Depending on the what the end result should provide, one can utilize a "gel-based and non-gel-based approach" by which the proteins can be characterized, the one deals in resolving proteins by 2 Dimensional Poly Acrylamide Gel Electrophoresis, (2D-PAGE) and Mass Spectrophotometry (MS) that operates by investigating the mass to charge ratio of the target protein, e.g. MALDI- TOF. The other involves the processing of the isolated protein sample through high thorough-put technology including Multi-Dimensional Protein Identification Technology (MUDPIT), Liquid Chromatography (LC), and the combination of Liquid Chromatography to Tandem mass spectrometry (MS/MS) but not limiting the methods to it-self.

1.7.1. Gel-based and non-gel based

Gel based proteomics have been frequently used due to the fact that it allows for differential expression analysis and thus allows for discrimination of protein expressions between healthy and otherwise diseased/induced or experimental sample sets. It incorporates the display of proteins on what is called acrylamide gels for simple visualization and mapping differences in a 2nd dimension (Monteoliva and Albar, 2004). Primarily proteins are separated by its molecular mass with one dimensional electrophoresis (1DE). Two-dimensional electrophoresis (2DE) involves the separation by Iso-electric focusing (IEF) first, then separation by molecular mass (Verrills, 2006). The proteins are separated according to their net charge (pI) on a pH strip in a horizontal direction, followed by a separation in the second dimension according to molecular mass vertically on an acrylamide gel and is very useful since it allows for reproducibility (Verrills, 2006). The 2D separation is usually done on a poly-acrylamide matrix and has variations or different mediums through which it can be done and the protein profile is visualized using a selective stain or dye such as Coomassie brilliant blue R-250 stain.

Complex protein mixtures can either be separated on “native gels” non-denaturing gels, Sodium Dodecyl Sulfate Poly acrylamide gels (SDS-PAGE), also known as denaturing gels, or be subjected to 2 dimensional difference gel electrophoresis (2DIGE). The difference between the three types of gels is that with native gels, even though used extensively for complex mixtures, separates the proteins with intact quaternary structure, size and charge, the SDS gels, does the opposite and is preferred method, since it removes the charge and linearizes the proteins before separation and thus separated according to the size or length of the proteins. The difference between these two methods and 2DIGE is the simple addition or labeling of proteins with radioactive isotopes.

Each method has its advantages and disadvantages which is dependent on what the aim of a study would be. While Gel- based methods are perfect for biomarker discovery it also allows for separation of large numbers of proteins and peptides and comparative analysis of two sample sets. Whereas gel-free methods are advantages for analysis of samples with very little quantities and are much more automated than the laborious gel based method. Table 1.2 outlines the differences between these two methods, their benefits and drawbacks.

UNIVERSITY of the
WESTERN CAPE

Table 1.2: Gel based and non-gel based methods compared (Modified from Dihazi and Műller, 2007).

	Gel-based methods		Gel free methods		
	2 Dimensional Gel Electrophoresis (2D electrophoresis)	Difference Gel Electrophoresis (DIGE)	Surface Enhanced Laser Desorption Ionization (SELDI)	Liquid Chromatography –Mass Spectrometry (LC-MS)	Capillary Electrophoresis-Mass spectrometry (CE-MS)
Advantages	-Separation of large number of proteins - Easy biomarker identification	- Allows for quantification -Separation of both experimental and control sample mix on single gel -Less amount of protein sample required -Use of internal standard possible	-Quantification possible -Automated -High throughput analysis -Less protein sample required	-Quantification possible -Automated -High throughput analysis -Less protein sample required -Sample fractionation LC	-Quantification possible -Automated -High throughput analysis -Less protein sample required - Sample fractionation CE
Disadvantages	-Requires large sample amounts -Lack of automation -limitation for large and small proteins -Loss of Hydrophobic proteins -Not useful screening tool -Not reproducible -Laborious	-Lack of automation -limitation for large and small proteins -Loss of Hydrophobic proteins -Expensive -Laborious	-Drawbacks in detecting low abundance proteins -Lack detection of high molecular weight proteins -Problematic protein and post translational modification identifications -Protein fractionation required	-Drawbacks in detecting low abundance proteins -Lack detection of high molecular proteins	-Drawbacks in detecting low abundance proteins -Lack detection of high molecular weight proteins -Proteins identification problematic -Deliver complex patterns -Pattern recognition not a single biomarker

1.7.2. Mass Spectrometry Technology (MS)

The identification of molecules, peptides, ions of any small or micro sized level has great importance when it comes to biological research. Mass Spectrometry or MS is a technology that identifies any micro molecule, compounds through the determination of the physical mass to particle charge ratio (*m/z*) by means of ionizing the molecules and then subjecting it into an electric field for detection (Proteomics Shared Research, 2011). Figure 1.4 shows the basic flow for mass spectrometry analysis. A sample is ionized in a chamber, followed by sorting or separations based on mass and lastly detection chamber that picks up each peptide and relates the result.

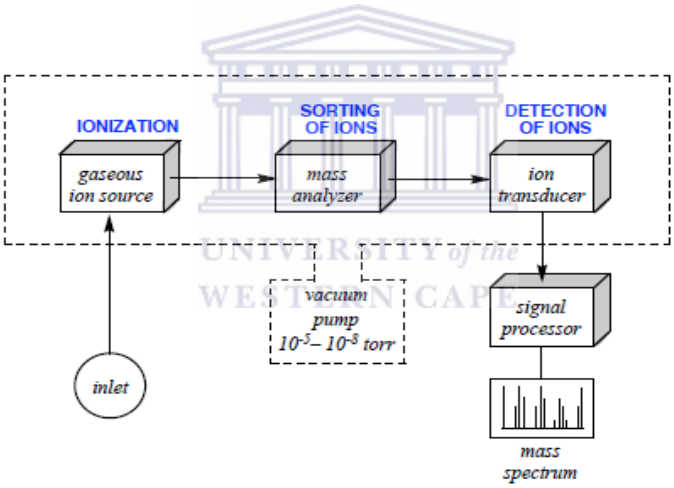


Figure 1.4: Mass spectrophotometer basic components, an ion source, mass analyzer and a detector (Aebersold and Mann, 2003).

There are various ways of ionizing a sample, of which two, Electrospray ionization (ESI) and Matrix Assisted Laser Desorption/Ionization (MALDI) is most commonly used, others includes, chemical ionization (CI), Atmospheric Pressure (APCI), Electron Impact (EI), Fast Atom Bombardment (FAB), Field Desorption/Field Ionization (FD/FI) and Thermospray Ionization (TI) (Proteomics Research Shared, 2011). The basic MS consists of an ion source,

a mass analyzer and the detector (Aebersold and Mann, 2003). The sample is ionized the analyzer measures or determined the *m/z* value and the detector registers the numbers of ions at each m/z value that appears as peaks in a spectrum (Aebersold and Mann 2003). Electrospray is a method that incorporates the streaming of liquid containing the sample of interest and applying an electrical charge to it, thereby ionizing it (Proteomic shared research, 2011). MALDI utilizes a laser which ionizes and vaporizes the sample, followed by a MS analysis. Likewise, as there are different Ion sources, there are different analyzers and detectors. The most common analyzers include Quadruples, Time-of-Flight Magnetic sectors and more while different detectors would include Electron multiplier and Faraday cup. The combination of these technologies thus renders potential information data or results and therefore the method will be employed in this study.

The methods employed in this thesis were used in hope of determining genetic variants that influences the outward phenotype amongst individuals, which includes association studies and proteomic studies. The association studies used single nucleotide polymorphism (SNPs) of a specific gene and the probability was determined and the probability of their association was calculated in the SA autistic population. In the proteomic study urine proteins were isolated and evaluated as potential biomarkers in SA African autistic individuals.

1.8. Aims of project

The aims of this project are twofold:

1. To study the association of SNPs rs4307059 and rs4327572 for the *CDH10* gene in the healthy and autistic SA population.
2. To generate a proteome profile for African autistic individuals that could shed light on the differential expression of proteins between SA autistic and a healthy group.

Chapter 2

Abstract

Autism spectrum disorders (ASD) is a complex behavioral disorder that are primarily of genetic origin (Schanen, 2006). The aim of part of this study was to investigate genetic association of SNPs rs4307059 and rs4327572 for the *CDH10* gene with autism in the SA population. A non-significant association was observed between these SNPs and autism status in SA population. The Chi-square test was used to study the association of these SNPs. The TT, CT, CC genotypes for rs4307059 had prevalence of 71.2%, 20.8% and 8% for the healthy subjects while the autistic population had distribution percentages of 65.4%, 28.7% and 5.9%. The rs4327573 genotype percentages for CC, CT and TT were 1.9%, 26.8 % and 71.3 % for healthy SA population compared to the autistic population percentages of 4.2 %, 27.8% and 68.1%. There were no significant association between both SNPs (rs4307059 and rs4327572) and the status of autism in the South African population.

Genetics of autism in a SA population

Table of Contents

2. Genetics of autism

2.1 SNP studies on autism in South Africa (SA)

An abundant amount of SNP association studies have been generated from SNP data of which most have been investigated in other populations (Manolio *et al.*, 2008, Hindorff *et al.*, 2009). The importance of identifying or making these associations of genes relates to the etiology of the ASD disorder. Predominantly the genetics of the SA population is diverse and requires these types of studies to exclude or include genes in the etiology of ASD in SA. The *CDH10* gene in particular has recently been implicated in ASD in the broader European and Italian population with strong association (Wang *et al.,* 2009; Prandini *et al.,* 2012).

2.1.1 The Cadherin 10 type 2 gene *(CDH10)*

The cadherins comprise a family of transmembrane glycoproteins (Juliano, 2002). The members of this family are categorized according to specificity of binding. Predominantly cadherins are homophilic and have roles in adhesion; other members may be heterophilic and have very poor binding (Redies and Takeichi, 1996). These molecules mediate calcium-dependent cell-cell adhesion rendering them very important for cell-cell interactions (Suzuki *et al.*, 1991). Up until 1991, only three types of cadherins had been characterized namely Epithelial (E)-, Neural (N)-, and Plasmental (P)-cadherin (Suzuki *et al.,* 1991). It's been nearly three decades since the identification of this molecule and now there are more than 20 members that comprise this family. Cadherins are classed based on molecular structure into Classical Cadherins (type I and II), Desmosomal Cadherins, Protocadherins and others (Redies and Takeichi., 1996). Table 2.1 is a list of cadherins members identified in the human genome.

Table 2.1: Cadherins members found in the human genome (Modified from Pòtter *et al.*, 1999)

Name [a]	Symbol [b]	Locus [c]	Main expression in	Type [d]
E-Cadherin, Uromodulin (mouse), L-CAM (chicken)	CDH1	16q22.1	Epithelial	I
N-Cadherin	CDH2	18q11.2	Neuronal cells	I
P-Cadherin	CDH3	16q22.1	Skin, prostate, placenta	I
R-Cadherin, cadherin-4	CDH4		Retina, brain	I
VE-Cadherin, cadherin-5	CDH5	16q22.1	Vascular endothelium, brain	II
K-Cadherin; cadherin-6	CDH6		Fetal kidney/brain	II
Cadherin-7	CDH7		Brain	II
Cadherin-8	CDH8		Brain	II
Cadherin-9, T1-cadherin	CDH9		Brain	II
Cadherin-10	CDH10		Brain	II
OB-Cadherin, cadherin-11 (mouse, rat, chicken)	CDH11	16q22.1	Osteoblasts, brain, testis, lung, placenta	II
Br-Cadherin, cadherin-12	CDH12	5p13-p14	Brain	II
H-Cadherin, cadherin-13, T-Cadherin (chicken, human)	CDH13	16q24.2	Heart, brain, aorta	O
M-Cadherin, cadherin-15	CDH15	16q24.3	Muscle, brain	I
Ksp-Cadherin	CDH16	16q21-q22	Kidney	O
HPT1, L1-cadherin (rat)	CDH17	8q22.1	Liver, intestine	O
Cadherin-18, cadherin-14	CDH18		Brain	II
Pcdh1, protocadherin 42	PCDH1	5q31-33?	Brain, melanoma cells	PC
Pcdh2, protocadherin 43	PCDH2	5q31-33?	Brain, fibroblast	PC
BH-Protocadherin	PCDH7	4p15	Brain, heart	PC
Desmoglein 1	DSG1	18q12.1	Skin	D
Desmoglein 2	DSG2	18q12.1	Epithelia	D
Desmoglein 3	DSG3	18q12.1	Skin	D
Desmocollin 1	DSC1	18q12.1	Skin	D
Desmocollin 2	DSC2	18q12.1	Epithelia	D
Desmocollin 3	DSC3	18q12.1	Skin	D
RET	RET	10q11.2	Neuronal derived cells	O
hFAT	FAT	4q34-q35	Epithelia, T-cell lines, Fibroblasts, keratinocytes	O
FIB1, FIB2, FIB3			Fibroblasts	O
ME1 to ME6			Melanoma cells, melanocytes	O

[a] Name of human cadherin or its species homologue
[b] HGMW-approved gene symbol according to GenBank
[c] Chromosomal location of the human gene
[d] Five subgroups of cadherins: I, II classical type I and II, respectively; D, Desmosomal cadherins; PC Protocadherins; O, Other related cadherins.

This family and its members still branches off into more subfamilies of which at least 80 members have been shown to be expressed within a single mammalian species, of these includes classic cadherins (Pötter *et al.*, 1999). Cadherins, the major groups, type II predominantly have been shown to be expressed in the brain in the frontal cortex region (Pötter *et al.*, 1999).

The brain is central to the central nervous system (CNS) and since autism is a "polygenic neurological disorder that affects the CNS" (Benayed *et al.*, 2009), thus investigation of connections within the CNS, can lead to clarity on the pathogenesis of autism. This CNS is complex and highly organized neural networks that generate brain function (Yagi and Takeichi, 2000). Ultimately cadherins are synaptic components with roles in neuronal circuitry, synaptic junction formation and synaptic plasticity (Suzuki *et al.*, 1997, Tang *et al.*, 1998, Honjo *et al.*, 2000, Manabe *et al.*, 2000, Tanaka *et al.*, 2000, Yagi and Takeichi, 2000) and are known to act alongside catenins, which are polypeptide molecules that strictly bind to these synaptic components (figure 2.1). Catenins are cytoplasmic proteins composed of β-catenins, α-catenins and p120-catenin. β-catenins binds to the tail of the cadherin molecule where it acts as an anchor for the α-catenins and p120- catenin binds to membrane region of the cadherin. These three molecules have functions of signal transducers, intracellular trafficking and regulating actin dynamics and lastly regulating cell locomotion (Niessen and Leckband, 2011).

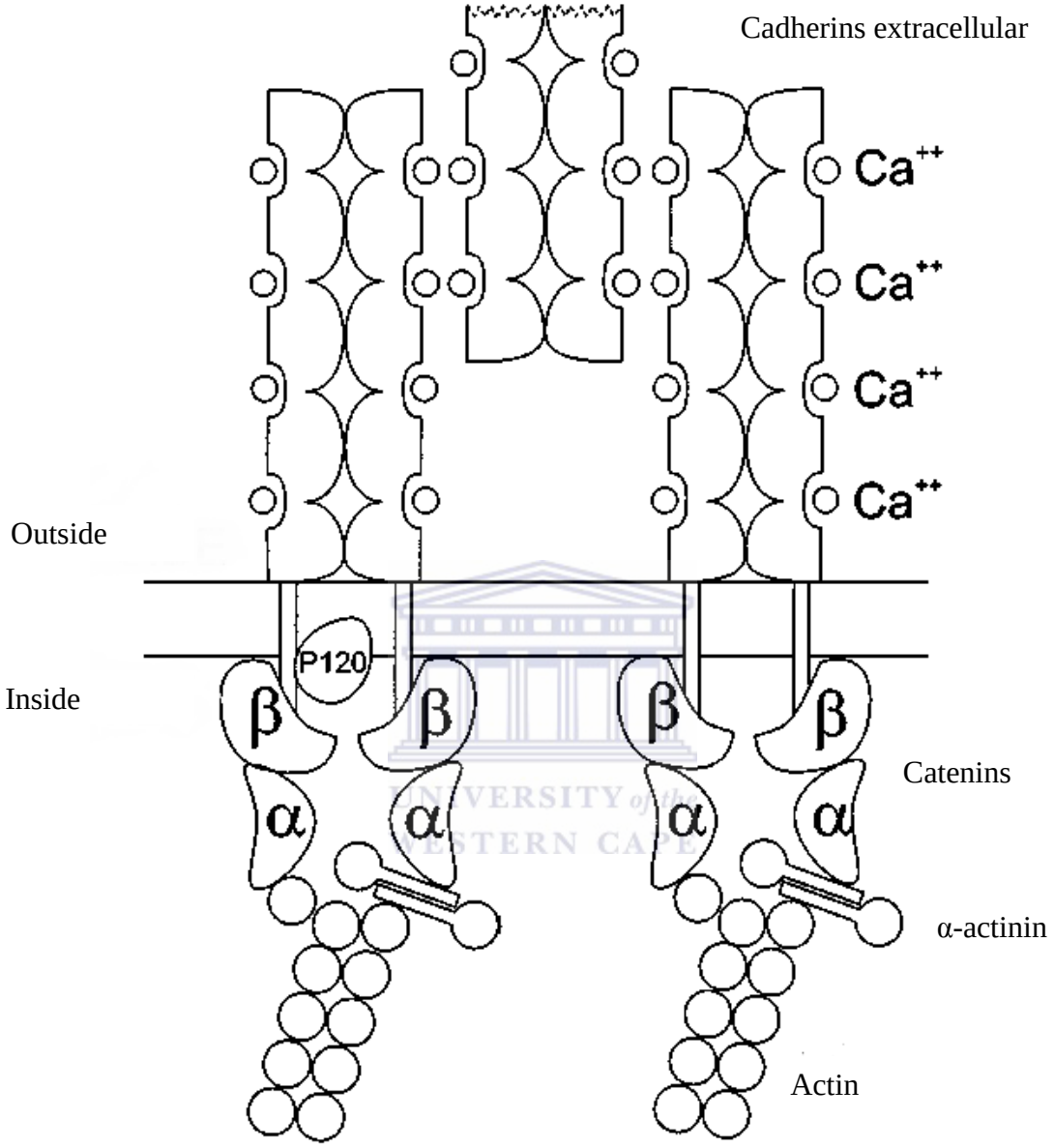


Figure 2.1: Cadherin structure and localization in synaptic cell-to-cell adhesion junctions. The cadherin-catenin complex are composed of the cadherin molecule bound to β- and α-catenins which aids binding to actin containing cytoskeleton (Juliano, 2002).

2.1.2 The genomic and proteomic organization of *CDH10*

CDH10 resides on chromosome 5 between locus regions 24,487,209 base pairs (bp) and 24,645,087 bp. The *CDH10* gene is 157.88kb in length and is found on the reverse strand of chromosome 5 and is conserved in other species such as primates (figure 2.2).

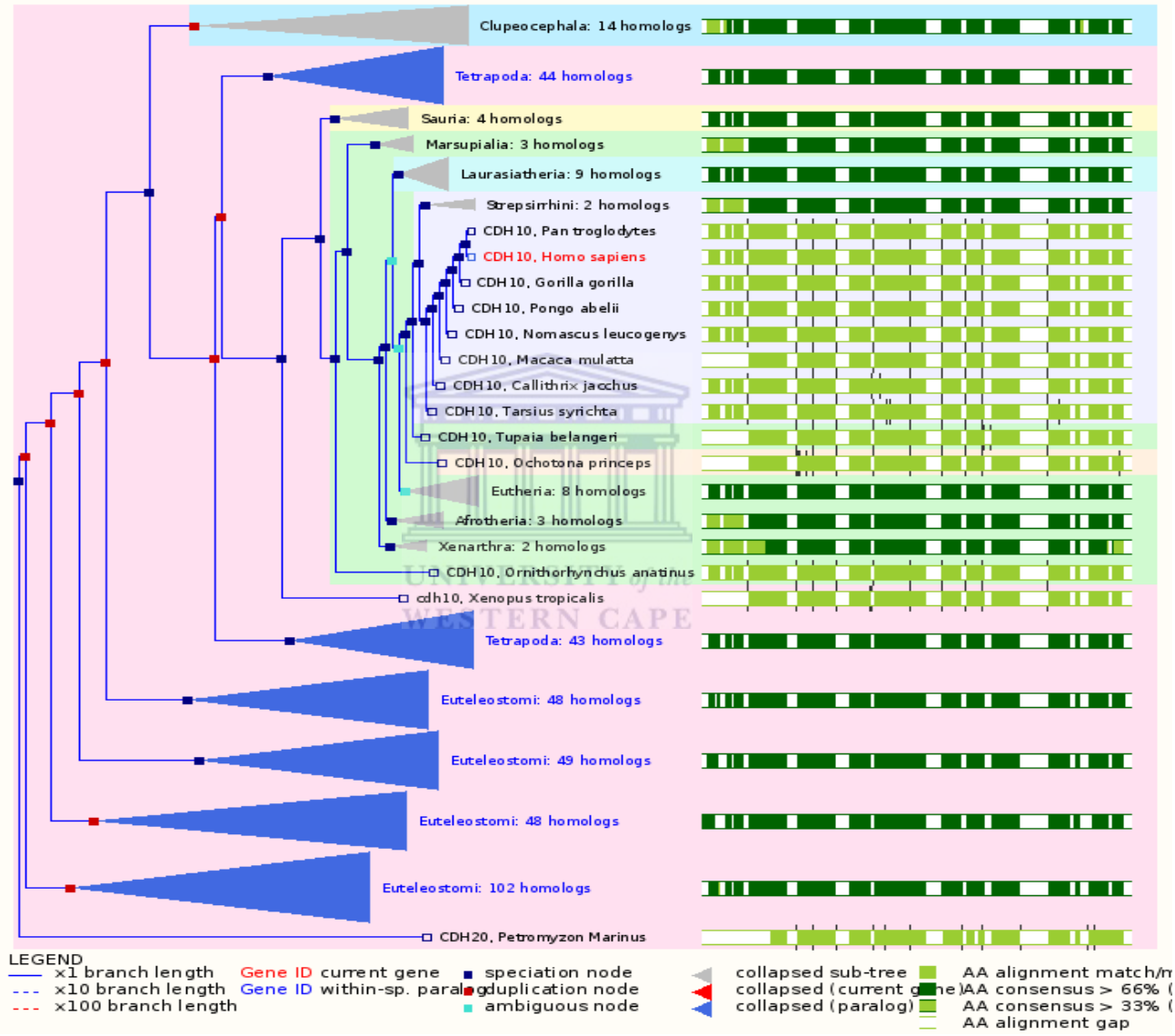


Figure 2.2 : Gene Ancestry and homology of cadherins (Ensemble, 2011)

One of the closest species homologs to the human is the common house mouse, *Mus musculus*, (Not shown in the figure above) with associating roles like cell adhesion, plasma membrane calcium ion binding (The gene ontology, 2011). The gene is conserved between,

its genomic organization and is closely linked and conserved between family members of the cadherin family as some members may derive from 12 exons, other vary up to 16 exons (Yagi and Takeichi, 2000). Classic cadherins including *CDH10* are highly conserved and has 5 tandemly repeated extracellular domains that are interrupted by multiple introns while some protocadherins are not (Hatta *et al.*, 1988, Gumbiner, 1996, Strehl *et al.*, 1998, Wu and Maniatis, 1999).

As previously mentioned, *CDH10* type II is a classic cadherin and is unique as it has up to 5 conserved extracellular domains (EC domains) and a single cytoplasmic domain, all connected by a transmembrane segment (Takeichi, 1990, Uemura, 1998, Wu and Maniatis, 1999). The EC domains engage specific to homophilic molecules rendering its specificity, as it may only bind to other EC domains with similar motif structures (Nose *et al.*, 1988, Wu and Maniatis, 1999) whereas the cytoplasmic domains renders the molecules to interact with β-Catenins in the plasma membrane that plays a pivotal role in "epithelial-mesenchymal transition during development" (Aberle *et al.*, 1996, Wu and Maniatis, 1999). The EC domains separates cadherin members from each other as the amount of EC domains may vary due to negatively charged sequenced motifs (DXD, DRE, and DXNDNAPXF) that have roles in Ca^{2+} binding (Takeichi 1990, Yagi and Takeichi,2000).

2.1.3 *CDH10* function and expression

CDH10, a cadherin, forms part of a multitude of cell-cell adhesion molecules that connects the most important cells required for brain signaling neurons. It is reported that synapses have physiological and adhesive functions and neurons communicates with each other through them (Yagi and Takeichi, 2000). Although found in synapses, cadherins and catenins have not been found to be present in the adherent junction zones of synapses yet appear in

surrounding areas (Yagi and Takeichi, 2000). These Cadherin and catenins complexes are aligned symmetrically over the pre- and postsynaptic plasma membrane and tend to act as a kind of "lock-and- key component for regulating specific interneuronal connections, as hypothesized by Sperry (1963)" (Uchida *et al.*, 1996, Yagi and Takeichi, 2000).

Data on this gene's expression is limited in comparison to other family members. The expression pattern for this gene has been investigated in other species such as chickens and rats. These studies monitored the CNV and the limbic system for gene expression whilst also studying the total RNA from different human samples (Kools *et al.*, 1999, Liu *et al.*, 2006). In a study conducted by Shimoyama *et al.*, (2000), cDNA clones were sequenced for the *CDH10* and *CDH7* gene. Results of this study showed that for *CDH10* and *CDH7*, the clones consisted of 3261 and 3060 nucleotides, respectively and that it shows very high homology of 96 and 94% to chicken. The expression for the *CDH* family is both spatially and temporally regulated, and research found that these molecules indicate vital roles in the development of the vertebrate nervous system (Nakagawa and Takeichi, 1995, Radice *et al.*, 1997, Lele *et al.*, 2002, Malicki *et al.*, 2003, Masai *et al.*, 2003, Liu *et al.*, 2006). Furthermore a test for cell-cell binding activity showed that *CDH10* not only associates with β-catenins but also with α-catenins which confirms the cadherin-catenin complex (Shimoyama *et al.*, 2000).

It is clear that the gene is relative more expressed in the brain as reported by Kools *et al.*, (1999). An investigation of this protein's RNA expression was observed in most of the brain-derived tissues which are constantly monitored when doing studies on ASD as it is for the most part, the most effected body part in ASD (Kools *et al.*, 1999).

2.1.4 Genomic location of *CDH10*

CDH10 resides on chromosome 5 (figure. 2.4), “one of the largest chromosomes” in the human genome with numerous intra-chromosomal repetitions but consists of very few gene densities (Schmutz *et al.,* 2004). A chromosome, 177.77 base pairs long that houses about 923 gene loci and is associated with at least 80 human diseases and associated to 66 specific loci (Schmutz *et al.*, 2004) (figure. 2.3). Most of these diseases are related to cancer, but also disorders of the mind (e.g.; Schizophrenia, Attention-deficit hyperactivity disorder (ADHD) right across to diseases of muscles (e.g.; myopathy, spinal muscular atrophy), skin (e.g.; Hermansky-Pudlak syndrome), and systems (e.g.; severe combined immunodeficiency) are also linked to this chromosome; (figure 2.4). An interesting fact is that schizophrenia and ADHD are found to have susceptible loci on this chromosome and are co-morbid disorders with autism as all the symptoms fall under the spectrum of ASD.

The sequencing of this chromosome in 2001 relayed clarity about the gene/s existing on the chromosome (Fig 2.3). Chromosome 5 is known for the gene/s families it harbors some include Interleukin gene family and Protocadherins. The 'interleukins' are a family of polypeptides belonging to the family of 'cytokines', i.e. cell communications (Malkovský *et al.,* 1988). Protocadherins, belong to the super family of cadherins. These glycoproteins play a significant role in the formation of neuronal connections in the human adult brain (Takeichi, 1995, Gumbiner, 1996, Shapiro and Colman, 1999, Treffert, 1999, Tepass, 1999, Wu and Maniatis, 2000).

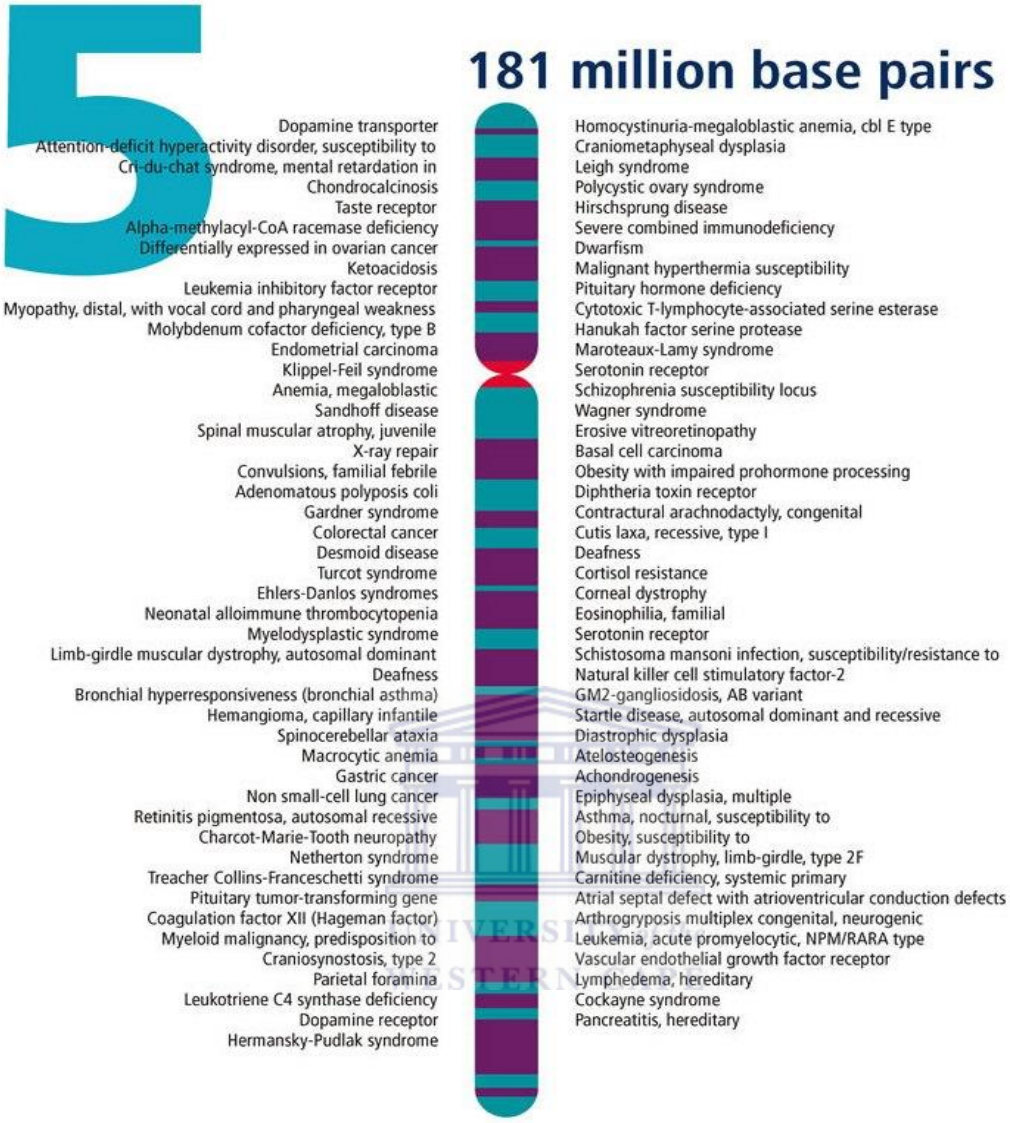
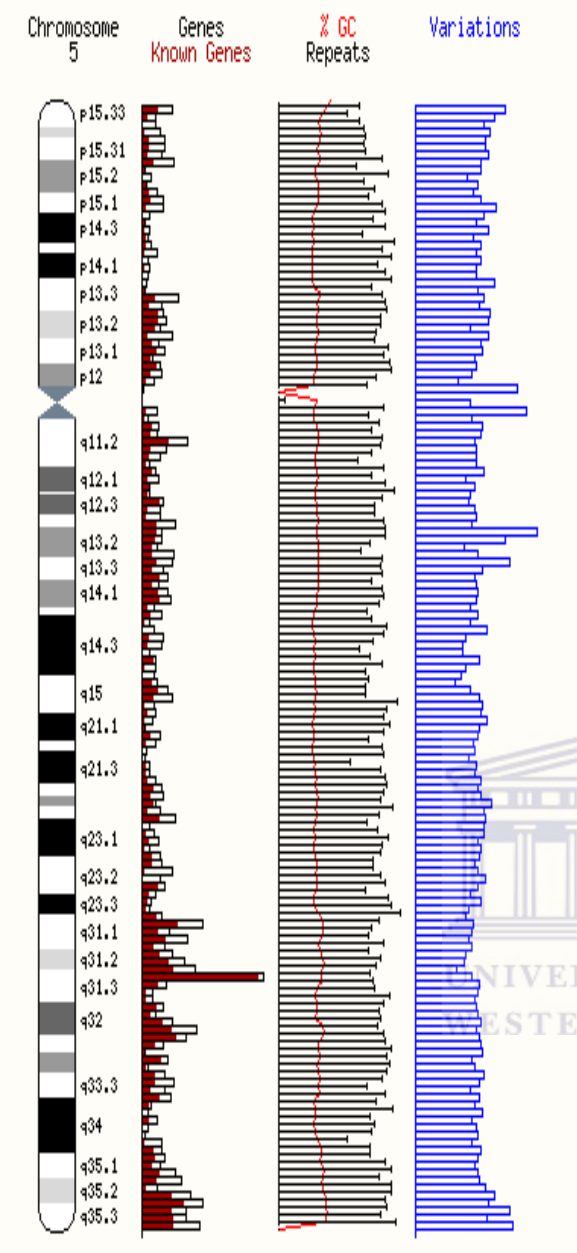


Figure 2.3: Chromosome 5 and gene- and trait associated diseases (The Gene Gateway Workbook, 2011)



Length (bps):	180,915,260
Known Protein-coding Genes:	845
Novel Protein-coding Genes:	27
Pseudogene Genes:	650
miRNA Genes:	83
rRNA Genes:	25
snRNA Genes:	106
snoRNA Genes:	61
Misc RNA Genes:	68
SNPs:	1,991,185

Figure 2.4: The fully sequenced human chromosome 5 summary, (Ensemble, 2011)

2.1.5 Protein dysfunction of *CDH10* and relevance to autism

In the case of a mutated gene, the end results sometimes in a mutated protein which in turn causes a problem as the protein cannot perform its original functions. Mutations of genes have been known to be the causative agent for various medical conditions. An example of

one such mutation would be that of the aryl hydrocarbon-interacting protein (*AIP)* gene. It has been shown that the mutation of this gene in pituitary adenomas to be a causative agent for acromegaly also known as 'gigantism' disease (Chahal *et al.*, 2011).

The *CDH10* gene encodes for a cell adhesion molecule previously stated in section 2.1.1 above. The protein expressed by this gene is required to interact with β-catenins, which means that in the case of an ineffective protein expressed, connection between membranes could be ineffective as well. Looking at the structure of the gene, the EC domain and CP domain and their functions, it is easy to deduce that loss of function or decreased function will cause overall disruption by inhibiting binding which would result in the underdevelopment of the CNS. *CDH10* along with *CDH9* and other cadherins have been investigated as these genes codes for "cell-adhesion molecules which has potential roles in the pathogenesis of ASDs" (Wang *et al.,* 2009). With a loss of function these molecules cannot assist in the development at the embryonic stage which results in typical symptoms displayed in ASDs. The study also concluded that *CDH10* gene showed more strong significant SNP association as compared to other genes like neurexin and contactin (Wang *et al.,* 2009).

2.1.6 *CDH10* SNP selection

According to the NCBI website, approximately 3298 SNPs exist for the *CDH10* gene. Disorders that have been associated with the gene would include pancreatic cancer, prostate cancer and ASD (The National Centre for Biotechnology Information, 2013). A study on neurodevelopment showed that CDH10 and CDH9 (with SNPs rs4307059) were associated with suicide (Chojnicka *et al.*, 2012).

A genome wide association study in 2009 focused on selected genes implicated in ASD. According to Wang *et al.,* (2009), the 5p14.1 locus of the fifth chromosome held strong genetic variants associated with ASD. This study was conducted on two cohorts; one of 780 families with children, second cohort of 1204 autistic individuals and 6491 non-autistic individuals with European ancestry. The study found six SNP variants between genes *CDH10* and *CDH9* yielded strong signaling for ASD pathogenesis of the SNPs rs4307059, had the most significant signal (P= $3.4X10^{-8}$). Amongst these there were five other SNPs with p values ranging between 1.7 $X10^{-5}$ to $6.2X10^{-7}$ in the discovery cohorts (Wang *et al.,* 2009).

SNPs (rs4307059 and rs4327572) were selected for this study for this reason. Rs4307059 and rs4327572 are two inter-genic SNPs found at positions 26003460 and 26008578 respectively on chromosome five (Wang *et al.,* 2009). The SNP rs4307059 has an orientation plus N with minor to major allele of C/T. Rs4327572 has an orientation plus N with minor to major allele of T/C.

2.2 Hypothesis

It is hypothesized that there is an association between the prevalence of SNPs (rs4307059 and rs4327572) of the *CDH10* gene with autism status in the SA population.

2.2.1 Aims

- The intention for this study was to determine if there is an association between SNPs (rs4307059 and rs4327572) of *CDH10* gene and the status of autism in the SA population.

2.2.2 Objectives

- To genotype the SNPs rs4307059 and rs4327572 of the *CDH10* gene in the healthy and autistic SA population
- To determine the association of allele and genotype distribution for the SNPs rs4307059 and rs4327572
- To determine the association of SNPs rs4307059 and rs4327572 with autism in SA population.

2.3 Materials and methods

Materials

15ml Greiner tubes	Greiner Bio-One GmbH, Frickenhausen Germany
384-well micro plates	Applied Biosystems, California USA
96-deep-well micro plates	Applied Biosystems, California USA
Applied Biosystems 7900HT Real time PCR	Applied Biosystems, California USA
BuccalAmp™ Extraction Kit	Epicentre® Biotechnologies, Wisconsin USA
Microfuge eppMotion 5070	Applied Biosystems, California USA
Nanodrop® ND-1000 Spectrophotometer	Nanodrop® Technologies ThermoScientific, Massachusetts USA
Perkin Elmer 7900 PCR system	Applied Biosystems California, USA
Swabs	Pur wraps MDCL Ltd. VWR International, Pennsylvania, USA

Quickextract™ DNA Extraction solution	Epicentre® Biotechnologies, Wisconsin USA
Taqman® SNP Custom Genotyping Assay	Applied Biosystems, California, USA
TaqMan® Universal Genotype Master Mix	Applied Biosystems, California, USA

Methods

2.3.1 Ethical clearance

The study protocol was approved by Western Cape Ethics Committee (SR: 5/09/32) of University of the Western Cape and Western Cape Education Department. Consent was also obtained from principals of schools and parents for the collection and genetic analysis of DNA samples, in accordance with the Declaration of Helsinki (Appendix I).

2.3.2 Sample selection criteria and collection

Healthy group: Samples were collected from neuro-typical healthy students (8-18 years of age) from local schools between the ages of 8 and 18 years of age. This group comprised of SA subjects of 101 white, 114 black and 118 mixed ancestries.

Autistic group: The diagnosis was carried out strictly on the basis of criteria outlined by the DSM-IV (Association, 1994). The Childhood Autism Rating Scale (CARS) was administered for the assessment of cases (Schopler *et al.,* 1986). Selection of subjects in the study was based on clinical and psychological evaluation by experienced psychologist and psychiatrist. The autistic group comprised of 86 white, 80 black and 68 mixed ancestries unrelated SA children.

Collections: The samples were collected from autistic individuals studying in schools for children with special needs in Cape Town and Gauteng Province whereas the controls samples were collected from schools in the Western Cape. The collection entailed the individual's cheek being swabbed with a sterile swab. This action, rubbing of the swab against the inside of the individual's cheek for one minute was done in duplicate. The swabs were placed in sterile 15 ml Greiner tubes then labeled and stored at -20^{o}C.

2.3.3 SNP selection and primer design

Due to a large number of identified SNPs *CDH10*, using the Applied Biosystems website a SNP assay could therefore be generated. The SNPs rs4307059 and rs4327572 for the gene *CDH10* were identified from literature of a previously conducted experiment in the European population by Wang *et al.,* (2009). Using SNP genotyping, Taqman® technology, the selected SNPs can be made to order using the NCBI database.

The Primers for these SNPs were pre-designed and obtained from the Applied Biosystems Website (Applied Biosystems, 2010).

The TaqMan® Custom genotyping assay were labeled VIC and FAM to the sequences

Rs4307059:

5‘-AGCTTTCACTGATGTGTCCGAATTG C-3‘and

5‘-TTCATGTAACCAGGATATTTTCCAT- 3‘

Rs4327572:

5‘-ATTTTATAAATACTTATAAAGCAAA C -3‘ and

5‘-TAAAACAGCAAAATATGAAAAAGACA -3’

2.3.4 DNA extraction

Using BuccalAmp™ Extraction Kit, DNA was extracted from the samples collected with swab. This was done by removing the swab tip and transferring it to a 2 ml Eppendorf tube and adding a 250 µl of QuickExtract™ DNA solution 1.0 followed by a 1 minute incubation period at 65^{o}C. Extraction continued with vortexing the tube for 15 seconds, incubated for 4 minutes at 98^{o}C with a further 15 second vortexing and stored at -20^{o}C until quantification.

2.3.5 DNA quantification

DNA concentration was determined by using a NanoDrop®ND-1000 Spectrophotometer.

2.3.6 PCR preparation

Following quantification, the DNA samples were diluted (final concentration 20 ng/µl) into a 96 deep-well plate with double distilled water. To the first row in the plate, 50 µl of ddH_2O was aliquoted to function as negative control, (row labeled A). A PCR reaction sample mix was then prepared for samples as follows:

1680 µl of Taqman ®Universal genotyping Master mix were added with 840 µl of double distilled water in a tube and vortexed.

84 µl of 20x Taqman® SNP Custom Genotyping Assay and 84 µl of double distilled water were added to an Eppendorf tube and vortexed. The tube was covered with foil as the assay is light sensitive. Following the vortexing, the samples were mixed and with the use of the microfuge eppMotion 5070 automated pipetting system, were then transferred to 384-well plates along with the master mix to a final volume of 5 µl.

2.3.7 PCR amplification, parameters and detection

The PCR reactions were amplified using a Perkin Elmer 7900 PCR with the following conditions. An initial holding step was carried out for 2 minutes at 50^{o}C followed by 40 cycles of: Denaturing for 10 minutes at 92^{o}C followed by annealing for 15 seconds at 60^{o}C and extension at 60^{o}C for 60 seconds. The resulting amplified products were read on the

Applied Biosystems 7900HT Real-Time PCR System and recorded. Although 333 controls and 234 autistic samples were genotyped, not all rendered a signal.

2.3.8 Statistical analysis

These frequencies were subjected to Chi Square testing for the C and T alleles in the SA healthy and the SA autistic groups. Probabilities for associations were done using computational methods (Graphpad.com, 2011) where a $P < 0.05$ are significant.

2.4 Results

For SNPS rs4307059, the present study group was comprised of typed case (188, unrelated autistic children) and typed controls (212, unrelated healthy children) where the respective figures for rs4327572 were 72 and 209 respectively. For statistical correction, some of the subjects were excluded from the actual number of collected subjects.

2.4.1 SNPs rs4307059 and rs4327572 of *CDH10* gene

The following results were obtained from the Taqman® SNPS custom Genotyping Assays.

2.4.1.1 rs4307059 Allele and genotype data

The allele and genotype data were collected for the SNP rs4307059 on the *CDH10* gene. The data were obtained for both the healthy and the autistic SA population and the clusters represented in the figure 2.5 (a and b) indicates the distribution for the genotypes in the healthy and autistic SA population. In the SA healthy population the percentage distribution of the TT, CT, and CC genotypes were 71.2%, 20.8% and 8% respectively whereas the autistic population had distribution percentages of 65.4%, 28.7% and 5.9% respectively (table 2.2). Statistically, no significant differences were observed between the healthy and the autistic. The T allele and C allele frequency were 81.6% and 18.4% for the healthy group while the respective values among autistic subjects were 79.8% and 20.2% (table 2.2).

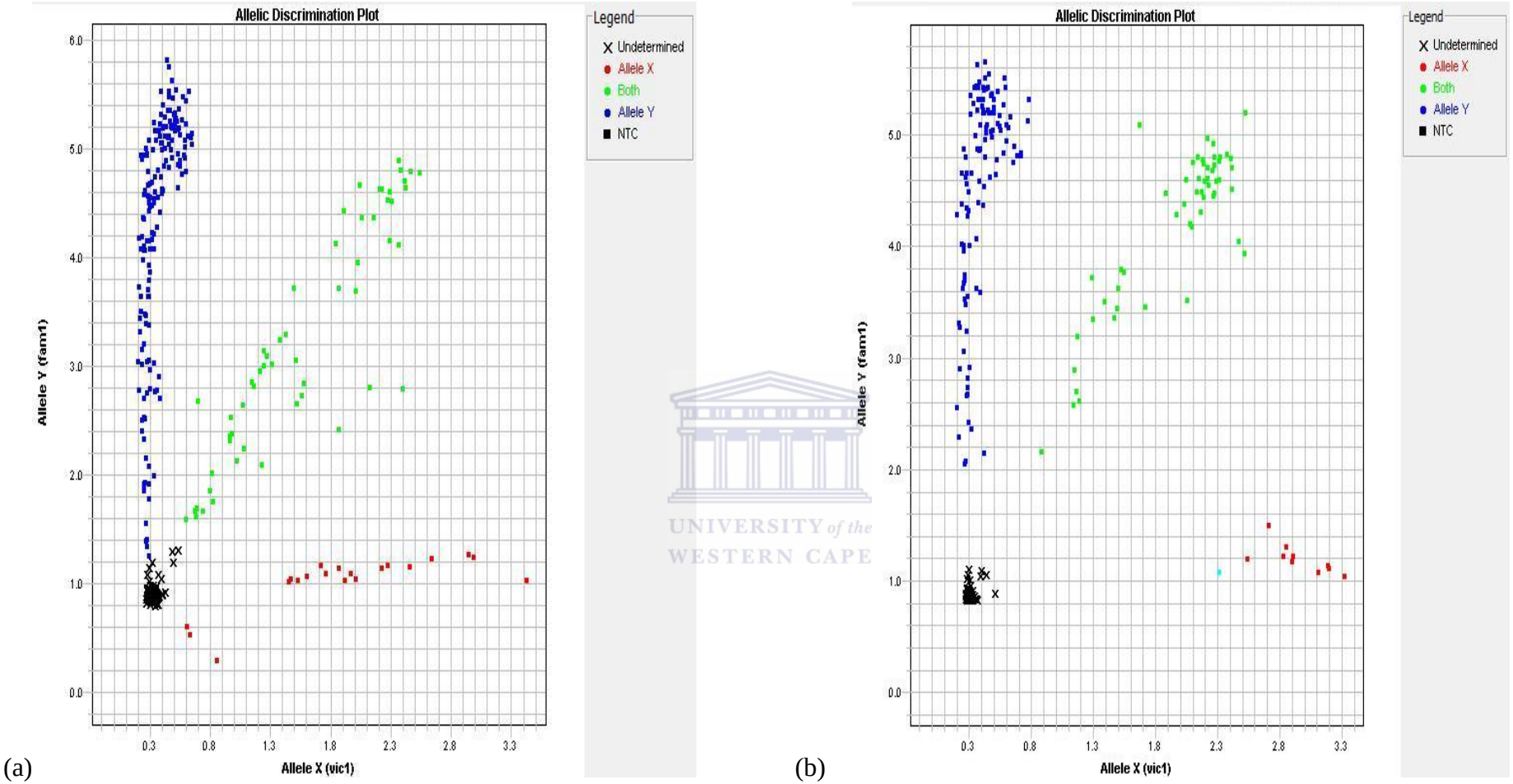


Figure 2.5: Allelic discrimination plot for the Taqman SNP rs4307059 Custom Genotyping assay for the (a) healthy and (b) autistic S A population. The blue, green and red cluster represents the homozygous and heterozygous genotypes which is TT, CT and CC respectively for this SNP in the healthy and autistic populations. The clusters thus represent the distribution for these genotypes in the population groups in this study.

Table 2.2: Genotype counts and allele frequencies in three ethnic groups, stratified by autism studies in SA population in normal individuals for rs4307059, (significant *P<0.05)

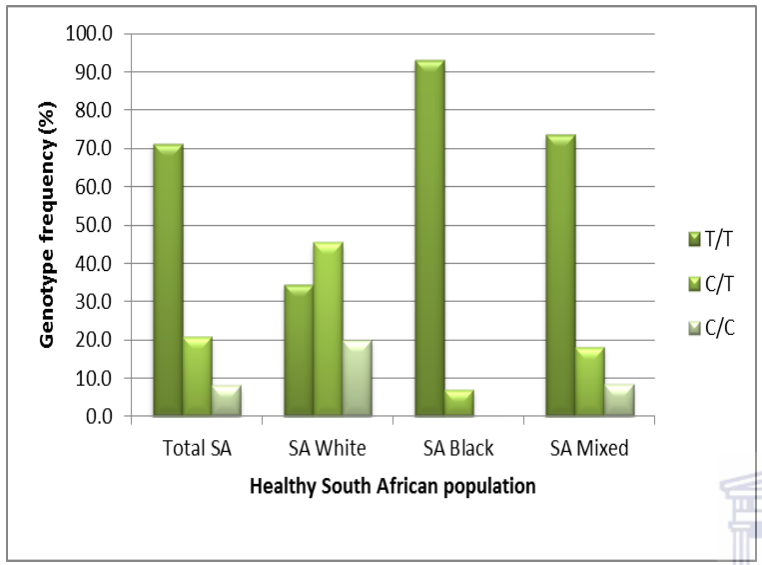
Ethnic group	Number	Genotype			P-Value	Allele (%)		P-Value
		TT (%)	CT (%)	CC (%)		T allele	C allele	
		Total South African						
Healthy	212	151(71.2)	44(20.8)	17(8.0)		346(81.6)	78(18.4)	0.8094
Autistic	188	123(65.4)	54(28.7)	11(5.9)	0.3054	300(79.8)	76(20.2)	
		SA White						
Healthy	55(26)	19(34.5)	25(45.5)	11(20)		346(81.6)	78(18.4)	0.1340
Autistic	69(37)	34(49.3)	28(40.6)	7(10.1)	0.1165	300(79.8)	76(20.2)	
		SA Black						
Healthy	85(40)	79(92.9)	6(7.1)	0		164(96.5)	6(3.5)	
Autistic	74(39)	65(87.8)	9(12.2)	0	0.5472	139(93.9)	9(6.1)	0.5638
		SA Mixed						
Healthy	72(34)	53(73.6)	13(18.1)	6(8.3)		119(82.6)	25(17.4)	
Autistic	45(24)	24(53.3)	17(37.8)	4(8.9)	0.0513	65(72.2)	25(27.8)	0.1672

Figures 2.6 and 2.7 illustrate genotype frequency percentage and the allele frequency percentage of autistic and healthy SA population. In the healthy SA white population the percentage distribution of the TT, CT, and CC genotypes were 34.5%, 45.5% and 20% respectively whereas the autistic population had distribution percentages of 49.3%, 40.6% and 10.1% respectively. Statistically, there were no significant differences observed between the healthy and the autistic genotypes for this group. The allele frequency for the T allele was higher with 81.6% and 18.4% for the C allele in the healthy group whereas the autistic had a slight decrease of the T allele having a 79.8% frequency and the C allele having an increase to 20.2%. There was no statistical significant association of distribution for the healthy and autistic genotyping in the SA population.

In the SA black healthy population, the percentage distribution observed for the TT, CT, and CC genotypes were 92.9%, 7.1% and 0%, respectively compared to the autistic population with distribution percentages of 87.8%, 12.2% and 0%. It was interesting to note that the black group does not have any CC genotype. In terms of statistical significance, no distinct significant differences were observed between the healthy and the autistic individuals. The frequency of the T allele was observed to be highest (96.5%) while the C allele had the lowest frequency (3.5%) as compared to other ethnic groups for the healthy group. The autistic of had a slight reduction in T allele frequency (93.9 %) and the C allele increased to a 6.1% frequency as compared to the expected frequencies in the healthy group. *P*-values among healthy and autistic individuals for allelic genotypic frequencies were 0.5638 and 0.5472, respectively. There was no statistical significant difference with respect to the allele and genotype frequencies for this ethnic group.

In the SA mixed race healthy population, the percentage distribution of the TT, CT, and CC genotypes were 73.6%, 18.1% and 8.3 % respectively, whereas the autistic population had distribution percentages of 53.3%, 37.8% and 8.9%. The T allele frequency had the same pattern with the higher percentages of 82.6 % and 17.4% C allele frequency in the healthy group. The autistic had a slight reduction with the T allele having a 72.2 % and the C allele, 27.8% frequency. The observed *P* value shows that there was no statistical significant difference among healthy and autistic subjects with respect to their genotype and allelic frequencies.

(a)



(b)

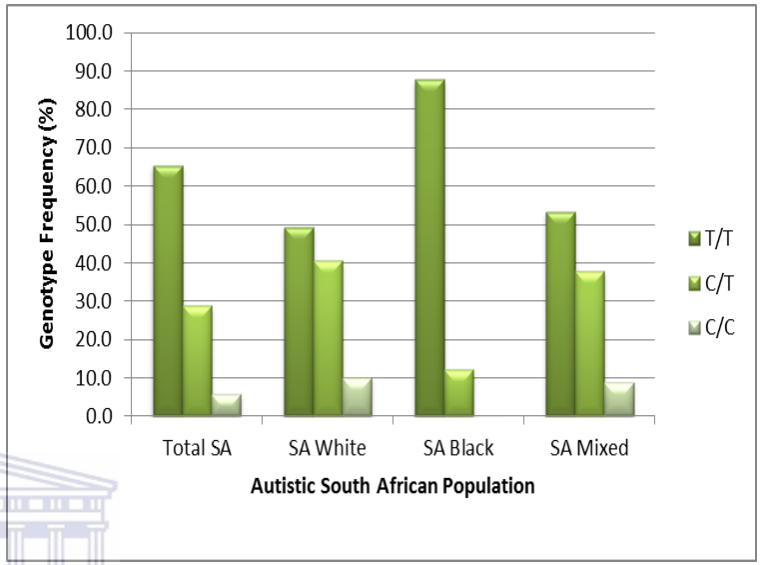
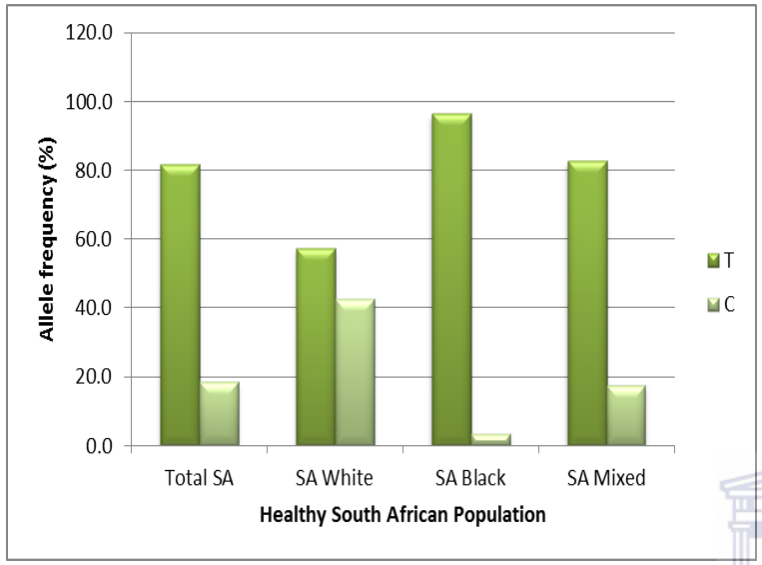


Figure 2.6: Percentage of genotype TT, CT and CC for the SNP rs4307059 of the *CDH10* in (a) healthy and (b) autistic SA population

(a)



(b)

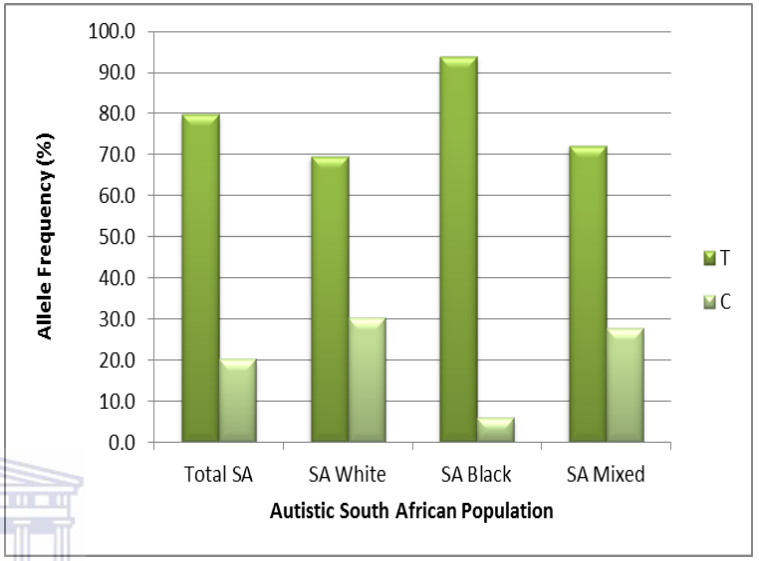


Figure 2.7: Percentage of the alleles C and T for the SNP rs4307059 of the *CDH10* gene in (a) healthy and (b) autistic SA population

2.4.1.2. rs4327572 Allele and genotype data

The allele and genotype data were collected for the SNP rs4327572 on the *CDH10* gene. The data was obtained similarly for both the SA healthy and the autistic population and the clusters represent the distribution of the genotypes in the healthy and autistic population (figure. 2.8).

In the SA healthy population, the percentage distribution of the TT, CT, and CC genotypes were 1.9%, 26.8% and 71.3% respectively, in comparison with the autistic population with distribution percentages of 4.2%, 27.8% and 68.1% (table 2.3). The frequency of the T allele was higher in autistic (18.1%) compared to the healthy (15.3%) while the allele frequency for the C allele was higher with 84.7% in healthy group and 81.9% in the autistic group. The allele and genotype frequency of this SNP indicated no significant association of either the allele or genotype frequency to autism with probabilities of $P = 0.7409$ and $P = 0.5516$.

In the SA white healthy population, the percentage distribution of the TT, CT, and CC genotypes was 1.7%, 35.6% and 62.7% respectively when compared to the white autistic group with a distribution percentage of 6.3%, 31.3% and 62.5% for those genotypes respectively. The T and C alleles had a frequency of 19.5% and 80.5% for the healthy group in comparison to the autistic with 21.9% and the C allele, 78.1 % frequency. No statistical significant association was observed for this SNP with autism in the white SA group.

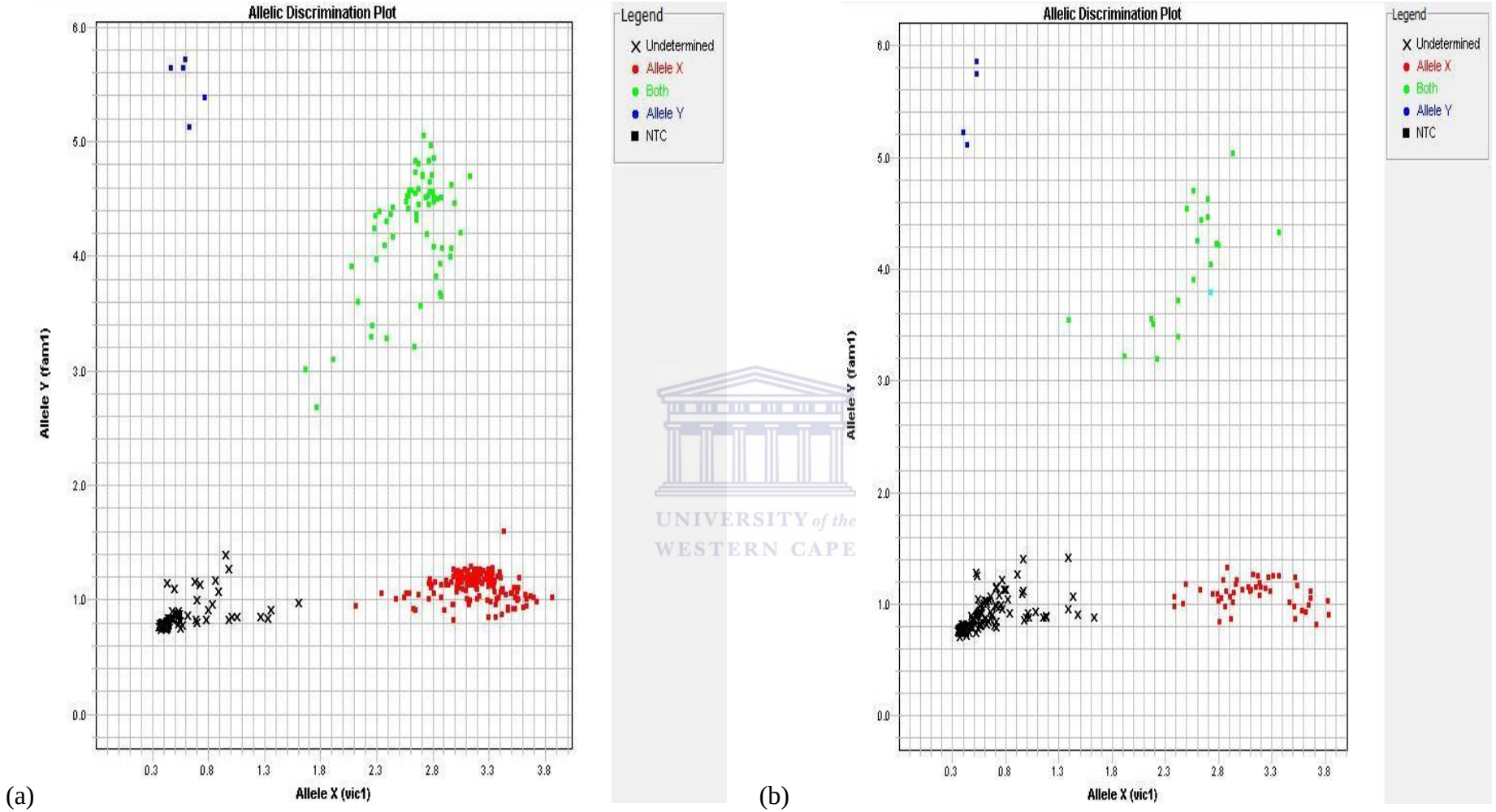


Figure 2.8: Allelic discrimination plot for the Taqman SNP rs4327572 Custom Genotyping assay for the (a) healthy and (b) autistic SA population

Table 2.3: Genotype counts and allele frequencies in the three different ethnic groups, stratified by autism studies in SA population and in compassionate normal individuals for rs4327572, (significant *P<0.05)

Ethnic group	Number	Genotype			P-Value	Allele (%)		P-Value
		TT (%)	CT (%)	CC (%)		T allele	C allele	
		Total South African						
Healthy	209	4(1.9)	56(26.8)	149(71.3)	0.5516	65(15.3)	354(84.7)	0.7409
Autistic	72	3(4.2)	20(27.8)	198(68.1)		26(18.1)	472(81.9)	
		SA White						
Healthy	59(28)	1(1.7)	21(35.6)	37(62.7)	0.4914	23(19.5)	95(80.5)	0.9298
Autistic	32(44)	2(6.3)	10(31.3)	20(62.5)		14(21.9)	50(78.1)	
		SA Black						
Healthy	67(32)	2(3)	13(19.4)	52(77.6)	0.7216	17(12.7)	117(87.3)	0.7718
Autistic	18(25)	0	3(16.7)	15(83.3)		3(8.3)	33(91.7)	
		SA Mixed						
Healthy	83(40)	1(1.2)	22(26.5)	60(72.3)	0.5012	24(14.5)	142(85.5)	0.6237
Autistic	22(31)	1(4.5)	7(31.8)	14(63.6)		9(20.5)	35(79.5)	

The percentage distribution of the TT, CT, and CC genotypes in the SA black healthy population were 3%, 19.4% and 77.6% whereas the autistic population observed distribution percentages of 0%, 16.7% and 83.3%. The T and C allele frequency witnessed a 12.7% and 87.3% frequency for the healthy group. Two interesting observations were that there were no TT genotypes in the black autistic group and that the % C allele is the highest in this group, i.e. 87.3% in the healthy and 91.7% in the autistic group. No significant association of genotype and allele frequencies with autism in the SA black population (*P*= 0.7216 and *P*=0.7718).

The SA mixed-raced healthy population had the percentage distribution of the TT, CT, and CC genotypes of 1.2%, 26.5% and 72.3 % respectively whereas the autistic population had 4.5%, 31.8% and 63.6% respectively. The allele frequency for the T allele observed 14.5% and the C allele an 85.5 % frequency for the healthy group. When one compares the % T and C allele between the SA mixed autistic group with SA black autistic group, the T allele is higher i.e. 20.5% compared to 14.5%. In the Case of the C allele, % C allele is higher in the SA black autistic compared to the SA mixed autistic group namely, 97.5% compared to 79.5%. No significant association of allele and genotypic frequencies with autism using chi square calculation (*P*=0.6237 and *P*=0.5012).

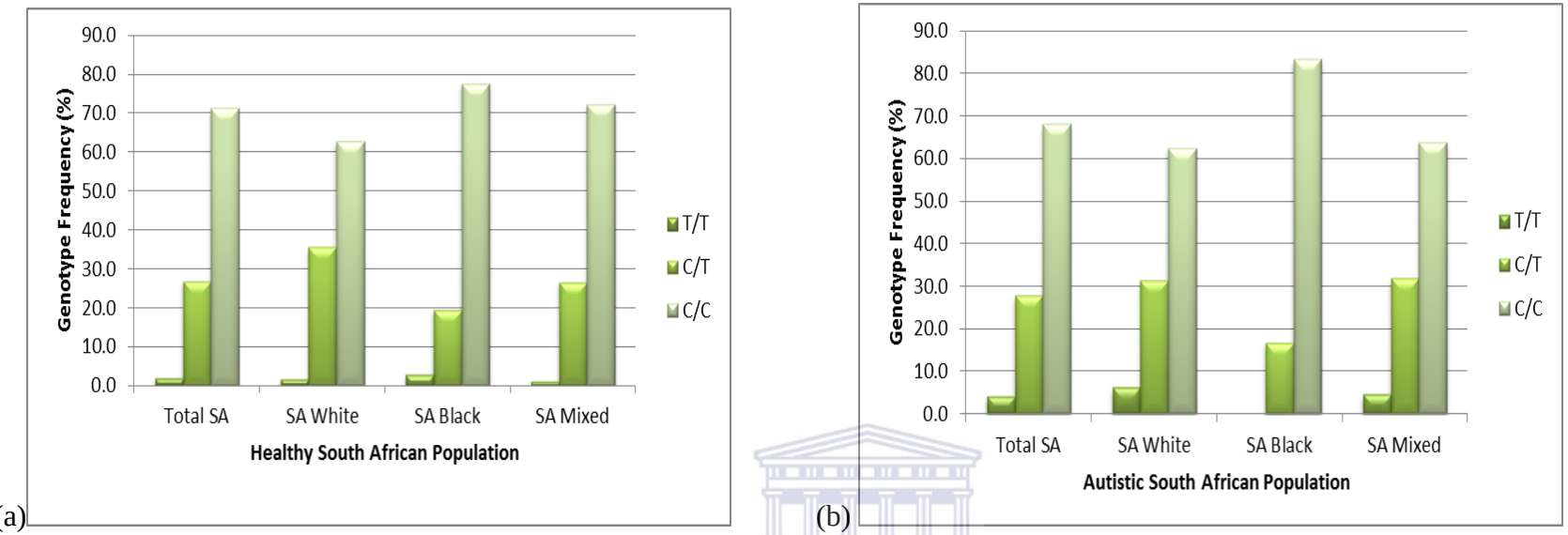
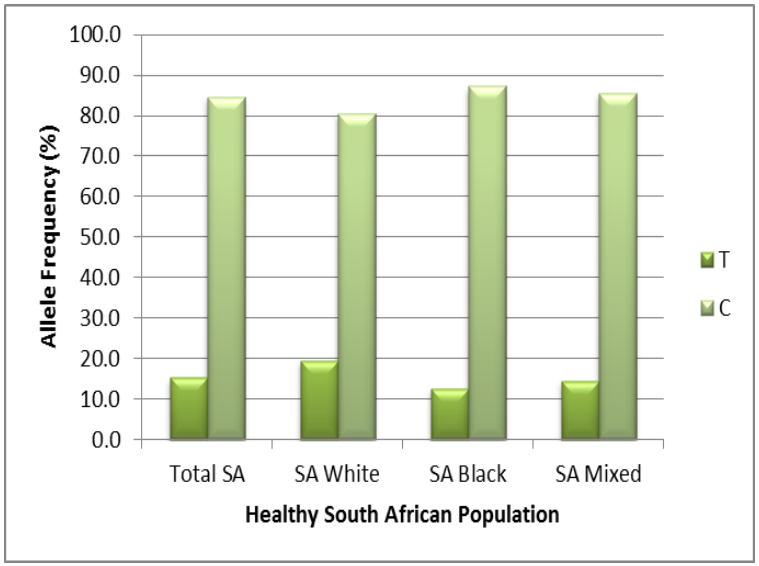


Figure 2.9: Percentage of genotype frequency CC, CT and TT for the SNP rs4327572 of the *CDH10* in (a) healthy and (b) autistic SA population

(a)



(b)

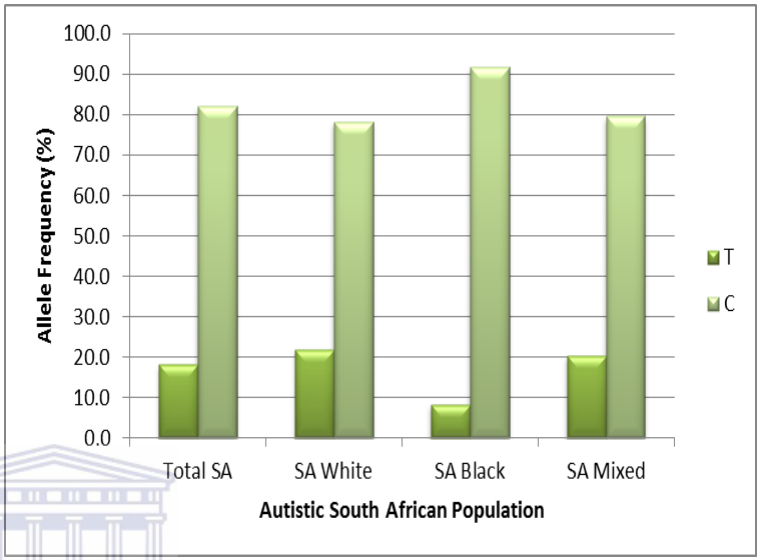


Figure 2.10: Percentage of allele frequency of C and T for the SNP rs4327572 of the *CDH10* gene in (a) healthy and (b) autistic SA population

2.5 Discussion

The present study was aimed at identifying the genotyping frequencies and allelic distribution of the SNPs rs4307059 and rs4327572 of *CDH10* gene in the different ethnic groups for both healthy and autistic population groups. Furthermore it also aimed to identify the association of these genotypes and alleles with autism.

2.5.1 SNP analysis rs4307059

The present study represents the first report on genetic association studies on the *CDH10* gene with the two SNPs rs4307059 and rs4327572 in a SA population. The distribution for the SNP rs4307059 has been observed globally for statistical purposes but not in association studies with disorders. To date, only two studies have found association between the SNP rs4307059 and the etiology of autism (Wang *et al.*, 2009, Prandini *et al.*, 2012). Figure 2.11 gives a representation of the T and C allele distribution of rs4307059 globally.

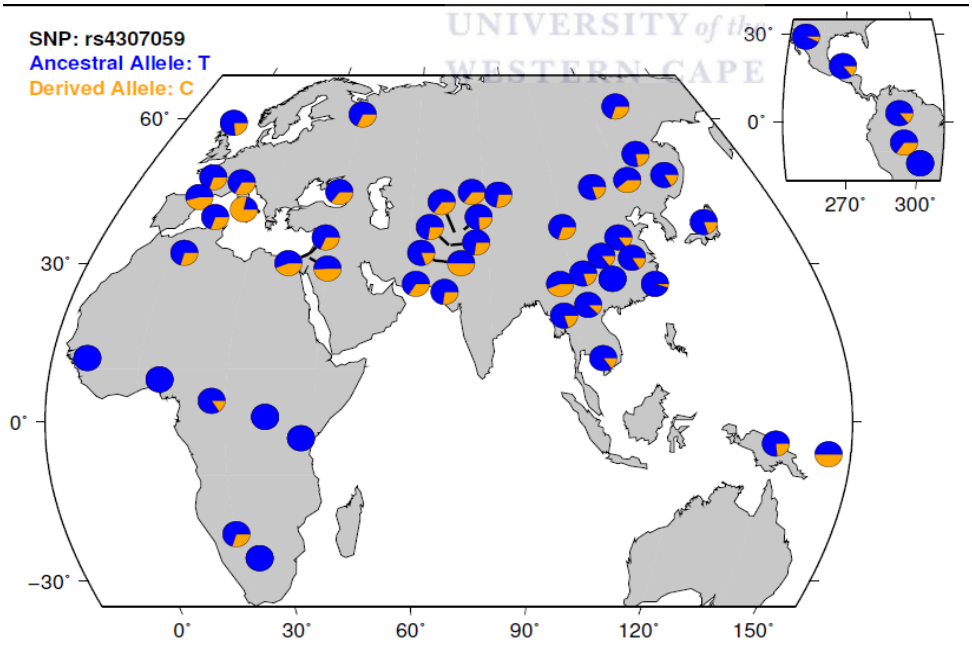


Figure 2.11: Global distribution and prevalence of the SNP rs4307059 (Applied Biosystems, 2012)

In the present study, the T allele is more prevalent in all ethnic groups, be it healthy or autistic. The healthy black group has the highest prevalence of the T allele at 96.5% with an only 3.5% C allele. The genotypes TT, CT and CC observed similar frequencies over all between the SA total groups in terms of ratios, with the TT genotypes more prevalent , CC least frequent and the CT with intermediate results in the healthy than the autistic as seen in table 2.2. A similar pattern of genotype ratios was observed within the different ethnic groups with comparable frequencies ratios between the healthy and autistic.

The *CDH10* gene has been investigated in neuroscience since it has such a major function with brain development especially in the fetal brain (Abrahams and Geschwind, 2008; Wang *et al.*, 2009). Expression of the gene has been observed with in-situ hybridization showing that it is localized in the brain of fetus mainly in the frontal cortex region previously associated with ASD similarly like the *CNTNAP2* which are "well-established to be involved with ASD" (Abrahams and Geschwind, 2008, Wang *et al.*, 2009). Two independent studies, reported a significant association for the T allele and the etiology for autism with a strong probability of P=3.4X10^{-8} and P= 0.017 in two different populations (Wang *et al.*, 2009, Prandini *et al.*, 2012). The degree of freedom for their study was established at 2 where the p-values are determined at 0.05% significance or cut off at 5.991. The ancestral allele T was found to be the risk variant in the European study with 0.65 frequency whereas the frequency of this allele for all African populations is 1.00 (Mcclellan and King, 2010). The results for their study indicated that the population was in Hardy Weinberg equilibrium. When compared to the European study, it was observed for this SNP rs4307059 to be significant with a P=3.4X10^{-8}, very strong signal association for common variants with susceptibility to ASD (Wang *et al.*, 2009). In an Italian study, done on families present of ASD, the T allele

demonstrated a strong signal for association with a probability of *P*=0.017 (Prandini *et al.*, 2012).

2.5.2 SNP analysis rs4327572

The SNP rs4327572 along with rs4307059 have been reported by Wang *et al.*, (2009) to have been significant in terms of the frequencies of the genotypes for the ancestral and minor alleles, T and C. The genotypes TT, CT and CC observed a similar trend with regards to the frequencies over all between the SA total groups as seen in table 2.3. The TT genotype observed as the least throughout the different populations in the healthy although the autistic observed higher incidence than compared to the healthy. Throughout this study, the T allele showed slightly higher prevalence in the autistic as compared to the respective healthy groups with the exception of the black autistic observing 8.3% T allele occurrence but 12.7% in the healthy. TT genotype was observed in in 3% of healthy black population but it was completely absent in autistic population.. This result could have been affected by the number of typed samples TT genotype might have different frequencies in a bigger sample size. Likewise the CC genotype had the opposite pattern at being observed at a slightly higher prevalence in the healthy groups with the exception of the black when compared to the autistic. Similarly the C allele had prevalence to some extent higher in the healthy groups with exception of the black group observing a spike of 91.7% in the autistic to of 87.3% in the healthy. The CT genotype had observed small difference between the SA healthy (26.8%) and autistic (27.8%) groups including the different ethnic groups (table 2.3). These results thus indicate that the T allele, the minor allele shows more prominence in these population groups and is thus associated with the susceptibility to ASD. When compared to the European study done by Wang *et al.*, (2009) from which this currents study was inspired. It illustrated a strong signal with a *P*-value of $P = 6.2X10^{-7}$(Wang *et al.*, 2009). The total group

size used for these groups should have rendered statistical significant results either positive or not for the allele, however, in this instance only 69.20% of the healthy group and 29.15% of the autistic group was successfully genotyped.

The results observed for these SNPs were conducted strictly in a SA population, and even though the numbers were low, these results observed could appear to be affected by the genetics that confer to the population diversity. The greater part of studies reveals that black populations harbour more genetic diversity than non-Black populations for mitochondrial DNA sequences, Y chromosome SNPs, Autosomal MicroSatelites as well as Autosomal SNPs, to name a few (Excoffier, 2002). These populations have greater levels of genetic diversity, extensive population substructure, and less linkage disequilibrium (LD) among loci than non-African populations and possess a number of genetic adaptations resulting from diverse climates, diets, and the exposure to infectious diseases (Campbell and Tishkoff, 2008). Other factors that may be used to account for this massive diversity could also include, "older onset of population demographic expansion in Africa, population size fluxes and periodic extinctions of populations outside Africa" (Excoffier, 2002). However all these are "plausible and are not mutually exclusive" to the black population diversity (Excoffier, 2002).

Lack of significance could also be accounted for by the fact that the SNPs location, that being in the intronic region of the *CDH10* gene may be selective active in certain populations. In general, intronic regions are described as "junk DNA" that disrupts earlier sustained genes (De Souza *et al.*, 1997). Proximity to the gene in question can lead to interference as De Souza *et al.*,(1997) discussed the exon theory of genes predicted proteins to be the results

from simple compact elements, and that the elements corresponding to exons that introns defines the boundaries of such elements (De Souza *et al.*,1997).

2.6 Conclusions

The objectives were to genotype these SNPS's, to determine the allele and genotype frequencies and association with autism. To a degree, these were met, the results for this study indicated that the SNP rs4307059 and rs4327572 were successfully genotyped but not significant by association in the susceptibility for ASD. The probability for the T allele for rs4307059 were *P*=0.8094 and the C allele for rs4327572 were *P*= 0.7409 in the SA healthy and autistic population. When compared to the other population study for the SNP rs4307059, the associations were not made clear for ASD to the SA population. This study thus indicates that the *CDH10* gene does not indicate relevance to the etiology for ASD in the SA population. In conclusion, the hypothesis that an association between the prevalence of these SNPs (rs4307059 and rs4327572) of the *CDH10* gene and autism status in the SA population existed was not found.

2.6.1 Limitation and shortcoming of the present study

Both Wang *et al.,* (2009) and Prandini *et al.,* (2012) have used massive cohort sizes. In the first paper Wang *et al.,* (2009) genotyped 1,204 autistic and 6,491 control subjects with strong association signals for these SNPs and 824 subjects in an Italian study by Prandini *et al.,* (2012). These studies could positively identify association statistically due to the positive hits estimated, benefits of having such a huge group. The present study had numbers far less and could account for the estimation of associations and prevalence observed in the SA population for both SNPs. The SNP rs4327572 did not have a strong association in the SA population and could be as a result to the fact that few numbers were positively genotyped for

the SNP. Both of these SNPs were obtained from the same company and used to genotype the same population group but yielded different positive hits. This could be relayed to the functionality of the SNP since prior to the genotyping, tests were done to assess the degree of functionality of each SNP.

2.6.2. Future studies

The SNPs rs4307059 and rs43275727 were selected from a study done in another population based on the strong signal for the association observed. The results obtained were to an extent similar, in future; a larger sample size could yield different results or yield more significant results. The SA population ethnic groups, if distinctly identified with respect to the mixed groups, the ancestry for those groups could shed light and the diversity that goes with it but also may add to the etiology of ASD in the SA populations.

Chapter 3

Abstract

The SA population is a diverse nation, with the African community comprising of up to 70% of the total population and have an interesting genetic history. The proteomics study aimed at isolating urinary proteins from African South-African male subjects was conducted. This included the purification of the urine protein samples as well as additional steps to remove high abundance proteins. The purified protein samples were subjected to 1D and 2D SDS-PAGE electrophoresis and protein identifications using Electrospray ionization (ESI) mass spectrometry. The proteomics study results revealed differential expression of proteins between healthy and autistic South Africans. While none of the proteins had direct association to Autism Spectrum Disorder (ASD), most had functions related to other diseases of which only few could be co-morbid disorders to ASD. This study also shows that more investigation with a larger scale of the study could reveal more distinct results for the selected group but also for different ethnic groups observed in the SA population. The results revealed 7 positively identified protein spots with a range of molecular weights and pI's. The proteins identified are Uromodulin, Vitelline membrane outer layer protein homologue, kinninogen-1, Alpha-1-Antitrypsin, Ig Kappa chain region C, and CD59 glycoprotein.

Proteomics of autism in a SA sample group

Table of Contents

3 Urinary proteomics in the discovery of biomarkers for autism

3.1 Proteomic studies on urine

It has been demonstrated that proteomic studies can be done using urine as source due its properties (Wu *et al.*, 2010a). The term "proteomics" and its technologies have been explained in the previous section (Chapter 1). Biomarkers are biological indicators that are "used as an index of the intensity of a disease or other physiological state in the organism" (Majkić-Singh, 2011).

The completion of the Human Genome project has shifted the focus towards the understanding of function and structure of the human genome through the identification of protein structure, protein-protein interaction and their differential expression (Bae *et al.*, 2003). In the post genomic era, proteomics has been considered increasingly important (Pandey and Mann, 2000). It plays a significant role in clinical useful applications through the study of changes in protein expression levels or post translational modifications (Vidal *et al.*, 2005).

Proteomics is a promising approach for the detection and identification of novel biomarkers using various biological fluids such as urine, plasma serum and saliva. Out of those, urine is considered to be the most attractive proteomic sample due to several advantages. Urine is a filtrate of blood and therefore contains a similar protein composition to that of blood. This makes it more useful for experimentation of protein profiles and it may hold information about the changes that may take place within the body (Barratt and Topham, 2007). These protein changes highlight the role of peptides as potential biomarkers. It is noninvasive and easily obtained. Urinary proteins and peptides are reported to be stable and can undergo

insignificant proteolysis within several hours after the collections (Theodorescu and Mischak, 2007). Thus it allows for collection of samples without inflicting bodily harm to the subject, in large quantities and or repeatedly. The stability of these proteins favors urine as use as an interesting source for biomarkers as it can be stored for several years without significant alteration to the proteome (Roelofsen *et al.*, 2007, Wu *et al.*, 2010a). Many studies on urinary proteins have been reported since 1970's.

The urine proteome however differs from one individual to another and is composed of extracellular proteins, plasma membrane proteins and lysosomal proteins (Adachi *et al.*, 2006). Many proteins have been identified using different proteomic techniques such as 1D and 2D gel electrophoresis, mass spectrometry (MS), Liquid chromatography-MS/MS and many others. Pieper (2004), Pitsikun (2004), Castagna (2005), Sun (2005), and Wang (2006) and their collaborators have identified 150, 295, 383, 226, 225 unique proteins (gene products) respectively. Furthermore, nearly 1000 protein spots have not been identified when resolved on 2D gel (Adachi *et al.*, 2006).

Through the advances in proteomic technology and mass spectrometry, protein discovery in urine have increased in its identifications surpassing the numbers previously reported. Using high accuracy mass spectrometry, Kentsis *et al.*,(2009) identified up to 2362 proteins with more than 1000 not previously reported in other studies. The proteins annotated were found to be associated with 27 common and more than 500 rare human diseases (Kentsis *et al.*, 2009). Another interesting factor observed in their study was that most of these identified, proteins appeared to correspond to expression from the urogenital tract and most likely formed in the kidneys or bladder (Kentsis *et al.*, 2009). Furthermore of those identified, 336

proteins were found to be filtered through the glomerular filter suggesting that those proteins were from a different organ or area of origin other than that of the kidney organ. This finding proposes the use for urine proteomics for studies of human organ and disease etiology (Kentsis *et al.,* 2009). The possibility of identifying proteins from other organs could produce interesting and significant data in the study of autism without the risk of harm or discomfort to sufferers. A study conducted in 2012 reports on anti-neuronal antibodies extracted from venous blood. This study aimed at measuring these antibodies as signs for autoimmunity in the brain and the relationship to the severity of autism as classed using the Childhood Autism Rating Scale. The presences of these proteins were investigated and were found to be significantly correlated to the severity of autism in the study group (Mostafa and Al-Ayadhi, 2012).

There are two ways for characterizing urine, a gel-based and gel-free method because urine is easily accessible by noninvasive methods and contains bodily peptides, proteins and amino acids related to renal systems (Dihazi and Müller, 2007). By using the gel-based method (1D SDS PAGE), Adachi *et al.*,(2006) showed that the urine proteome contains more than 1500 proteins including a large proportion of membrane proteins. These proteins were digested from the SDS PAGE gels and subjected to HPCL and MS with positive identification.

With a gel based approach in proteomics coupled with MS technology it can and has thus been used to identify biomarkers or biological indicators from this source. These techniques offer the ability to examine the expression of up to 10,000 protein spots which can be studied (Thongboonkerd *et al.*, 2002). Identification of diagnostic biomarkers for autism is very

important since early detection and/or rapid diagnostics for this disorder is non-existent currently. The aim of this study was to investigate this possibility.

3.1.1. Proteomic studies and autism

In recent years proteomics approaches have been used in aid of understanding the etiology for autism. Since autism has been previously demonstrated to have associations with metabolic aberrations, immunological functional and gastrointestinal disturbances though the exact mechanical significance is unknown (Chugani *et al.*, 1999, Kidd, 2002, Ashwood *et al.*, 2003, Yap *et al.*, 2010). Using a non-gel based proteomic approach on blood serum, Corbett *et al.*, (2006) identified differentially expressed apolipoproteins and complement proteins between autistic and control subjects. Furthermore Yap *et al.*, (2010) revealed novel findings when subjecting autistic urine samples to nuclear magnetic resonance (NMR) spectroscopy. This study reported significant metabolic differences in the urine compositions between healthy and autistic children. Emond *et al.*, (2013) also reported that there had been a significant difference in urinary metabolic profiles between autistic and healthy subjects.

3.2.Hypothesis

It is hypothesized in this study that there are differences in the proteome profile of urine of autistic children when compared to healthy children.

3.2.1 Aims

- To identify the differentially expressed proteins among autistic and healthy subjects.

3.2.2. Objectives

- To collect urine samples from both healthy and autistic children
- To extract proteins from urine samples
- To generate reproducible proteome profiles using 1D and 2D SDS-PAGE electrophoresis
- To identify differentially expressed protein using mass spectrometry

3.3. Materials and methods

Materials

1.0 mm Spacer glass plates	Bio-Rad, California, USA
2-Mercaptoethanol	Sigma Aldrich, Missouri, USA
(3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate) CHAPS	Sigma Aldrich, Missouri, USA
Short glass plates	Bio-Rad California, USA
Acetone	Kimix, Cape Town, SA
Acetonitrile	Kimix, Cape Town, SA
Acrylamide (30% Bis acrylamide)	Bio-Rad California, USA
Agarose	Bio-Rad California, USA
Agilent LC-ESI-QTOF MS	Agilent technologies, California, USA
Ammonium bicarbonate	Bio-Rad California, USA
Ammonium persulphate	Bio-Rad California, USA
Ampholytes	Bio-Rad California, USA
Aurum™ Affigel® Blue Mini kit/column	Bio-Rad California, USA
Bio-Rad Bradford assay dye	Bio-Rad California, USA
Bovine serum albumin	Sigma Aldrich, Missouri, USA
Bromophenol blue	Sigma Aldrich, Missouri, USA

Coomassie brilliant blue R-250 (CBB R-250)	Bio-Rad California, USA
DTT (Dithiothreitol)	Fermentas, Thermo scientific, Massachusetts, USA
Ettan™ IPGphorII™ IEF	Amersham Biosciences, GE Healthcare UK
Fermentas unstained Protein ladder	Fermentas, Thermo scientific, Massachusetts, USA
Glacial acetic acid	Merck, Darmstadt, Germany
Glycerol	Merck, Darmstadt, Germany
Glycine	Bio-Rad California, USA
Hydrochloric acid	Merck, Darmstadt, Germany
Iodoacetamide	Sigma Aldrich, Missouri, USA
IPG strips	Bio-Rad California, USA
Lyophilizer Labconco	Labconco, Kansas, USA
(N,N,N´,N´-tetramethylethylenediamine) TEMED	Sigma Aldrich, Missouri, USA
PDQuest software basic	Bio-Rad California, USA
Pharos FX™ Plus Imager, (Quantity One imaging Software 4.6.9)	Bio-Rad California, USA
Phenylmethanesulfonylfluoride (PMSF)	Sigma Aldrich, Missouri, USA
Plus one dry strip covers fluid	Amersham Biosciences, GE

	Healthcare, UK
Propan-2-ol	Merck, Darmstadt, Germany
Sodium dodecyl sulfate	Bio-Rad California, USA
Sorvall RC5C centrifuge	Thermo Scientific, Massachusetts, USA
Speed vacuum SC100 Thermo Savant	Thermo Scientific, Massachusetts, USA
Thiourea	Sigma Aldrich, Missouri, USA
Tris-HCl base	Bio-Rad, California, USA
Trypsin solution	Promega, Wisconsin, USA
Urea	Sigma Aldrich, Missouri, USA
Whatman™ filter paper	GE Healthcare, UK

Buffers and solutions

- 1 X SDS Buffer: 1 part of 10X SDS buffer diluted with 9 parts dH_2O
- 1% Sealing agarose gel: 1 g agarose; 0.04% bromophenol blue stock ; dissolved in 100 ml 1X SDS buffer solution
- 10% Ammonium persulphate: 0.1 g APS dissolved in 1 ml dH_2O; prepared fresh every two weeks
- 10% Sodium dodecyl sulphate: 10 g SDS dissolved in 100 ml dH_2O
- 10X SDS Buffer: 1.92 M glycine (144 g); 0.25M Tris-HCl (30.2 g); 0.1% SDS (10 g); dH_2O to final volume 1L

- 12% Separating gel: 30% acrylamide (37.5:1) (4 ml); 10% sodium dodecyl sulphate 91.25 ml); 3 M Tris-HCl pH 8.8 (1.25 ml); 10% ammonium persulphate (freshly prepared) (0.05 ml); 0.02 ml TEMED and 3.43 ml dH_2O to final volume of 10ml.
- 2 X SDS reducing buffer pH 6.8: 0.2 M Tris-HCl pH 6.8 (1.21 g); 40% glycerol (20 ml); 10% SDS (5 g); 0.02% bromophenol blue (0.01 g); dH_2O; At time of use, 10% 2-mercaptoethanol (0.15 ml per 1.5 ml aliquot)
- 25mM Ammonium bicarbonate/50% Acetonitrile: 50mM ammonium bicarbonate stock plus 1part volume of 100% acetonitrile
- 5% Stacking gel: 30% acrylamide (37.5:1) (0.8 ml); 10% sodium dodecyl sulphate 0.625 ml); 1M Tris-HCl pH 6.8 (0.625 ml); 10% ammonium persulphate (0.025 ml); 20 µl TEMED and 2.905 ml dH_2O to final volume of 5 ml
- 50% Acetonitrile: 5 ml acetonitrile; 5 ml milliQ H_2O
- 50mM Ammonium bicarbonate: 0.08 g ammonium bicarbonate; 20 ml milliQ H_2O (Prepared fresh before use)

- 70% Acetonitrile: 1400 µl acetonitrile; 600 µl milliQ water
- Bio-Rad Bradford protein assay dye: 1 part of protein assay diluted with 4 parts of dH_2O
- Bovine serum albumin: 0.005 g of BSA dissolved in 1ml of urea solubilizing buffer
- Bromophenol blue stock solution: 1% (100 mg) bromophenol blue, 0.05 M Tris base (60 mg), dH_2O to 10 ml
- Coomassie staining solution I: 10% glacial acetic acid; 0.025% (w/v) CBB R-250; 25% propanol; double distilled H_2O to final volume 1L
- Coomassie staining solution II: 10% glacial acetic acid; 0.003125% (w/v) CBB R-250; 10 % propanol; double distilled H_2O to final volume 1L

- Coomassie staining solution III: 10% glacial acetic acid; 0.003125% (w/v) CBB R-250; double distilled H_2O to final volume 1L
- Destaining solution: 10% glacial acetic acid; 1% glycerol; double distilled H_2O to final volume 1L
- Low application buffer : 20 mM Tris-HCl pH 8.3; dH_2O
- Urea solubilizing buffer: 7 M urea, 2 M thiourea, 4% CHAPS dissolved in 100 ml distilled H_2O

Methods

3.3.1 Ethical clearance

This study has been approved by the UWC ethics committee (SR: 5/09/32) and permission was received from the Western Cape Education Department (*Appendix I*). Furthermore, permission was obtained from the school principals to conduct the study at the respective schools. Informed consent was obtained from the parents of the respective children participating in the study (Appendix *I*).

3.3.2 Study design and population

The project was designed as a case-control comparative study involving two groups of children diagnosed with autism (experimental group) and without any history of autism (control group). The inclusion criteria for the experimental group were to be children/pupil's at school with the diagnosis of autism made by a professional doctor and/or psychologist. The study was gender specific to males of African ancestry. Children/pupil's had to be South Africans aged between 6 and 14 years and residing in the Western Cape region. The Informed written consent from the parent as well as the school principal and teachers were taken for this study. The criteria of selection of the healthy group matched that of the experimental

group in terms of the age requirement, gender specificity, and race. Informed consent from parents and the school principal were also taken. The healthy group had to display the absence of any mental disabilities and any history of mental illnesses for the comparative study.

3.3.3 Sample collection

All samples were collected in Cape Town from the Western Cape region. For the experimental group, participants were recruited from Nolunthando School situated in Khayelitsha. The healthy group participants were recruited from the Lantana Primary School in Mitchell's plain. A teacher assisted the male subjects during the process of the second urination for the day and assured all samples were midstream collections. The urine was collected into a sterile 100 ml glass bottles that contained a proteanases inhibitor Phenylmethanesulfonylfluoride (PMSF) at a 1 mM concentration to a urine volume of 60 ml. All samples were obtained on the school premises. The bottles were kept on ice until they reached the laboratory and were stored at -80^{o}C.

3.3.4 Sample preparation

The frozen urine samples were thawed and were clarified by centrifugation at 3,000 x *g* for 10 minutes and then filtered through a 0.45 µM syringe filter. Proteins were extracted from the urine filtrate utilizing the acetone method (Khan and Packer, 2006). A urine-to-acetone ratio of 1:4 was used and the samples were placed at -20°C for an incubation period of at least 12 hours. Subsequent to incubation, the samples were centrifuged for 30 minutes using a Sorvall RC5C centrifuge at 16,887 x *g* and the precipitated pellet was collected. The pellet was then subjected to a post precipitation 80% acetone wash and allowed to air dry. The cleaned pellet was suspended in a solubilizing buffer containing 7 M urea, 2 M thiourea and

4% 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) and placed on a vortex overnight at 4^{o}C. The resulting solubilized samples were then centrifuged at 20,000 x *g* for 10 minutes and the supernatant recovered. A 200-fold dialysis of the recovered supernatant was done overnight at 4^{o}C against 18 Ωcm deionized water. The protein sample was recovered and frozen at -80^{o}C for 20 minutes followed by freeze drying using a Labconco lyophilizer. The samples were collected the following day, labeled and stored in lyophilized form at -20^{o}C.

3.3.5 Depletion of high abundance albumin proteins

In this section half of each lyophilized sample for both healthy and autistic groups were subjected to an albumin depletion step using the Aurum™ Affi-gel® Blue Mini Kit according to the manufacturer's instruction manual. The beads were drained of residual buffer and washed twice with 1 ml of Low Application Buffer. The lyophilized samples were suspended in the same Low Application Buffer and transferred onto the Affi-gel® Blue Mini Kit columns and allowed to be filtered via gravitational force into sterile 1.5 ml Eppendorf tubes. The filtrate collected from the column (albumin free) were re-dialyzed overnight at 4^{o}C against 18Ωcm deionized water, lyophilized and stored as described in section 3.3.4.

3.3.6 Protein quantification

All the lyophilized protein samples (healthy and autistic samples both depleted and un-depleted of albumin) were re-suspended in Solubilizing Buffer. These were quantified using the Bradford reagent using bovine serum albumin (BSA) as a standard. Using the Bradford (1976) method, a standard curve was generated where BSA was used with the following concentrations, 0.5µg/µl, 1µg/µl, 2 µg/µl, 3 µg/µl , 4 µg/µl and 5 µg/µl. To a cuvette 5 µl of

each standard was added to 995 µl of Bradford reagent and incubated for 15 minutes at room temperature and the absorbance read on a spectrophotometer at 595 nm.

For the re-suspended protein samples, two fold dilutions were prepared of which 5µl of each of the series were mixed with 995 µl of Bradford reagent. The samples were incubated for 15 minutes and the absorbance was read at 595 nm using a spectrophotometer. All samples were quantified using the Bradford method (Bradford, 1979) and then resolved on a 12% 1D SDS-PAGE gel.

3.3.7 One dimensional (1D) SDS-PAGE electrophoresis analysis

With a concentration determined the re-suspended protein samples were subjected to 1D analysis. SDS-PAGE gels were prepared with a 12% separating gel and a 5% stacking gel set to cast in a 1.0 mm glass plates (Laemmli, 1970). With calculating the final weight of protein in the well to 10 µg, 15 µg and 20 µg, samples were diluted with an equal volume of 2X SDS reducing buffer, boiled for 5 minutes in a water-bath at 95°C. The samples were subjected to a pulse centrifugation prior to loading adjacent to an unstained Fermentas protein ladder. The gels were initially electrophoresed at 50 V for 10 minutes followed by 100 V until the bromophenol blue dye reached the bottom of the gel. The gels were then stained for visualization and imaging.

3.3.8 Visualization and imaging

The Coomassie Brilliant blue R-250 staining and a destaining solution was used for visualizing protein bands and spots. The SDS-PAGE gels were exposed to three steps staining and destaining in the following manner:

Coomassie Brilliant Blue Stain I: The SDS-PAGE gel was submerged into 250 ml of this stain, heated for 30 seconds in a microwave and left at room temperature on a shaker for 30 minutes. Following this, the Coomassie stain I was decanted.

Coomassie Brilliant Blue stain II: Subsequent to the first stain I step, the SDS-PAGE gel was then submerged into 250 ml of this stain, heated for 30 seconds in a microwave, and left at room temperature on a shaker for 30 minutes. The second stain was then decanted in preparation for the final staining step 3.

Coomassie stain III: Following stain II, the SDS-PAGE gel was submerged in 250 ml of stain III, heated for 30 seconds in the microwave, and left at room temperature for 30 minutes on a shaker. The stain was then decanted, and the gels were destained in 250 ml of destaining solution with shaking at room temperature until distinct protein bands were visualized.

Imaging: The SDS-PAGE-gels were viewed on Pharos FX™ Plus Molecular Imager, utilizing Quantity One® 1D analysis software, and followed by analysis using the PDQuest software basic 8.0.1 Gel Imager.

3.3.9 Two dimensional (2D) SDS-PAGE electrophoresis analyses

There are four main steps in 2D analyses; these include in-gel rehydration, iso-electrical focusing (first dimension), equilibration and then the second dimension run which is the molecular weight separation of proteins (Thongboonkerd *et al.*, 2002).

In-gel rehydration: Protein samples of 50 and 100 µg were used for 2D gel optimization. The IPG strips can only harbor 1.25 µl of carrier Ampholytes and 150 µg of protein and total 125 µl of complete mixture. To supplement the balance of total mixture, solubilizing buffer was added containing 8% dithiothreitol (DTT). The samples were subjected to a pulse centrifugation and loaded onto a level re-swelling tray. The samples were soaked into a 7 cm long IPG strip with pH 4-7 overnight at room temperature covered in Amersham mineral oil.

Iso-electrical Focusing (IEF): Following overnight re-swelling in the tray, the IPG strips were rinsed with distilled water and placed on an Ettan™ IPGphorII™ IEF machine with the respective ends (positive and negative) aligned and covered with mineral oil. The IEF run was set to a three step program, step 1: 250 V for 10 minutes, step 2: 4000 V for 1 hour and step 3 4000 V until it reached 12 000 V/hours at 20°C.

Equilibration: This step involved the equilibration of the focused IPG strips in Equilibration Buffer I and II. Following IEF, the IPG strips were removed and washed gently by pouring distilled water over them and transferred into a tray. Each strip was then incubated in Equilibration Buffer I for 10 minutes on a shaker at room temperature after which Equilibration Buffer I was decanted and Equilibration Buffer II added following same procedure as Equilibration Buffer I (Ngara and Ndimba, 2011). The IPG strips were then washed gently using 1 X SDS buffer and prepared for the final step in 2D SDS PAGE.

Second Dimension electrophoresis: In the second dimension the IPG strips were subjected to separation based on molecular weight. The strips were placed on top of a 12% separating SDS-PAGE gel, which were aligned such that the marker was placed at the negative end of the strip. A square cut (4 mm X 4 mm) piece of Whatman™ filter paper was soaked in Fermentas Unstained Protein ladder and placed adjacent to the strip. To seal off the strip, a 1% agarose solution was loaded, sealing the strip in place. The strip was then set to run on the gel for 100 V until the bromophenol blue dye reached the bottom of the gel. Once completed, the SDS-PAGE gels were recovered and stained for visualization and imaging as described in section 3.3.7.

3.3.10 PDQuest analysis

The PDQuest basic 8.0.1 2D analysis software was used to conduct 2D SDS PAGE analysis between the two samples sets used in this study. This software was used since it allows one to determine differential expression, quantitative differences and statistical significance between respective groups. In this analysis four parameters were used to compare the healthy and autistic samples. A quantitative parameter was used to determine the fold change in expression with and upper limit of 1.5 fold and lower limit of 0.5 fold. A statistical parameter, the student's t-test was used with a 90% significance cut-off. Based on these parameters only a few spots were identified in this program and used for further analysis. The last two were Boolean parameters which allows for the analysis of spots between the group by combining and comparing the groups. Those selected were the union of sets and the other the intersection between all the sets.

3.3.11 Protein excising and identifications

3.3.11.1 In-gel digestion

Spots of interest were selected on the bases of PDQuest software basic (8.0.1) analysis. The digestion of the spots was done sequentially in 5 steps: washing, destaining, dehydration, desiccation and digestion.

Washing: Spots were manually excised by use of a sterile spot-cutter and transferred to sterile micro-centrifuge low binding Eppendorf tubes. To the gel pieces, 200 µl of distilled water was added and vortexed slowly for 5 minutes. The water was removed and the step repeated thrice. Destaining: To the gel pieces 50% (v/v) 50 mM ammonium bicarbonate and 50% (v/v) acetonitrile solvent were added and vortexed for 5 minutes and repeated until the gel plugs were colorless to clear. The solution was discarded and clear gel pieces subjected to dehydration. Dehydration: To the clear gel pieces 100 µl of 100% (v/v) acetonitrile solvent was added and incubated by vortexing slowly for 10 minutes and solution was discarded. Desiccation: The gel pieces were dried with the use of a Speed Vacuum SC100. Digestion: The washed and dried gel pieces were subjected to tryptic digestion. The peptides were digested with 15 µl of 10 ng/ml trypsin solution and incubated at 37^{o}C for 12 (minimum) to 16 hours (maximum). Samples were recovered and the peptides extracted. Extraction: To the tryptic protein plugs, 20 µl of 70% acetonitrile was added and incubated on low setting vortex for 30 minutes and the supernatant collected. This step was repeated twice and the supernatant collected and subjected to desiccation.

3.3.11.2 Identifications using Agilent LC-ESI QTOF MS

Protein spots were digested with trypsin and identified by use of an Agilent 6530 LC-ESI-QTOF Mass Spectrophotometer. The identification of selected protein spots were carried out by Dr. Zac Macdonald in the Molecular and Cell Biology Laboratory at the University of Cape Town. The MS results were viewed using the Agilent Spectra Mill software version 4.0, and matched with sequences in the SwissProt database. The process of identification included the drying of the digested spots followed by resolubilization in a specific loading buffer and injected onto the LC/MS system. For a detailed description of the method and parameters for protein identification refer to *Appendix II.*

3.4 Results

The acetone precipitation and protein extraction method was utilised to obtain urinary proteins from both the healthy and autistic South African groups. This included the check for protein concentration per gel, additional treatments and purification steps. Purification steps had to be implemented to reduce the high ion levels present since urine is a complex mixture of solutes, ions and other materials. This was done by dialyzing the samples against 18Ωcm distilled water at 4^{o}C overnight and followed with lyophilizing to insure a concentrated sample. Protein samples were quantified using the Bio-Rad Bradford assay and results for all samples ranged between 0.31 and 4.45 µg/µl. Three concentrations were used for gel loading, 10 µg, 15 µg and 20 µg to establish a standard concentration for the 1D SDS-PAGE profile. The 15 µg final protein content gave the best resolution for the 1D SDS-PAGE gel profile. Technical triplicates were done for each sample to ensure profiles were consistent (data not shown).

The use of an albumin depletion kit was considered to be essential to meet our aims in this project. Albumin proteins are known to mask smaller and less abundant proteins and thus their removal was incorporated. This meant that part of this experiment was to observe the difference in the protein profiles in urine samples with and without albumin. Figures 3.1 and 3.2 show the urinary protein profiles of six healthy and six autistic subject pre and post albumin treatment of samples.

3.4.1 1D SDS-PAGE analysis

A comparison was assessed between the healthy and autistic groups, to establish any differences between the groups, both un-depleted and depleted groups.

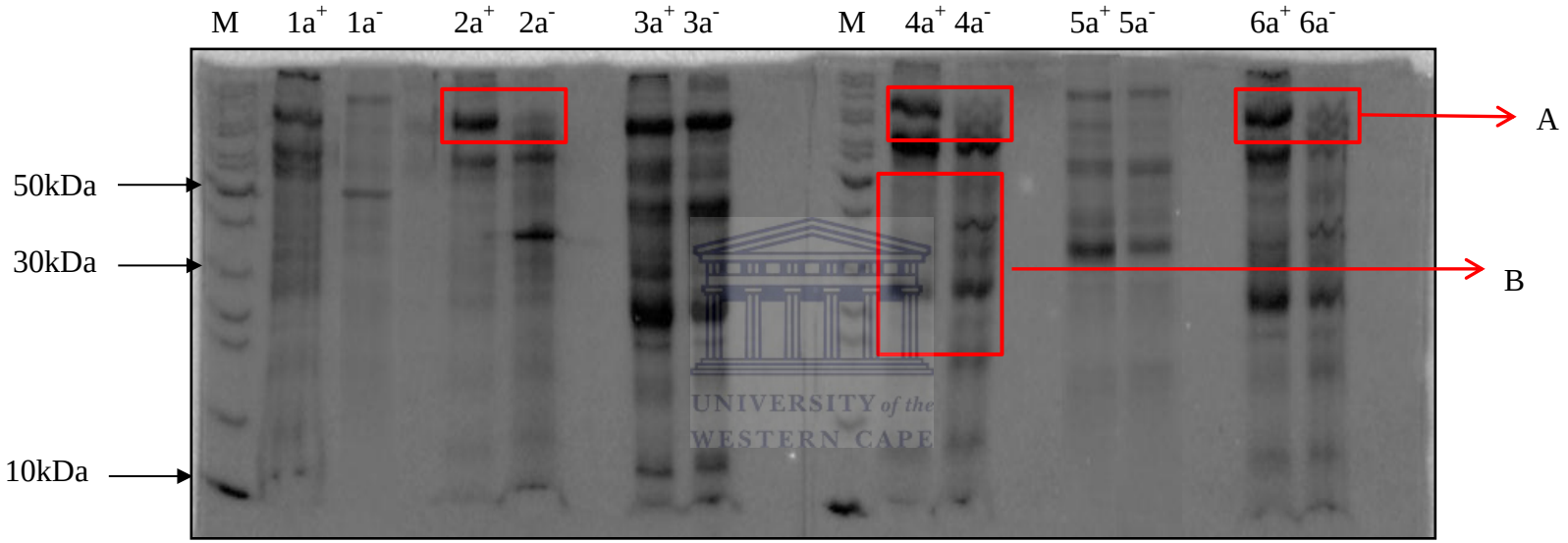


Figure 3.1: Healthy un-depleted and depleted of albumin urine protein 1D SDS-PAGE profiles. Subject's sample 1 to 6 was loaded in the order observed with the un-depleted sample first followed by the depleted sample for each subject. The un-depleted of albumin samples are indicated by the a^+ and depleted of albumin indicated by a^- . This was done for all the samples through 6. M represents the protein marker lane (kDa). Red boxes depict the difference between the same sample before and after Albumin depletion. Legend (A) directs to bands which were to indicate the presence or absence of albumin. In the pre albumin treatment distinct bands were visible which in the corresponding post albumin treated lane were not visible to absent. Legend B directs to bands for subject four, pre and post albumin treatment profiles. In lane a^+ for subject four shows ample protein bands marked in the red box. In lane a^- for the same subject post albumin treatment, reveals more bands or enhanced intensities for those proteins in that part of the gel.

Figure 3.1 depicts the protein profiles of six normal subjects. This figure shows protein profiles pre and post albumin depletion. In the figure, sample (1) was loaded (pre albumin depletion) and post albumin depletion. Similarly, in next lanes following samples 2 to 6 were loaded in the same order. In the lanes (subject's 2, 4 and 6) marked by the red square blocks, differences were observed in band intensities

Figure 3.2 represents the proteins profiles for the experimental group. All samples were loaded for each subject the pre and post albumin treatment for the sample in the same manner for the healthy group. The red boxes (legend C) show the results observed from the use of albumin removal treatment. The bands appear to be more abundant or present in the depleted lanes than the un-depleted lanes of the same sample.

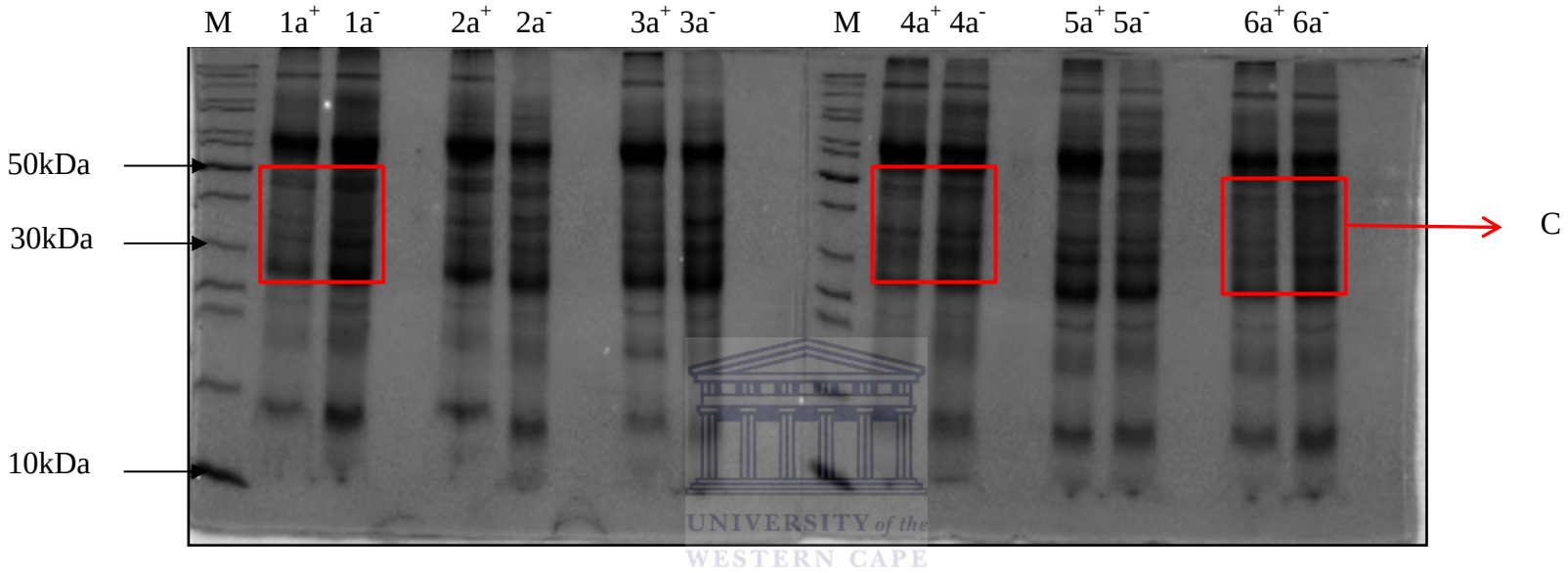


Figure 3.2: Autistic un-depleted and depleted of albumin urine protein 1D SDS-PAGE gel profiles. As seen subject samples 1 to 6 were loaded with the albumin containing and albumin free samples for each subject. The a^+ here represents the albumin containing sample and the a^- represents the albumin free sample. This was done for all the subjects through 6. M represents the protein marker lane (kDa). Differences observed shown in there red boxes (C).

3.4.2 2D SDS-PAGE analysis

Following electrophoresis of samples by 1D SDS-PAGE, the samples were then analyzed on 2D SDS-PAGE gels to determine their respective quality and protein profiles. The samples were resolved on 2D SDS-PAGE gels using 7 cm length of IPG strips of pH4-7. The strips were electrophoresed in the second dimension on 12% SDS-PAGE gels at 100 V for 90 minutes. At first, the un-depleted samples were subjected to 2D focusing, but results showed poor resolution and proteins did not migrate on the strip possibly due to the high salt/ion concentration. The protein samples for both healthy and autistic groups in both forms, un-depleted and depleted of albumin were then subjected to the desalting method that included the dialysis against 18Ωcm deionized water followed by lyophilization. The figures 3.3, 3.4, 3.5 and 3.6 show the protein profiles obtained after all the preparations, the albumin removal and desalting steps. Technical triplicates were made for each sample to provide consistent data.

Figure 3.3 displays the protein profiles for all the healthy South African male samples without the removing of the albumin. The profiles positively displayed proteins; in some sections of the gels some clusters were observed. Other observations made were the presence of large protein spots. This made it difficult to determine whether or not one or more than one protein could be present for that size and/or pI.

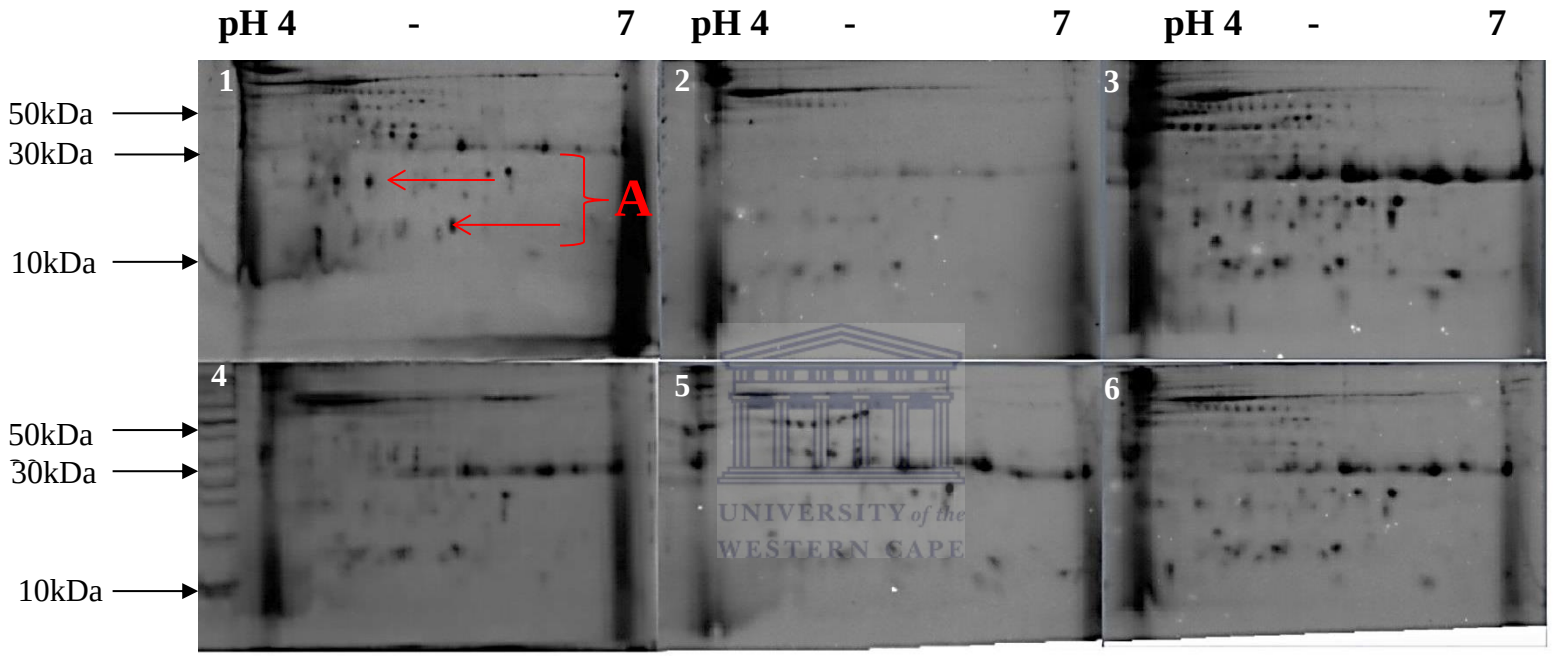


Figure 3.3: Healthy albumin un-depleted urine protein profiles resolved on 12% 2D SDS-PAGE gels pH 4-7. Protein marker bands are demarcated on the left in kDa. The arrows in sample 1 indicate protein spots which are observed differently in the post albumin step sample (as in figure 3.4). The six samples resolved on 12% PAGE gels, parameters mentioned previously at a 100 µg final concentration, were arranged in the order in which they were collected. From left to right in the top row are the samples 1 to 3 and bottom row samples 4 to 6. This was done similarly for all other proceeding samples.

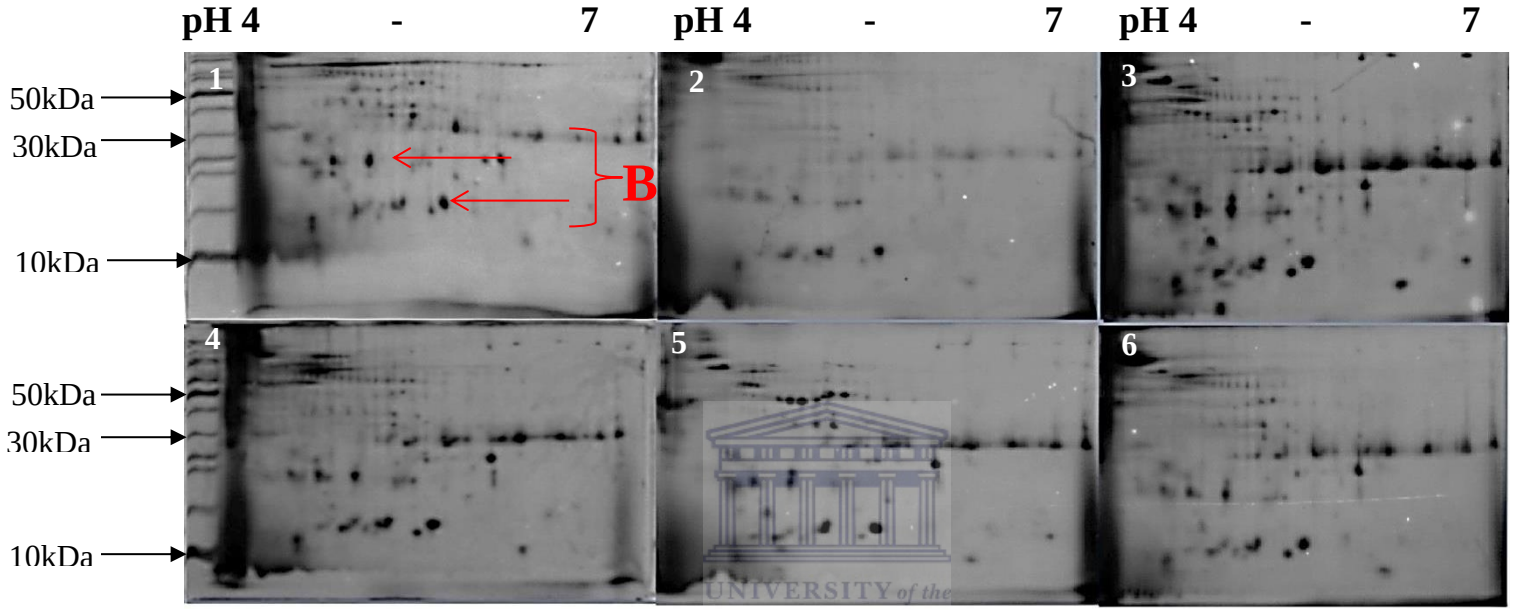


Figure 3.4: Albumin depleted urine protein profiles among healthy subjects resolved on 12% 2D SDS-PAGE gels pH4-7. Protein marker bands are noted on the left in kDa. As mentioned all samples were electrophoresed at a 100 µg final protein content onto a 12% SDS PAGE gel. From left to right in the top row are the samples 1 to 3 and bottom row samples 4 to 6. The arrows for point B indicates to protein spots which have observed a difference in resolution when compared to the un-depleted counterpart in figure 3.3 for sample 1.

Figure 3.4 displays the protein profiles for all the healthy South African male subject samples which have undergone the removal of the abundant protein albumin. The six samples were electrophoresed under the exact same conditions with regard to gel percentage, protein concentrations, electrophoresis parameters and staining regime. The expected outcome for these samples was to observe a difference in protein spots with regards to number of spots and resolution of smaller spots. The red arrow in figure 3.3 and 3.4 shows a portion of one sample that indicates the difference between an albumin removed sample and an un-depleted sample. The label A (figure 3.5) shows a few protein spots which were more intensified shown by point B (figure 3.6). The dark spot predicted to be albumin which has a molecular weight of 67 kDa were expected to appear in the top part of the gels. Between the groups, higher molecular weight protein spots were observed in the un-depleted groups in this area of the gels, when compared to their counterpart in the albumin depleted groups, some of these appear to have either a reduced amount of protein in that area or nothing at all.

Figure 3.5 displays the protein profiles for all the autistic African male samples which have undergone the removal of the abundant proteins albumin. The six samples were processed with the exact same parameters as the healthy un-depleted and depleted of albumin sample. The arrangement of samples was displayed in order of collection as well with samples 1 to 3 in the top row left to right and 4 to 6 in the bottom. The treated samples for this experimental group are shown in figure 3.6. To the naked eye, when observing for differences in profiles, some were seen in particular a very distinct presence of more spots after albumin removal treatment. In figure 3.6, sample 3 appeared to have more spots (point D) than seen in the un-depleted sample 3 (point C) in figure 3.5. From these gels and their comparison, it could be stated that the albumin removal may have removed the albumin and associated proteins and may have enhanced or unmasked the resolution for lower molecular weight proteins.

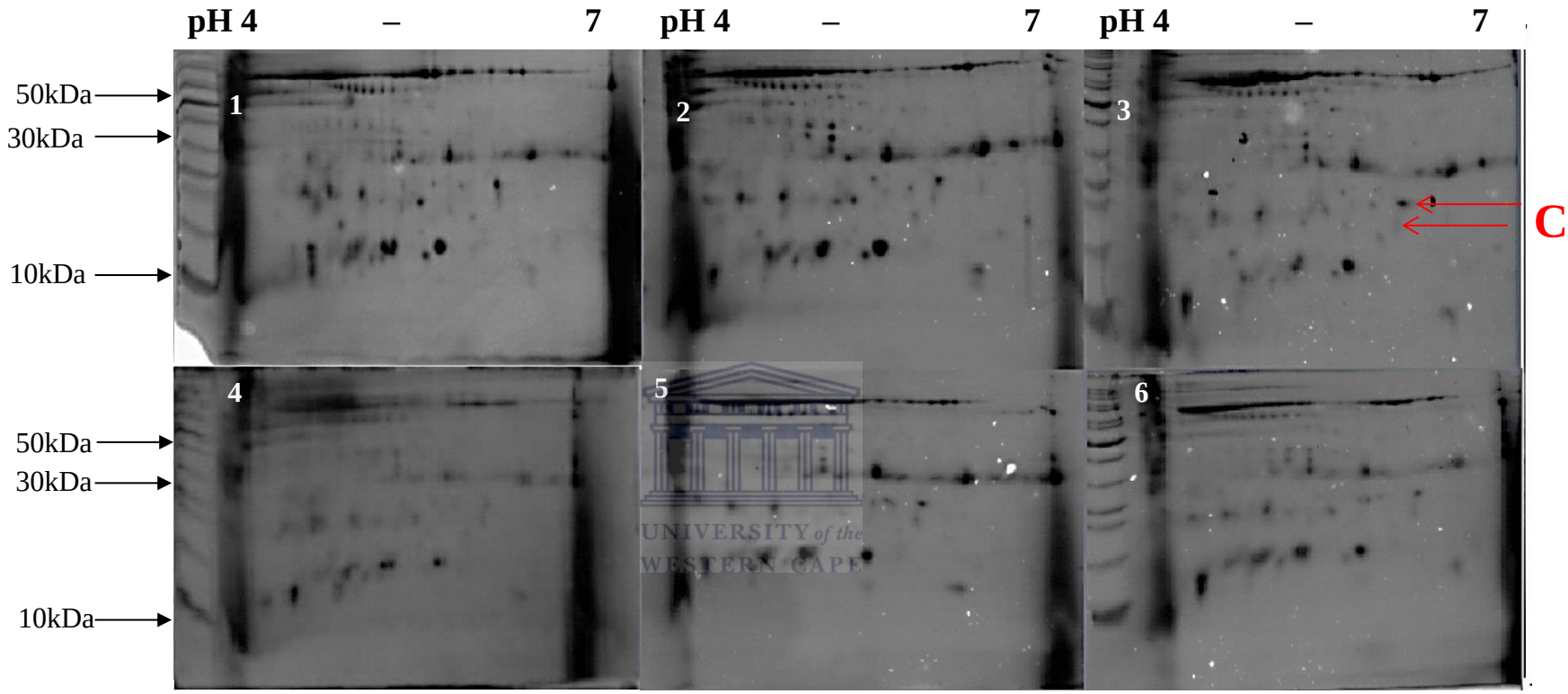


Figure 3.5: Autistic albumin un-depleted urine protein profiles resolved on 12% 2D SDS-PAGE gels, pH4-7. Protein marker bands are demarcated on the left side in kDa. Seen above label C indicates protein spots which observes different profile in the depleted group for sample 3

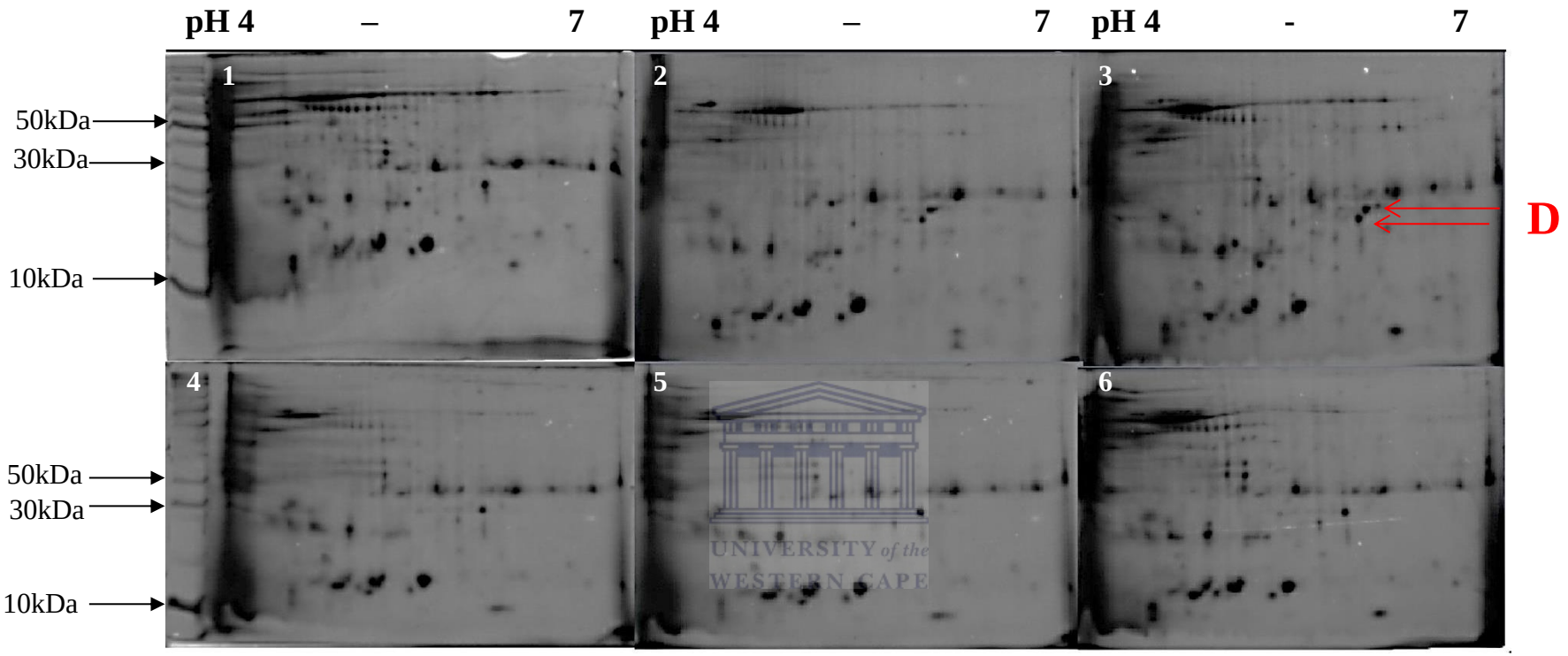


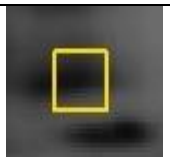
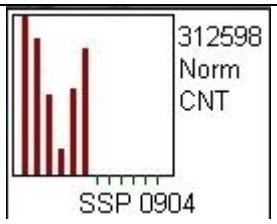
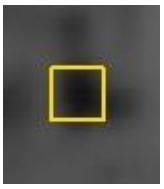
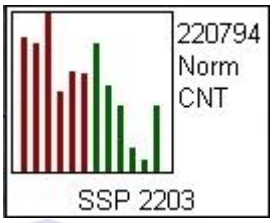
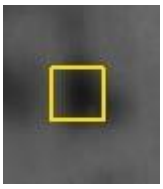
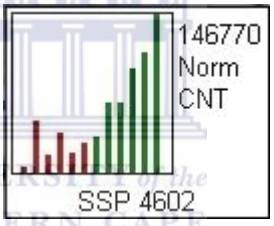
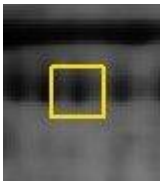
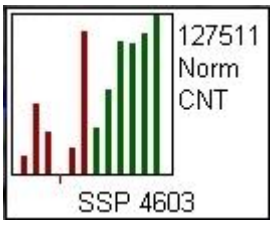
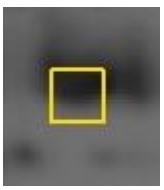
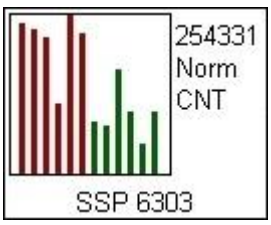
Figure 3.6: Autistic albumin depleted urine protein profiles resolved on 12% 2D SDS-PAGE gels pH4-7. Protein marker bands are demarcated on the left in kDa. The arrows in sample directs to a cluster of protein spots not previously observed for that sample pre albumin depletion.

3.4.3 Protein profile comparisons

PDQuest™ basic version 8.0 2D analysis software was used to match spots between sample gels as well as measure those abundant with quantity differences, and a statistical test the student's t-test. The PDQuest quantity set parameter was used to compare expression patterns with an upper limit of 1.5 fold and lower limit of 0.5 as described in section 3.3.10.

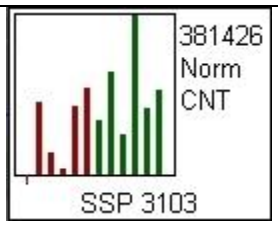
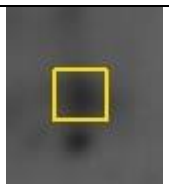
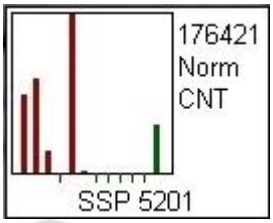
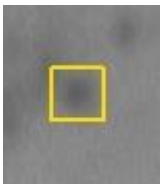
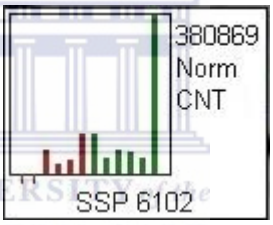
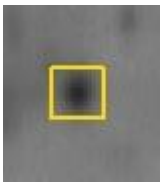
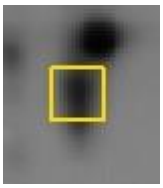
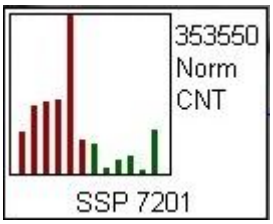
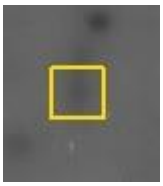
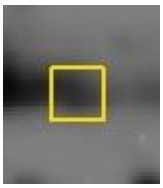
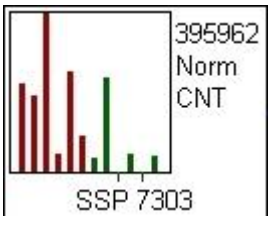
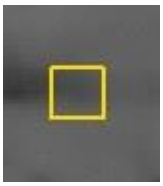
The protein spots were analysed from both the undepleted group and the depleted group (healthy and autistic) to establish whether or not the abundant proteins may affect the profiles for the different groups. The un-depleted healthy protein profiles were analysed against the autistic un-depleted profiles and the healthy depleted profiles against the autistic depleted group. The data obtained from the software presented spots for both the undelepleted and depleted groups. From those only a few were either statistically significant, quantitiatively significant or in some cases considered both. The following tables provides a list of spots identified as significant using this software. The spots in table 3.1 and tables 3.2 displays the spots for the un-depleted and depleted groups nominated from PDQuest. Figures 3.7a and 3.7bIn figure 3.7a and 3.17b , 10 spots were selected for identification. These figures indicate the spots from the two individual gels for spots one to five and six to ten. Spots one to five in figure 3.7a were the albumin depleted sample whereas spots in figure 3.7 b indicate the sample from prealbumin depletion.

Table 3.1: Protein spots selected for identification of albumin depleted samples for the healthy and autistic groups.

Spot number	Healthy	Spot review	Autistic
1			
2			
3			
4			
5			

The yellow squares represent the presence of a protein spot (some may differ in resolution) whereas the yellow circles represent the absence of a spot. Bar graphs: red bars- healthy and green bars- autistic. The bar graphs display the expression patterns of healthy and autistic samples for the respective protein spots *SSP- Sample spot Protein and the occurrence of the protein within the samples.

Table 3.2: Protein spots selected for identification of albumin un-depleted samples for the healthy and autistic groups.

Spot number	Healthy	Spot review	Autistic
6			
7			
8			
9			
10			

The yellow squares represent the presence of a protein spot (some may differ in resolution) whereas the yellow circles represent the absence of a spot.

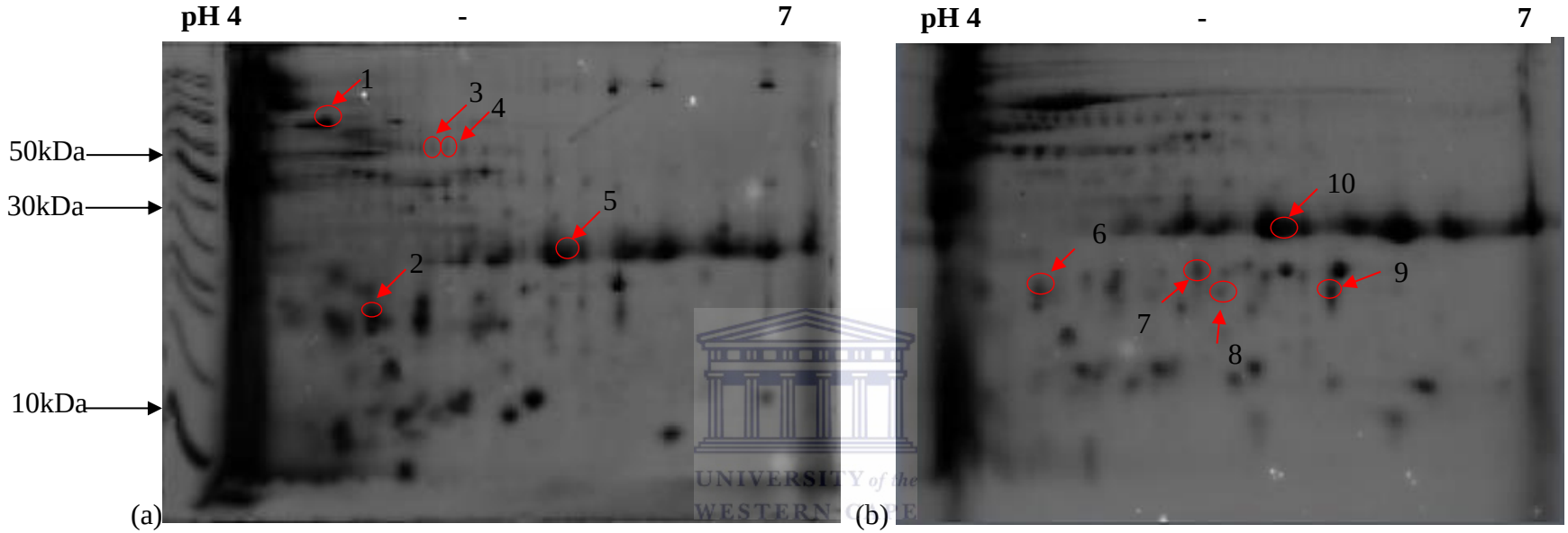


Figure 3.7a: Healthy SA albumin depleted urine protein profile resolved on 12% 2D SDS-PAGE gel with selected spots for analysis.
Figure 3.7b: Healthy SA un-depleted of albumin urine protein profile resolved on 12% 2D SDS-PAGE gel with the respective spots for analysis.

3.4.4 Protein Identification by MS/MS mass spectrometry analysis

The results yielded were a list of potential protein identities in an MS Excel sheet format. Positive identification for protein spots subjected to mass spectrometry is based on a scoring system of the MOWSE score. This score is calculated as -10Log (P), where P stands the absolute probability (P). The score is based on the likelihood that the experimental sequences observed matched to database sequences (e.g. Mascot, MOWSE (HGMP), PeptIdent) is a random event. The list contains the unique score for each possible protein, the percentage of peptide matches between the theoretical and actual proteins as well as the percentage of coverage across the protein map compared to a database. The output data were obtained and tabulated accordingly in tables 3.3 and 3.4. Table 3.3 presents the list of spots identified from the MS analysis. The table includes the estimated and theoretical pI/MW (kDa) for each spot with the confidence score obtained as well as the amount of peptides matched during the identification process. The 10 spots were analyzed however 3 spots were identified as contaminants instead and were labeled red (figure 3.3). Table 3.4 presents the list of positively identified proteins from the MS analysis. Based on the PDQuest software, these spots were either quantitatively significant or statistically significant based on the student's t-test. This table also provides indication list of proteins identified in t two groups

Table3.3: List of proteins identified by mass spectrometry from healthy and autistic samples

Spot Number [a]	Protein Name	Accession number [b]	Estimated pI/MW(kDa)[c]	Theoretical pI/MW(kDa) [d]	Confidence score ($p<0.05$)[e]	Peptides matched/ coverage (%)
1	Uromodulin	P07911	4.4/85	5.05/69.71	262.69	15/26.5
2	Vitelline membrane outer layer protein homologue	Q7Z5L0	4.7/18	4.90/21.52	78.1	4/28.7
3	Kininogen-1	P01042	5/60	6.34/71.91	170.67	10/16.7
4	Alpha-1-antitrypsin	P01009	5.1/60	5.37/46.71	148.97	8/24.1
5	Ig Kappa chain region C	P01834	5.6/40	5.58/11.60	125.84	6/79.2
6	CD59 glycoprotein	P13987	4.5/20	6.02/14.17	67.02	4/25
(7)	Keratin type II cytoskeletal 1		5.5/20		113.47	
(8)	Keratin type I cytoskeleton 10		5.5/15		146.62	
(9)	Keratin, type I cytoskeleton 10		6/16		195.72	
10	Ig Kappa chain C region	P01834	5.8/30	5.58/11.60	130.36	6/79.2

[a] Spot numbers from figure 7 a) and b)
[b] Accession numbers acquired from Swissprot/Uniprot
[c] Estimated pI/MW from 2D gel
[d] Theoretical pI/MW computed from Expasy database
[e] Significant scores were scores > 65 in Mascot.
() Spots classed as contamination, no positive identification were observed besides Keratin

Table 3.4: List of positively identified proteins, from the healthy and autistic samples

Spot Number [a]	Protein Name [b]	Accession number [c]	Protein spot presence in group [d]		PDQuest analysis/significance [e]	
			Healthy	Autistic	Student's t-test	Quantitatively Significant
1	Uromodulin	P07911	++++++	----------	Yes	No
2	Vitelline membrane outer layer protein homologue	Q7Z5L0	++++++	++++++	Yes	No
3	Kininogen-1	P01042	++++++	++++++	No	Yes
4	Alpha-1-antitrypsin	P01009	+++++	++++++	No	Yes
5	Ig Kappa chain region C	P01834	++++++	++++++	No	Yes
6	CD59 glycoprotein	P13987	+++++	++++++	Yes	Yes
10	Ig Kappa chain C region	P01834	++++++	++++	Yes	Yes

[a] Spot numbers from figure 7 a) and b)

[b] Protein name assigned from Swissprot/Uniprot

[c] Accession numbers acquired from Swissprot/Uniprot

[d] The presence of the protein spots found and excised from the experimental groups. The + indicates the level of protein observed in the group (++++++, present in all; ++++, present in only four; +, present in only one; -, none observed)

[e] Spot selection based on two tests/analysis parameters. The Student T-test with a 90 % significance and an quantitative parameter of significance. (Quantitative limits were selected in an outer limit method, with 1.50 fold upper and 0.5 fold lower limits)

3.5 Discussion

Research in proteomic studies has been of great importance due to its potential in biomarker discovery for diagnostics and drug development. In this project, proteins were extracted from urine of healthy and autistic African-South African individuals. This was done to establish a urine proteome profile with success and furthermore to assess the usefulness of these proteomes resolved in 2D SDS-PAGE gels. Successful biomarkers have been identified with other diseases, an example would be breast cancer and the expression of the estrogen receptor (ER) (Weigel and Dowsett, 2010). This has not yet been developed in autism even though few studies have been undertaken but up to date no reliable biomarker candidate exists (Momeni *et al.*, 2012). Urine proteomics have been shown to produce positive results with the benefits of the type of biological material. However one should realize that there are limitations. Firstly the presence of impurities in the urine which affect the isolation of the proteins and secondly, to reach optimum conditions is a lengthy process.

Due to the nature of physical and chemical variability of urine samples important purification steps were required for the preparation of the proteins. As urine is excreted as waste and impurities from the body are present. The presence of the high abundance proteins such as albumin becomes a concern because the smaller proteins are masked therefore preventing the identification of potential biomarker candidate proteins. Furthermore urine protein degrades easily due to handling as well as proteinases present in the sample. These factors pose a risk and have to be addressed.

Urine is known to contain a very high salt content which interferes in protein separation (Ferreira, 2007). Other drawbacks are the high abundance of certain proteins such as Albumin and Immunoglobulins (Igs). These molecules mask other proteins that are present in low concentrations by altering the migration of all proteins during electrophoresis. For early detection of diabetic nephropathy and cardiovascular disease, micro-albuminuria is deemed a clinical important marker (Barratt and Topham, 2007). Another main component present is the Tamm Horsfall Protein or known by the alias Uromodulin (Carvalho *et al.*, 2002). Uromodulin is formed and secreted in the thick ascendant limb of the loop of Henle, and was found in amounts ranging from 20 to 200 mg (per 24 hours) (Carvalho *et al.*, 2002).

The conditions under which urine is investigated also plays a big role on the final product. Differences observed between proteomes are dependent on the following factors: health status, stress, sex, diet and physiological status and conditions. Other differences are also observed within an individual; depending on the timing of collections and the influences on the system during collections (Khan and Packer, 2006). Table 3.5 displays the end result from different collection sets with its benefits as well as the drawbacks.

Table 3.5: Urine collections and differences observed time dependent, (Modified from Dihazi and Műller, 2007)

Urine collection method	Advantage	Disadvantage
First morning void	Free of : Dietary influences Changes due to physical events	Hypertonic and concentrated Bacterial contamination Protein degradation Difficulty reproducing protein pattern
Second morning void	Less protein degradation Better patient control High reproducibility of results Free of : dietary influences, changes due to physical events Stable proteome	Does not represent the process taking place over a 24 hour period
24 hour void collection	Monitoring of kidney function over a long period	Lack of patient control High protein degradation Lack of reproducibility Contamination(overgrowth of contaminating microorganism) Useful only when all urine is collected for the 24 hour period Standardisation for proteomics almost impossible
Random spot collection	Flexibility of collection(at any time of day) Less protein degradation Better patient control Easy to handle for proteomics	Does not represent the processes taking place over a 24hour period Proteome is depending on dietary and physical activity
Catheterizations of the bladder		Invasive Risk of introducing infection Risk of traumatizing the urethra and bladder Urine proteome contamination with blood cell proteins

In the present study, urine protein profiles were mapped and resolved on 1D and 2D SDS-PAGE gels. Protein spots were then mapped and proteins of interest identified using Electrospray Ionization (ESI) Mass spectrometry and the output data matched to the Swissprot database. The profiles were also representatives of albumin depleted urine samples against the opposite being albumin containing samples. The removals of major proteins ideally aim at identifying proteins of smaller size and quantity that goes undetected. Here

albumin was removed, and from figure 3.1 scrutinizing the healthy and autistic samples, it was observed that the healthy samples appear affected by the step. In figure 3.2 the autistic 1D profiles do not seem to appear affected by the step. Any differences observed could be noticed by the increase in bands throughout the samples in the un-depleted lanes as well as the depletion or decrease of band intensities in the un-depleted samples, figure 3.2. Also observed and reported here is that the method of acetone extractions does produce positive/useful results for urine extractions even with the suffering of salt interference which was eliminated with further steps of clarification by dialysis and lyophilizing.

This study was done to evaluate the potential of 2D SDS PAGE followed by mass spectrometry analysis of the urine proteome and the use thereof in understanding ASD. As indicated the method can improve providing optimum conditions when albumin depletion step is incorporated, thus increasing the chance of the identification of lower molecular weight proteins.

Though this was done and some differences were observed, more steps could be incorporated such as Immunoglobulin removal, uromodulin exclusion, and more importantly with regards to the groups (both healthy and autistic) itself, larger sample size and precaution to technical errors and contaminations. Though the Uromodulin displayed unexpected results, removing it can be just as vital, since it does fall under the concept of ‘larger more abundant proteins” masking smaller proteins.

Using PDQuest™, 10 spots were selected based on the significance observed in this program and peptide sequences were analyzed by ESI-MS. The work shows relatively good proteome profiles when compared to previous studies, marker proteins have been observed one such can be considered the Uromodulin protein for reasons to be discussed. The protein output can also been used as indicator of the condition of the kidneys, and indirectly could be used to estimate the value of the results although not done in this study.

The results shown as in table above reports 7 identities of best matched observed between the experimental sequences and those in the database used. These spots were either observed in one group and not the other or observed to have been enhanced in either the control over the autistic and vice versa. The cut off for a significant ion score and for mass spectrometry identification in this analysis set was 65, with a threshold of 5% ($p<0.05$).

3.5.1 Uromodulin/Tamm-Horsfall protein (THP) (UROM_Human: P07911)

A protein discovered nearly 50 years ago, and considered as the most abundant protein is a gene product with an aptitude to work together with the immune system. Furthermore it has roles in renal inflammatory responses and urate metabolism (Turner, 2003, Sedor, 2010, Schmid *et al.*, 2010). Uromodulin is also known as the Tamm-Horsfall protein. These proteins are heavily glycosylated proteins, and under reducing conditions present themselves at ~90kDa (Van Rooijen *et al.*, 1999). Healthy individuals' usually produce about 50 mg of Uromodulin per day and are regarded as safe and protected against ascending urinary tract infections, reducing the risk of kidney stone formations (Schmid *et al.*, 2010).

In this study, this protein could have served as an indicator for healthy profiles by determining the output per day and could in turn add weight to the strength of how well the study was conducted or executed. Individuals suffering a defect or mutation in the gene are diagnosed with familial juvenile hyperuricemic nephropathy (*FJHN*) an autosomal dominant disorder and also a form of medullary cystic kidney disease (*MCKD2*) (Sedor, 2010).

Here it was observed that the expression of this molecule was only present in the healthy individuals but not in the autistic (mass spectrometry confidence score of 262.69). This result does however not necessarily mean a direct link to autism since the function of this protein primarily focuses on kidney regulations. The fact that it was not observed in the autistic group in this work does need to be further verified and confirmed. If this is indeed the case it appeals as to whether or not these individuals may have kidney diseases or lesser forms of infections. Though kidney failures have been associated with autism, it has only been associated through mitochondrial disease. Autism mitochondrial disease (AMD) is a form of ASD where children suffering of AMD have a malfunction in the generation of cellular energy (Davi, 2010).

3.5.2 Vitelline membrane outer layer protein 1 homologue (VMO1_Human: Q7Z5L0)

This protein has two variants according to the Uniprot database, few studies have been conducted on the human *VMO1* gene, however it share close homology with the chicken homologue. A study reported close homology for this protein between human, cow, sheep and camel (Shamsi *et al.,* 2011). Gene ontology shows this protein to have roles in the formation of the chorion-eggshell and is required to maintain the shape of the egg in *Drosophila melanogaster* (The Gene ontology, 2012). Furthermore even though this protein

was observed in all of the samples, the healthy group displayed an enhanced expression pattern when compared to the autistic group. Very little was revealed from this identification and may render this molecule as a candidate for further investigation subject to significance and function.

3.5.3 Kininogen-1 (KNG1_Human: P01042)

Kininogen-1 has been reported to be differentially expressed in Down Syndrome (DS) individuals (Heywood *et al.,* 2011). It was shown that this protein was up-regulated in DS pregnancies when compared to control pregnancies utilizing narrow pH range 2DIGE electrophoresis and MALDI-TOF and Tandem MS/MS (Heywood *et al.*, 2011). Autism have been identified or diagnosed independently from DS however, some individuals suffering of DS may be autistic as well, and therefore the two conditions have been reported to exist in an individual (Vatter, 1998).

The *KNG1* gene has been studied, and there are two types of products known, high molecular weight (HMW) and low molecular weight (LMW) single chains glycoproteins (Kitamura *et al.*, 1985). The results observed here displayed a confidence score of 170.67 with 10 unique peptides matched. Although the methods differ between this study and the DS study, similarly patterns were observed with regards the expression of this protein spot in the autistic group (more enhanced spot) compared to the healthy (refer table 3.1: Albumin depleted healthy and autistic selected spots).

The importance of this protein sheds light on the potential biomarker roles in DS. According to the Uniprot database, *KNG1* are thiol protease-, thrombin and plasma-induced aggregation of thrombocytes inhibitors. Individuals absent of the HMW protein have been shown to have abnormalities of the Hageman factor-dependant pathway of coagulation, kinin formation, fibrinolysis and bradykinin discharge (Liu *et al.*, 1977, Wu *et al.*, 2010b).

3.5.4 Alpha-1antitrypsin (A1AT_Human: P01009)

Alpha-1-antitrypsin is considered the most "abundant circulating serine protease inhibitor present in serum with concentrations of 85–250 mg/dL", (Russo *et al.,* 2009). This protein has major roles in liver and gastrointestinal (GI) regulations and a mutation in this gene is associated with diseases of the liver, lung and GI (Russo *et al.,* 2009). *A1AT* deficiency is a genetic condition and is classed as an autosomal recessive disorder (Kok *et al.,* 2007). The primary function for this protein is the role of "inhibiting the destruction of neutrophil elastase, and the protection against pulmonary damage" (Kok *et al.,* 2007). What marks this particular spot as interesting is the fact that the study conduct by Russo *et al.,*(2009) showed that autistic children had lower levels of this protein in serum when compared to normal children, and that those individuals were suffering from GI diseases, respiratory problems as well as hyperbilirubinemia as well as cancer. In the present study, although different sample type, the *A1AT* protein spots were observed to be more enhanced than observed in the healthy group (refer table 3.1: Albumin depleted healthy and autistic selected spots). Though autistic individuals were studied, the current study suggests and association and contradicts the Russo study. The results obtained could be due to the group size, Russo *et al.,*(2009) study's incorporated 71 members of 16 families of autism to 18 controls and concludes "suggestive association between autistic children with regressive disease and *A1AT* deficiency".

3.5.5 Immunoglobulin (Ig) Kappa chain C region (IGKC_Human: P01834)

This protein was identified for spots 5 and 10 above and in both instances was observed to be qualitatively significant as the controls displayed more enhanced levels of expression when compared to the autistic group. This pattern was observed in both un-depleted and depleted of albumin groups. What could be considered interesting is the fact that both spots were isolated from two different gels, and that the MS results showed a difference in the estimated pI/MW values for each spot (5 and 10). The spots 5 and 10 were collected from the same portion in the gel but from different gels, one after albumin depletion (5) and the other with albumin (10). The difference in MS results could thus suggest a modification or cleavage of protein occurred during the albumin depletion step. Immunoglobulins (Ig) are molecules comprised of two of both light and heavy chains, of which kappa is a type of light chain and furthermore have two domains or regions, the variant (V) and constant (C) domains (Sitnikova and Nei, 1998). They function as antibodies in response to the host immune system (Nextprot beta, 2012). The fact that Ig kappa chain C region is expressed more in the healthy than the autistic resonates the importance of the protein to the immune system and its fragility.

3.5.6 CD59 glycoprotein (CD59_Human: P13987)

CD59 is a membrane glycoprotein of 18-20-kDa that acts as a protector for human blood and vascular cells against cell disruption/death and is assisted through the human serum complement proteins C5b-9, (Ninomiya and Sims, 1992). This protein is an inhibitor of the membrane attack complex (MAC) of complement found on blood cells, endothelia and epithelial cells (Meri *et al.*, 1996). A deficiency of this protein results in chronic hemolysis and peripheral demyelinating disease (Nevo *et al.*, 2013). The results observed for this protein revealed a greater level of expression (enhanced spot intensities) in the autistic group

than seen in the healthy group. Though the score for this protein were 67.02, it had 25% coverage (equivalent to 4 unique peptide matches). The presence of this protein observed more in the autistic than the healthy does however require more evaluation and investigation.

Given the fact that this study aimed at identifying differential proteins, to be considered as a significant factor, is the population in which this study has been conducted. It's known that the South African population exhibits an extreme genetic diversity when compared to the rest of the world especially that of African lineage and within the African population. Though these changes were observed in this group, stringent methods were incorporated to keep the results as reasonable as possible. Since the group was boys, plenty of differences could be observed. The healthy boys of that age are very active, and with the different backgrounds could display numerous differences, yet this study required that urine samples be collected at the time of second void. This allowed for minimal protein degradation, and reproducibility which counts in favor of these types of studies. From table 3.1, it has been demonstrated the benefits of the different collections, and it can be seen that all the different collection time points have both advantages and disadvantages. Depending on the objective of the study, it can be chosen accordingly. When looking at the 1D and 2D gel images, it can be observed, that this collection choice did in fact have an accepted amount of differences, using PDQuest, it could be analysed and a few of corresponding spots across the groups were found.

3.6 Conclusion

The aim of this study was to establish urine proteome profiles of African-SA healthy and autistic individuals and in doing so, execute a comparative analysis of those profiles. The comparison was achieved, and spots were selected based on the significance observed in the t-test and quantitative analysis sets used in PDQuest. Though the results seemed distinct enough, secondary analyses could/may have been employed on different software for more accurate readings and stringent results. The purpose of this study ultimately was to see differences that could be used downstream for diagnostics. In conclusion, the hypothesis that differences in proteome profiles of urine of autistic children would differ compared to healthy children was found. This method could thus render potential proteins with further investigation and could be used for diagnostics in the future.

3.6.1 Limitations of the study

With all the benefits of urine as a source for protein studies it has been limiting in terms of results. The extraction of proteins is a very sensitive step considering that the sample is subject to protein degradation, bacterial contamination and interference from the actual collections and storage, (Dihazi and Müller, 2007). The size of the study plays a role in the results obtained. Only 6 subjects were used in both the healthy and autistic. Given the size of the SA population, a group size of 6-12 is miniscule and therefore the results cannot be seen as representative for the entire SA population.

With respect to the identified spots, it is hard to conclusively state that these results are significant. The selective ethnic group results may differ drastically from another ethnic group type and may yield a whole different set of results. The proteins identified here also

may have undergone modifications, truncations, degradations, and thus may be preliminary results since no specifications was done for the identification of proteins, and no selective staining were incorporated. Typically a variety of versions can be observed between genes and their products as well the amounts produced and "good correlation has been observed in only one third of the observed entities only" (Burkard *et al.*, 2011). Thus the results or observed abundance of these proteins cannot be seen as direct expression from the respective genes.

These results also add weight to the fact that autism is a multifactorial disorder. Several spots were identified of which a few can be related to ASD symptoms and cases depicting the diversity of the disorder and the spectrum, or as mentioned the umbrella concept. The spectrum does also pose a hard road, since no two individuals suffer the same symptoms, making it hard to focus on a few proteins. It would be premature to assume that a few proteins can render insight into the disorder.

3.6.2 Future studies

The way forward in research for autism in SA requires a more in-depth study. The sample sizes used in this study was very small taking into consideration the increase of autistic cases reported within the SA population and the increase of the population itself. A bigger sample group would thus generate more useful results. Furthermore, as a starting project, a general screening of proteins could be investigated in attempt to establish a more general profile, not excluding differences observed between the groups. An important factor in this study is the different ethnic groups. South Africa is composed of multiple ethnicities and genetic diversity within these ethnicities exists. Since most studies were done in other populations,

the default setting should be done befitting the South African population based on its genetic background.

Bibliography

ABERLE, H., SCHWARTZ, H. & KEMLER, R. 1996. Cadherin-catenin complex: Protein interactions and their implications for cadherin function. *Journal of Cellular Biochemistry,* 61**,** 514-523.
ABRAHAMS, B. S. & GESCHWIND, D. H. 2008. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Reviews Genetics,*9**,** 341-355.
ADACHI, J., KUMAR, C., ZHANG, Y., OLSEN, J. V. & MANN, M. 2006. The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biology,* 7**,** R80.
AEBERSOLD, R. & MANN, M. 2003. Mass spectrometry-based proteomics. *Nature,* 422**,** 198-207.
ARMSTRONG, D., DUNN, J. K., ANTALFFY, B. & TRIVEDI, R. 1995. Selective dendritic alterations in the cortex of Rett syndrome. *Journal of Neuropathology & Experimental Neurology,* 54**,** 195-201.
ASHWOOD, P., ANTHONY, A., PELLICER, A. A., TORRENTE, F., WALKER-SMITH, J. A. & WAKEFIELD, A. J. 2003. Intestinal lymphocyte populations in children with regressive autism: evidence for extensive mucosal immunopathology. *Journal of Clinical Immunology,* 23**,** 504-517.
ASSOCIATION, A. P. 1994. Diagnostic and statistical manual of mental disorders American Psychiatric Association. *Washington, dc,* 210.
ATLADÓTTIR, H. Ó., HENRIKSEN, T. B., SCHENDEL, D. E. & PARNER, E. T. 2012. Autism after infection, febrile episodes, and antibiotic use during pregnancy: an exploratory study. *Pediatrics,* 130**,** e1447-e1454.
BAE, M. S., CHO, E. J., CHOI, E. Y. & PARK, O. K. 2003. Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *The Plant Journal,* 36**,** 652-663.
BAILEY, A., LE COUTEUR, A., GOTTESMAN, I., BOLTON, P., SIMONOFF, E., YUZDA, E. & RUTTER, M. 1995. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological medicine,* 25**,** 63-78.
BAIRD, G., CASS, H. & SLONIMS, V. 2003. Diagnosis of autism. *BMJ: British Medical Journal,* 327**,** 488.
BAIRD, G., SIMONOFF, E., PICKLES, A., CHANDLER, S., LOUCAS, T., MELDRUM, D. & CHARMAN, T. 2006. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *The Lancet,* 368**,** 210-215.
BAKARE, M. & MUNIR, K. 2011. Autism spectrum disorders (ASD) in Africa: a perspective. *African Journal of Psychiatry,* 208**,** 14-21.
BAMBINI-JUNIOR, V., RODRIGUES, L., BEHR, G. A., MOREIRA, J. C. F., RIESGO, R. & GOTTFRIED, C. 2011. Animal model of autism induced by prenatal exposure to valproate: behavioral changes and liver parameters. *Brain Research,* 1408**,** 8-16.
BARON-COHEN, S., LOMBARDO, M. V., AUYEUNG, B., ASHWIN, E., CHAKRABARTI, B. & KNICKMEYER, R. 2011. Why are autism spectrum conditions more prevalent in males? *PLoS Biology,* 9**,** e1001081.
BARRATT, J. & TOPHAM, P. 2007. Urine proteomics: the present and future of measuring urinary protein components in disease. *Canadian Medical Association Journal,* 177**,** 361-368.
BAUMAN, M. L. 2010. Medical comorbidities in autism: challenges to diagnosis and treatment. *Neurotherapeutics,* 7**,** 320-327.
BECKER, K. G. 2012. Male gender bias in autism and pediatric autoimmunity. *Autism Research,* 5**,** 77-83.
BENAYED, R., CHOI, J., MATTESON, P. G., GHARANI, N., KAMDAR, S., BRZUSTOWICZ, L. M. & MILLONIG, J. H. 2009. Autism-Associated Haplotype

Affects the Regulation of the Homeobox Gene,< i> ENGRAILED 2</i>. *Biological Psychiatry,* 66**,** 911-917.
BENVENUTO, A., MARCIANO, S., CAPUANO, I. & CURATOLO, P. 2013. An update on autism spectrum disorders in children. *Minerva Pediatrica,* 65**,** 19-36.
BJØRKLUND, G. 1995. Mercury and acrodynia. *The Journal of Orthomolecular Medicine,* 10**,** 145-146.
BLEULER, E. 1911. Dementia praecox oder Gruppe der Schizophrenien. *Handbuch der Psychiatrie*.
BOLTON, P., MACDONALD, H., PICKLES, A., RIOS, P., GOODE, S., CROWSON, M., BAILEY, A. & RUTTER, M. 1994. A CASE - CONTROL FAMILY HISTORY STUDY OF AUTISM. *Journal of Child Psychology and Psychiatry and Allied Disciplines,* 35**,** 877-900.
BOTTINI, N., DE LUCA, D., SACCUCCI, P., FIUMARA, A., ELIA, M., PORFIRIO, M. C., LUCARELLI, P. & CURATOLO, P. 2001. Autism: evidence of association with adenosine deaminase genetic polymorphism. *Neurogenetics,* 3**,** 111-113.
BOYLE, C. A., BOULET, S., SCHIEVE, L. A., COHEN, R. A., BLUMBERG, S. J., YEARGIN-ALLSOPP, M., VISSER, S. & KOGAN, M. D. 2011. Trends in the prevalence of developmental disabilities in US children, 1997-2008. *Pediatrics,* 127**,** 1034-1042.
BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry,* 72**,** 248-254.
BROADSTOCK, M. & DOUGHTY, C. 2003. The effectiveness of pharmacological therapies for young people and adults with Autism Spectrum Disorder (ASD). *New Zealand Health Technology Asssesment Report,* 6.
BURKARD, T. R., PLANYAVSKY, M., KAUPE, I., BREITWIESER, F. P., BURCKSTUMMER, T., BENNETT, K. L., SUPERTI-FURGA, G. & COLINGE, J. 2011. Initial characterization of the human central proteome. *Biomedical Central Systems Biology,* 5**,** 17.
BUTLER, M., DASOUKI, M., ZHOU, X., TALEBIZADEH, Z., BROWN, M., TAKAHASHI, T., MILES, J., WANG, C., STRATTON, R. & PILARSKI, R. 2005. Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *Journal of Medical Genetics,* 42**,** 318-321.
CAMPBELL, M. C. & TISHKOFF, S. A. 2008. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics,* 9**,** 403-433.
CARRASCO, G. A., VAN DE KAR, L. D., SULLIVAN, N. R., LANDRY, M., GARCIA, F., MUMA, N. A. & BATTAGLIA, G. 2006. Cocaine-mediated supersensitivity of 5-HT sub 2A sub receptors in hypothalamic paraventricular nucleus is a withdrawal-induced phenomenon. *Neuroscience,* 143**,** 7-13.
CARVALHO, M., MULINARI, R. & NAKAGAWA, Y. 2002. Role of Tamm-Horsfall protein and uromodulin in calcium oxalate crystallization. *Brazilian Journal of Medical and Biological Research,* 35**,** 1165-1172.
CASTAGNA, A., CECCONI, D., SENNELS, L., RAPPSILBER, J., GUERRIER, L., FORTIS, F., BOSCHETTI, E., LOMAS, L. & RIGHETTI, P. G. 2005. Exploring the Hidden Human Urinary Proteome via Ligand Library Beads. *Journal of Proteome Research,* 4**,** 1917-1930.
CHAHAL, H. S., STALS, K., UNTERLÄNDER, M., BALDING, D. J., THOMAS, M. G., KUMAR, A. V., BESSER, G. M., ATKINSON, A. B., MORRISON, P. J.,

HOWLETT, T. A., LEVY, M. J., ORME, S. M., AKKER, S. A., ABEL, R. L., GROSSMAN, A. B., BURGER, J., ELLARD, S. & KORBONITS, M. 2011. AIP Mutation in Pituitary Adenomas in the 18th Century and Today. *New England Journal of Medicine,* 364**,** 43-50.

CHANDRAMOULI, K. & QIAN, P.-Y. 2009. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human Genomics and Proteomics,* 1.

CHARNEY, D. S., BARLOW, D. H., BOTTERON, K., COHEN, J. D., GOLDMAN, D., GUR, R. E., LIN, K.-M., LÓPEZ, J. F., MEADOR-WOODRUFF, J. H. & MOLDIN, S. O. 2002. Neuroscience research agenda to guide development of a pathophysiologically based classification system.

CHASTE, P. & LEBOYER, M. 2012. Autism risk factors: genes, environment, and gene-environment interactions. *Dialogues in Clinical Neuroscience,* 14**,** 281.

CHEUNG, C., YU, K., FUNG, G., LEUNG, M., WONG, C., LI, Q., SHAM, P., CHUA, S. & MCALONAN, G. 2010. Autistic disorders and schizophrenia: related or remote? An anatomical likelihood estimation. *PloS One,* 5**,** e12233.

CHOJNICKA, I., STRAWA, K., FUDALEJ, S., FUDALEJ, M., PAWLAK, A., KOSTRZEWA, G., WOJNAR, M., KRAJEWSKI, P. & PŁOSKI, R. 2012. Analysis of Four Genes Involved in the Neurodevelopment Shows Association of rs4307059 Polymorphism in the Cadherin 9/10 Region with Completed Suicide. *Neuropsychobiology,* 66**,** 134-140.

CHUGANI, D. C., SUNDRAM, B. S., BEHEN, M., LEE, M.-L. & MOORE, G. J. 1999. Evidence of altered energy metabolism in autistic children. *Progress in Neuro-Psychopharmacology and Biological Psychiatry,* 23**,** 635-641.

CLARKSON, T. W. & MAGOS, L. 2006. The toxicology of mercury and its chemical compounds. *CRC Critical Reviews in Toxicology,* 36**,** 609-662.

COMINGS, D. E., WU, S., CHIU, C., MUHLEMAN, D. & SVERD, J. 1996. Studies of the c-Harvey-Ras gene in psychiatric disorders. *Psychiatry Research,* 63**,** 25-32.

CORBETT, B., KANTOR, A., SCHULMAN, H., WALKER, W., LIT, L., ASHWOOD, P., ROCKE, D. M. & SHARP, F. 2006. A proteomic study of serum from children with autism showing differential expression of apolipoproteins and complement proteins. *Molecular Psychiatry,* 12**,** 292-306.

COSTA, L. G., ASCHNER, M., VITALONE, A., SYVERSEN, T. & SOLDIN, O. P. 2004. Developmental neuropathology of environmental agents. *Annual Review of Pharmacology and Toxicology,* 44**,** 87.

CURTIN, C., ANDERSON, S. E., MUST, A. & BANDINI, L. 2010. The prevalence of obesity in children with autism: a secondary data analysis using nationally representative data from the National Survey of Children's Health. *BioMedical Central Pediatrics,* 10**,** 11.

DAVI, A. 2010. Has your child with autistic symptoms been properly screened for a subset of mitochondrial disease known as oxphos?...probably bot. *The Autism File*.

DAVIDSON, P. W., MYERS, G. J. & WEISS, B. 2004. Mercury exposure and child development outcomes. *Pediatrics,* 113**,** 1023-1029.

DE SOUZA, S. J., LONG, M., SCHOENBACH, L., ROY, S. W. & GILBERT, W. 1997. The correlation between introns and the three-dimensional structure of proteins. *Gene,* 205**,** 141-144.

DIHAZI, H. & MÜLLER, G. A. 2007. Urinary proteomics: a tool to discover biomarkers of kidney diseases.

ELSABBAGH, M., DIVAN, G., KOH, Y. J., KIM, Y. S., KAUCHALI, S., MARCÍN, C., MONTIEL-NAVA, C., PATEL, V., PAULA, C. S. & WANG, C. 2012. Global

prevalence of autism and other pervasive developmental disorders. *Autism Research,* 5**,** 160-179.
EMOND, P., MAVEL, S., AÏDOUD, N., NADAL-DESBARATS, L., MONTIGNY, F., BONNET-BRILHAULT, F., BARTHÉLÉMY, C., MERTEN, M., SARDA, P. & LAUMONNIER, F. 2013. GC-MS-based urine metabolic profiling of autism spectrum disorders. *Analytical and Bioanalytical Chemistry***,** 1-10.
ETZRODT, J., KRISHNA-K, K. & REDIES, C. 2009. Expression of classic cadherins and δ-protocadherins in the developing ferret retina. *BioMedical Central Neuroscience,* 10**,** 153.
EXCOFFIER, L. 2002. Human demographic history: refining the recent African origin model. *Current Opinion in Genetics & Development,* 12**,** 675-682.
FARINA, M., ROCHA, J. B. & ASCHNER, M. 2011. Mechanisms of methylmercury-induced neurotoxicity: evidence from experimental studies. *Life Sciences,* 89**,** 555-563.
FEERO, W. G., GUTTMACHER, A. E., MEFFORD, H. C., BATSHAW, M. L. & HOFFMAN, E. P. 2012. Genomics, intellectual disability, and autism. *New England Journal of Medicine,* 366**,** 733-743.
FERREIRA, L. 2007. Use of Urinary Proteomics for the Discovery of Biomarkers of Renal Disease.
FLINTOFT, L. 2009. Complex disease: Autism clues from genome-wide studies. *Nature Reviews Genetics,* 10**,** 346-347.
FOLL, M. & GAGGIOTTI, O. 2006. Identifying the environmental factors that determine the genetic structure of populations. *Genetics,* 174**,** 875-891.
FOLSTEIN, S. E. & RUTTER, M. L. 1988. Autism: Familial aggregation and genetic implications. *Journal of autism and developmental disorders,* 18**,** 3-30.
FRAZER, A. & HENSLER, J. G. 1999. Serotonin. *Basic Neurochemistry,* 6**,** 335-346.
GADOW, K. D., DEVINCENT, C. J., SIEGAL, V. I., OLVET, D. M., KIBRIA, S., KIRSCH, S. F. & HATCHWELL, E. 2013. Allele-specific associations of 5-HTTLPR/rs25531 with ADHD and autism spectrum disorder. *Progress in Neuropsychopharmacoly and Biological Psychiatry,* 40**,** 292-297.
GHARANI, N., BENAYED, R., MANCUSO, V., BRZUSTOWICZ, L. & MILLONIG, J. 2004. Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder. *Molecular Psychiatry,* 9**,** 474-484.
GILLBERG, C. 1998. Chromosomal disorders and autism. *Journal of Autism and Developmental Disorders,* 28**,** 415-425.
GILLBERG, C., STEFFENBURG, S. & SCHAUMANN, H. 1991. Is autism more common now than ten years ago? *The British Journal of Psychiatry,* 158**,** 403-409.
GLESSNER, J. T., WANG, K., CAI, G., KORVATSKA, O., KIM, C. E., WOOD, S., ZHANG, H., ESTES, A., BRUNE, C. W. & BRADFIELD, J. P. 2009. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature,* 459**,** 569-573.
GRANT, B. 2011. Opening a Can of Worms-A father's determination to help his son resulted in an experimental treatment for autism that uses roundworms to modulate inflammatory immune responses. Can the worms be used to treat other diseases? *Scientist,* 25**,** 42.
GRANT, R. & NOZYCE, M. 2013. Proposed changes to the American Psychiatric Association diagnostic criteria for autism spectrum disorder: implications for young children and their families. *Maternal and Child Health Journal,* 17**,** 586-592.

GUIDOTTI, A., AUTA, J., DAVIS, J. M. & ET AL. 2000. Decrease in reelin and glutamic acid decarboxylase67 (gad67) expression in schizophrenia and bipolar disorder: A postmortem brain study. *Archives of General Psychiatry,* 57**,** 1061-1069.

GUMBINER, B. M. 1996. Cell adhesion: The molecular basis of tissue architecture and morphogenesis. *Cell,* 84**,** 345-357.

HATTA, K., NOSE, A., NAGAFUCHI, A. & TAKEICHI, M. 1988. Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion molecule: its identity in the cadherin gene family. *The Journal of Cell Biology,* 106**,** 873-881.

HERZING, L. B. K., KIM, S.-J., COOK, E. H. & LEDBETTER, D. H. 2001. The Human Aminophospholipid-Transporting ATPase Gene ATP10C Maps Adjacent to UBE3A and Exhibits Similar Imprinted Expression. *The American Journal of Human Genetics,* 68**,** 1501-1505.

HEYWOOD, W. E., MADGETT, T. E., WANG, D., WALLINGTON, A., HOGG, J., MILLS, K. & AVENT, N. D. 2011. 2D DIGE analysis of maternal plasma for potential biomarkers of Down Syndrome. *Proteome Science,* 9**,** 56.

HINDORFF, L. A., SETHUPATHY, P., JUNKINS, H. A., RAMOS, E. M., MEHTA, J. P., COLLINS, F. S. & MANOLIO, T. A. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences,* 106**,** 9362-9367.

HIRANO, S., SUZUKI, S. T. & REDIES, C. 2003. The cadherin superfamily in neural development: diversity, function and interaction with other molecules. *Frontiers in Bioscience: A Journal and Virtual Library,* 8**,** d306.

HONJO, Y., NAKAGAWA, S. & TAKEICHI, M. 2000. Blockade of cadherin-6B activity perturbs the distribution of PSD-95 family proteins in retinal neurones. *Genes to Cells,* 5**,** 309-318.

HULPIAU, P. & VAN ROY, F. 2009. Molecular evolution of the cadherin superfamily. *The International Journal of Biochemistry & Cell Biology,* 41**,** 349-369.

IMPAGNATIELLO, F., GUIDOTTI, A. R., PESOLD, C., DWIVEDI, Y., CARUNCHO, H., PISU, M. G., UZUNOV, D. P., SMALHEISER, N. R., DAVIS, J. M., PANDEY, G. N., PAPPAS, G. D., TUETING, P., SHARMA, R. P. & COSTA, E. 1998. A decrease of reelin expression as a putative vulnerability factor in schizophrenia. *Proceedings of the National Academy of Sciences,* 95**,** 15718-15723.

INGRAM, J. L., STODGELL, C. J., HYMAN, S. L., FIGLEWICZ, D. A., WEITKAMP, L. R. & RODIER, P. M. 2000. Discovery of allelic variants of HOXA1 and HOXB1: genetic susceptibility to autism spectrum disorders. *Teratology,* 62**,** 393-405.

INITIATIVE, A. R. 2005. Treatment Options for Mercury/metal Toxicity in Autism and Related Developmental Disabilities: Consensus Position Paper. San Diego, CA: Autism Research Initiative.

JAMAIN, S., QUACH, H., BETANCUR, C., RASTAM, M., COLINEAUX, C., GILLBERG, I. C., SODERSTROM, H., GIROS, B., LEBOYER, M., GILLBERG, C. & BOURGERON, T. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genetics,* 34**,** 27-29.

JULIANO, R. 2002. Signal transduction by cell adhesion receptors and the cytoskeleton: functions of integrins, cadherins, selectins, and immunoglobulin-superfamily members. *Annual Review of Pharmacology and Toxicology,* 42**,** 283-323.

JUNAID, M. A., KOWAL, D., BARUA, M., PULLARKAT, P. S., SKLOWER BROOKS, S. & PULLARKAT, R. K. 2004. Proteomic studies identified a single nucleotide polymorphism in glyoxalase I as autism susceptibility factor. *American Journal of Medical Genetics Part A,* 131**,** 11-17.

KADESJÖ, B., GILLBERG, C. & HAGBERG, B. 1999. Brief report: autism and Asperger syndrome in seven-year-old children: a total population study. *Journal of Autism and Developmental Disorders,* 29**,** 327-331.
KAUFMANN, W. E., CORTELL, R., KAU, A. S., BUKELIS, I., TIERNEY, E., GRAY, R. M., COX, C., CAPONE, G. T. & STANARD, P. 2004. Autism spectrum disorder in fragile X syndrome: communication, social interaction, and specific behaviors. *American Journal of Medical Genetics Part A,* 129**,** 225-234.
KENTSIS, A., MONIGATTI, F., DORFF, K., CAMPAGNE, F., BACHUR, R. & STEEN, H. 2009. Urine proteomics for profiling of human disease using high accuracy mass spectrometry. *Proteomics-Clinical Applications,* 3**,** 1052-1061.
KHAN, A. & PACKER, N. H. 2006. Simple urinary sample preparation for proteomic analysis. *Journal of Proteome Research,* 5**,** 2824-2838.
KIDD, P. M. 2002. Autism, an extreme challenge to integrative medicine. Part 1: the knowledge base. *Alternative Medicine Review,* 7**,** 292-316.
KIM, S., KIM, J. H., PARK, M., CHO, I. H. & YOO, H. J. 2007. Family-based association study between< i> GRIK2</i> polymorphisms and autism spectrum disorders in the Korean trios. *Neuroscience Research,* 58**,** 332-335.
KIM, Y. S., LEVENTHAL, B. L., KOH, Y.-J., FOMBONNE, E., LASKA, E., LIM, E.-C., CHEON, K.-A., KIM, S.-J., KIM, Y.-K. & LEE, H. 2011. Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry,* 168**,** 904-912.
KITAMURA, N., KITAGAWA, H., FUKUSHIMA, D., TAKAGAKI, Y., MIYATA, T. & NAKANISHI, S. 1985. Structural organization of the human kininogen gene and a model for its evolution. *Journal of Biological Chemistry,* 260**,** 8610-8617.
KNIGHT, J. C. 2009. Genetics and the general physician: insights, applications and future challenges. *Quarterly Journal of Medicine,* 102**,** 757-772.
KOGAN, M. D., BLUMBERG, S. J., SCHIEVE, L. A., BOYLE, C. A., PERRIN, J. M., GHANDOUR, R. M., SINGH, G. K., STRICKLAND, B. B., TREVATHAN, E. & VAN DYCK, P. C. 2009. Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics,* 124**,** 1395-403.
KOK, K., WAHAB, P., HOUWEN, R., DRENTH, J., DE MAN, R., VAN HOEK, B., MEIJER, J., WILLEKENS, F. & DE VRIES, R. 2007. Heterozygous alpha-I antitrypsin deficiency as a co-factor in the development of chronic liver disease: a review. *Netherlands Journal of Medicine,* 65**,** 160-106.
KOOLS, P., VANHALST, K., VAN DEN EYNDE, E. & VAN ROY, F. 1999. The human cadherin-10 gene: complete coding sequence, predominant expression in the brain, and mapping on chromosome 5p13–14. *FEBS letters,* 452**,** 328-334.
KORVATSKA, E., VAN DE WATER, J., ANDERS, T. F. & GERSHWIN, M. E. 2002. Genetic and Immunologic Considerations in Autism. *Neurobiology of Disease,* 9**,** 107-125.
LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.
LAURITSEN, M. B., PEDERSEN, C. B. & MORTENSEN, P. B. 2005. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. *Journal of Child Psychology and Psychiatry,* 46**,** 963-971.
LEE, D. Y. & KONG, X. 2012. Potential Treatment of Autism with Traditional Chinese Medicine. *North American Journal of Medicine and Science Jul,* 5**,** 189.
LELE, Z., FOLCHERT, A., CONCHA, M., RAUCH, G.-J., GEISLER, R., ROSA, F., WILSON, S. W., HAMMERSCHMIDT, M. & BALLY-CUIF, L. 2002. Parachute/n-

cadherin is required for morphogenesis and maintained integrity of the zebrafish neural tube. *Development,* 129**,** 3281-3294.
LEWIS, C. M. & KNIGHT, J. 2012. Introduction to genetic association studies. *Cold Spring Harbor Protocols,* 2012**,** 297-306.
LI, H., YAMAGATA, T., MORI, M. & MOMOI, M. Y. 2005. Absence of causative mutations and presence of autism-related allele in< i> FOXP2</i> in Japanese autistic patients. *Brain and Development,* 27**,** 207-210.
LI, N., CHEN, G., SONG, X., DU, W. & ZHENG, X. 2011. Prevalence of autism-caused disability among Chinese children: a national population-based survey. *Epilepsy & Behavior,* 22**,** 786-789.
LIU, C., SCOTT, C., BAGDASARIAN, A., PIERCE, J., KAPLAN, A. & COLMAN, R. 1977. Potentiation of the function of Hageman factor fragments by high molecular weight kininogen. *Journal of Clinical Investigation,* 60**,** 7.
LIU, Q., DUFF, R. J., LIU, B., WILSON, A. L., BABB-CLENDENON, S. G., FRANCL, J. & MARRS, J. A. 2006. Expression of< i> cadherin10</i>, a type II classic cadherin gene, in the nervous system of the embryonic zebrafish. *Gene Expression Patterns,* 6**,** 703-710.
LU, A., DAI, X., MARTINEZ-AGOSTO, J. A. & CANTOR, R. M. 2012. Support for calcium channel gene defects in autism spectrum disorders. *Molecular Autism,* 3**,** 18.
LUHRMANN, T. M. 2007. Social defeat and the culture of chronicity: or, why schizophrenia does so well over there and so badly here. *Cult Med Psychiatry,* 31**,** 135-72.
MADDALENA, A., RICHARDS, C. S., MCGINNISS, M. J., BROTHMAN, A., DESNICK, R. J., GRIER, R. E., HIRSCH, B., JACKY, P., MCDOWELL, G. A. & POPOVICH, B. 2001. Technical standards and guidelines for fragile X: the first of a series of disease-specific supplements to the Standards and Guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics. *Genetics in Medicine,* 3**,** 200-205.
MAHAFFEY, K. R., CLICKNER, R. P. & BODUROW, C. C. 2004. Blood organic mercury and dietary mercury intake: National Health and Nutrition Examination Survey, 1999 and 2000. *Environmental Health Perspectives,* 112**,** 562.
MAJKIĆ-SINGH, N. 2011. What is a biomarker? From its discovery to clinical application.
MALICKI, J., JO, H. & PUJIC, Z. 2003. Zebrafish N-cadherin, encoded by the< i> glass onion</i> locus, plays an essential role in retinal patterning. *Developmental Biology,* 259**,** 95-108.
MALKOVSKÝ, M., SONDEL, P., STROBER, W. & DALGLEISH, A. 1988. The interleukins in acquired disease. *Clinical and Experimental Immunology,* 74**,** 151.
MANABE, T., TOGASHI, H., UCHIDA, N., SUZUKI, S. C., HAYAKAWA, Y., YAMAMOTO, M., YODA, H., MIYAKAWA, T., TAKEICHI, M. & CHISAKA, O. 2000. Loss of cadherin-11 adhesion receptor enhances plastic changes in hippocampal synapses and modifies behavioral responses. *Molecular and Cellular Neuroscience,* 15**,** 534-546.
MANOLIO, T. A., BROOKS, L. D. & COLLINS, F. S. 2008. A HapMap harvest of insights into the genetics of common disease. *Journal of Clinical Investigation,* 118**,** 1590-1605.
MARUI, T., HASHIMOTO, O., NANBA, E., KATO, C., TOCHIGI, M., UMEKAGE, T., ISHIJIMA, M., KOHDA, K., KATO, N. & SASAKI, T. 2004. Association between the neurofibromatosis-1 (NF1) locus and autism in the Japanese population. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 131B**,** 43-47.
MASAI, I., LELE, Z., YAMAGUCHI, M., KOMORI, A., NAKATA, A., NISHIWAKI, Y., WADA, H., TANAKA, H., NOJIMA, Y. & HAMMERSCHMIDT, M. 2003. N-

cadherin mediates retinal lamination, maintenance of forebrain compartments and patterning of retinal neurites. *Development,* 130**,** 2479-2494.
MCCAULEY, J. L., OLSON, L. M., DOWD, M., AMIN, T., STEELE, A., BLAKELY, R. D., FOLSTEIN, S. E., HAINES, J. L. & SUTCLIFFE, J. S. 2004. Linkage and association analysis at the serotonin transporter (SLC6A4) locus in a rigid-compulsive subset of autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 127B**,** 104-112.
MCCLELLAN, J. & KING, M.-C. 2010. Genetic Heterogeneity in Human Disease. *Cell,* 141**,** 210-217.
MERI, S., LEHTO, T., SUTTON, C., TYYNELA, J. & BAUMANN, M. 1996. Structural composition and functional characterization of soluble CD59: heterogeneity of the oligosaccharide and glycophosphoinositol (GPI) anchor revealed by laser-desorption mass spectrometric analysis. *Biochemistry Journal,* 316**,** 923-935.
MEYER, U., FELDON, J. & DAMMANN, O. 2011. Schizophrenia and autism: both shared and disorder-specific pathogenesis via perinatal inflammation? *Pediatric Research,* 69**,** 26R-33R.
MIDDLETON, F. A., VARLINSKAYA, E. I. & MOONEY, S. M. 2012. Molecular Substrates of Social Avoidance Seen following Prenatal Ethanol Exposure and Its Reversal by Social Enrichment. *Developmental Neuroscience,* 34**,** 115-128.
MILES, J. H. 2011. Autism spectrum disorders—a genetics review. *Genetics in Medicine,* 13**,** 278-294.
MOMENI, N., BERGQUIST, J., BRUDIN, L., BEHNIA, F., SIVBERG, B., JOGHATAEI, M. T. & PERSSON, B. L. 2012. A novel blood-based biomarker for detection of autism spectrum disorders. *Translational Psychiatry,* 2**,** e91.
MONTEOLIVA, L. & ALBAR, J. P. 2004. Differential proteomics: an overview of gel and non-gel based approaches. *Briefings in Functional Genomics & Proteomics,* 3**,** 220-239.
MOONEY, S. M. & MILLER, M. W. 2001. Effects of prenatal exposure to ethanol on the expression of bcl-2, bax and caspase 3 in the developing rat cerebral cortex and thalamus. *Brain Research,* 911**,** 71-81.
MOSTAFA, G. A. & AL-AYADHI, L. Y. 2012. The relationship between the increased frequency of serum antineuronal antibodies and the severity of autism in children. *European Journal of Paediatric Neurology,* 16**,** 464-468.
MUHLE, R., TRENTACOSTE, S. V. & RAPIN, I. 2004. The Genetics of Autism. *Pediatrics,* 113**,** e472-e486.
MUTTER, J., NAUMANN, J., SCHNEIDER, R., WALACH, H. & HALEY, B. 2005. Mercury and autism: accelerating evidence. *Neuroendocrinology Letters,* 26**,** 439-446.
MYCHASIUK, R., RICHARDS, S., NAKAHASHI, A., KOLB, B. & GIBB, R. 2012. Effects of rat prenatal exposure to valproic acid on behaviour and neuro-anatomy. *Developmental Neuroscience,* 34**,** 268-276.
NAKAGAWA, S. & TAKEICHI, M. 1995. Neural crest cell-cell adhesion controlled by sequential and subpopulation-specific expression of novel cadherins. *Development,* 121**,** 1321-1332.
NAKAMURA, M., UENO, S., SANO, A. & TANABE, H. 2000. The human serotonin transporter gene linked polymorphism (5-HTTLPR) shows ten novel allelic variants. *Molecular Psychiatry,* 5**,** 32-38.
NEVO, Y., BEN-ZEEV, B., TABIB, A., STRAUSSBERG, R., ANIKSTER, Y., SHORER, Z., FATTAL-VALEVSKI, A., TA-SHMA, A., AHARONI, S. & RABIE, M. 2013. CD59 deficiency is associated with chronic hemolysis and childhood relapsing immune-mediated polyneuropathy. *Blood,* 121**,** 129-135.

NEWSCHAFFER, C. J., CROEN, L. A., DANIELS, J., GIARELLI, E., GRETHER, J. K., LEVY, S. E., MANDELL, D. S., MILLER, L. A., PINTO-MARTIN, J. & REAVEN, J. 2007. The epidemiology of autism spectrum disorders*. *Annual Review of Public Health,* 28**,** 235-258.

NGARA, R. & NDIMBA, B. K. 2011. Mapping and characterisation of the sorghum cell suspension culture secretome. *African Journal of Biotechnology,* 10**,** 253-266.

NIESSEN, C. M. & LECKBAND, D. 2011. Tissue organization by cadherin adhesion molecules: dynamic molecular and cellular mechanisms of morphogenetic regulation. *Physiological Reviews,* 91**,** 691-731.

NOSE, A., NAGAFUCHI, A. & TAKEICHI, M. 1988. Expressed recombinant cadherins mediate cell sorting in model systems. *Cell,* 54**,** 993-1001.

OH, I. S., PARK, A. R., BAE, M. S., KWON, S. J., KIM, Y. S., LEE, J. E., KANG, N. Y., LEE, S., CHEONG, H. & PARK, O. K. 2005. Secretome analysis reveals an Arabidopsis lipase involved in defense against Alternaria brassicicola. *The Plant Cell Online,* 17**,** 2832-2847.

ORNOY, A. 2009. Valproic acid in pregnancy: how much are we endangering the embryo and fetus? *Reproductive Toxicology,* 28**,** 1-10.

PANDEY, A. & MANN, M. 2000. Proteomics to study genes and genomes. *Nature,* 405**,** 837-846.

PERCY, A. K. 2001. Rett syndrome: clinical correlates of the newly discovered gene. *Brain and Development,* 23**,** S202-S205.

PERSICO, A. M., D'AGRUMA, L., ZELANTE, L., MILITERNI, R., BRAVACCIO, C., SCHNEIDER, C., MELMED, R., TRILLO, S., MONTECCHI, F., ELIA, M., PALERMO, M., RABINOWITZ, D., PASCUCCI, T., PUGLISI-ALLEGRA, S., REICHELT, K. L., MUSCARELLA, L., GUARNIERI, V., MELGARI, J. M., CONCIATORI, M. & KELLER, F. 2004. Enhanced APOE2 transmission rates in families with autistic probands. *Psychiatric genetics,* 14**,** 73-82.

PHILIPPE, A., MARTINEZ, M., GUILLOUD-BATAILLE, M., GILLBERG, C., RÅSTAM, M., SPONHEIM, E., COLEMAN, M., ZAPPELLA, M., ASCHAUER, H. & VAN MALDERGEM, L. 1999. Genome-wide scan for autism susceptibility genes. *Human Molecular Genetics,* 8**,** 805-812.

PIEPER, R., GATLIN, C. L., MCGRATH, A. M., MAKUSKY, A. J., MONDAL, M., SEONARAIN, M., FIELD, E., SCHATZ, C. R., ESTOCK, M. A., AHMED, N., ANDERSON, N. G. & STEINER, S. 2004. Characterization of the human urinary proteome: A method for high-resolution display of urinary proteins on two-dimensional electrophoresis gels with a yield of nearly 1400 distinct protein spots. *Proteomics,* 4**,** 1159-1174.

PISITKUN, T., SHEN, R.-F. & KNEPPER, M. A. 2004. Identification and proteomic profiling of exosomes in human urine. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 13368-13373.

PÖTTER, E., BERGWITZ, C. & BRABANT, G. 1999. The cadherin-catenin system: implications for growth and differentiation of endocrine tissues. *Endocrine Reviews,* 20**,** 207.

POURCAIN, B. S., WANG, K., GLESSNER, J. T., GOLDING, J., STEER, C., RING, S. M., SKUSE, D. H., GRANT, S. F., HAKONARSON, H. & SMITH, G. D. 2010. Association between a high-risk autism locus on 5p14 and social communication spectrum phenotypes in the general population. *The American Journal of Psychiatry,* 167**,** 1364.

PRANDINI, P., PASQUALI, A., MALERBA, G., MAROSTICA, A., ZUSI, C., XUMERLE, L., MUGLIA, P., DA ROS, L., RATTI, E. & TRABETTI, E. 2012. The association of

rs4307059 and rs35678 markers with autism spectrum disorders is replicated in Italian families. *Psychiatric Genetics,* 22**,** 177-181.
RADICE, G. L., RAYBURN, H., MATSUNAMI, H., KNUDSEN, K. A., TAKEICHI, M. & HYNES, R. O. 1997. Developmental defects in mouse embryos lacking N-cadherin. *Developmental Biology,* 181**,** 64-78.
RAMOZ, N., REICHERT, J. G., SMITH, C. J., SILVERMAN, J. M., BESPALOVA, I. N., DAVIS, K. L. & BUXBAUM, J. D. 2004. Linkage and association of the mitochondrial aspartate/glutamate carrier SLC25A12 gene with autism. *American Journal of Psychiatry,* 161**,** 662-669.
RAPIN, I. & TUCHMAN, R. F. 2008. Autism: definition, neurobiology, screening, diagnosis. *Pediatrics Clinics of North America,* 55**,** 1129-1146, viii.
REDIES, C. & TAKEICHI, M. 1996. Cadherins in the developing central nervous system: an adhesive code for segmental and functional subdivisions. *Developmental Biology,* 180**,** 413-423.
ROELOFSEN, H., ALVAREZ-LLAMAS, G., SCHEPERS, M., LANDMAN, K. & VONK, R. J. 2007. Proteomics profiling of urine with surface enhanced laser desorption/ionization time of flight mass spectrometry. *Proteome Science,* 5**,** 2.
RUSSO, A. J., NEVILLE, L. & WROGE, C. 2009. Low serum alpha-1 antitrypsin (AAT) in family members of individuals with autism correlates with PiMZ genotype. *Biomarker Insights,* 4**,** 45.
SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S. C., KAKOL, J. M., STEIN, L. D., MARTH, G., SHERRY, S., MULLIKIN, J. C., MORTIMORE, B. J. & WILLEY, D. L. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-933.
SANDERS, J. L. 2009. Qualitative or quantitative differences between Asperger's disorder and autism? Historical considerations. *Journal of Autism and Developmental Disorders,* 39**,** 1560-1567.
SCHANEN, N. C. 2006. Epigenetics of autism spectrum disorders. *Human Molecular Genetics,* 15**,** R138-R150.
SCHMID, M., PRAJCZER, S., GRUBER, L. N., BERTOCCHI, C., GANDINI, R., PFALLER, W., JENNINGS, P. & JOANNIDIS, M. 2010. Uromodulin Facilitates Neutrophil Migration Across Renal Epithelial Monolayers. *Cellular Physiology and Biochemistry,* 26**,** 311-318.
SEBAT, J., LAKSHMI, B., MALHOTRA, D., TROGE, J., LESE-MARTIN, C., WALSH, T., YAMROM, B., YOON, S., KRASNITZ, A. & KENDALL, J. 2007. Strong association of de novo copy number mutations with autism. *Science,* 316**,** 445-449.
SEDOR, J. R. 2010. Uromodulin and translational medicine: will the SNPs bring zip to clinical practice? *Journal of the American Society of Nephrology,* 21**,** 204-206.
SEIF ELDIN, A., HABIB, D., NOUFAL, A., FARRAG, S., BAZAID, K., AL-SHARBATI, M., BADR, H., MOUSSA, S., ESSALI, A. & GADDOUR, N. 2008. Use of M-CHAT for a multinational screening of young children with autism in the Arab countries. *International Review of Psychiatry,* 20**,** 281-289.
SHAMSI, F. A., CHEN, Z., LIANG, J., LI, K., AL-RAJHI, A. A., CHAUDHRY, I. A., LI, M. & WU, K. 2011. Analysis and comparison of proteomic profiles of tear fluid from human, cow, sheep, and camel eyes. *Investigative Ophthalmology and Visual Science,* 52**,** 9156-65.
SHAPIRO, L. & COLMAN, D. R. 1999. The diversity of cadherins and implications for a synaptic adhesive code in the CNS. *Neuron,* 23**,** 427-430.
SHARMA, J. R., ARIEFF, Z., GAMEELDIEN, H., DAVIDS, M., KAUR, M. & VAN DER MERWE, L. 2013. Association Analysis of Two Single-Nucleotide Polymorphisms of

the RELN Gene with Autism in the South African Population. *Genetic Testing and Molecular Biomarkers,* 17**,** 93-98.
SHIMOYAMA, Y., TSUJIMOTO, G., KITAJIMA, M. & NATORI, M. 2000. Identification of three human type-II classic cadherins and frequent heterophilic interactions between different subclasses of type-II classic cadherins. *Biochemical Journal,* 349**,** 159-167.
SITNIKOVA, T. & NEI, M. 1998. Evolution of immunoglobulin kappa chain variable region genes in vertebrates. *Molecular Biology and Evolution,* 15**,** 50-60.
SKAAR, D., SHAO, Y., HAINES, J., STENGER, J., JAWORSKI, J., MARTIN, E., DELONG, G., MOORE, J., MCCAULEY, J. & SUTCLIFFE, J. 2004. Analysis of the RELN gene as a genetic risk factor for autism. *Molecular Psychiatry,* 10**,** 563-571.
SPERRY, R. W. 1963. Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proceedings of the National Academy of Sciences of the United States of America,* 50**,** 703.
SPONHEIM, E. & SKJELDAL, O. 1998. Autism and related disorders: epidemiological findings in a Norwegian study using ICD-10 diagnostic criteria. *Journal of Autism and Developmental Disorders,* 28**,** 217-227.
STREHL, S., GLATT, K., LIU, Q. M., GLATT, H. & LALANDE, M. 1998. Characterization of Two Novel Protocadherins (PCDH8andPCDH9) Localized on Human Chromosome 13 and Mouse Chromosome 14. *Genomics,* 53**,** 81-89.
SUN, W., LI, F., WU, S., WANG, X., ZHENG, D., WANG, J. & GAO, Y. 2005. Human urine proteome analysis by three separation approaches. *Proteomics,* 5**,** 4994-5001.
SUTCLIFFE, J. S., DELAHANTY, R. J., PRASAD, H. C., MCCAULEY, J. L., HAN, Q., JIANG, L., LI, C., FOLSTEIN, S. E. & BLAKELY, R. D. 2005. Allelic Heterogeneity at the Serotonin Transporter Locus (SLC6A4) Confers Susceptibility to Autism and Rigid-Compulsive Behaviors. *The American Journal of Human Genetics,* 77**,** 265-279.
SUZUKI, S., SANO, K. & TANIHARA, H. 1991. Diversity of the cadherin family: evidence for eight new cadherins in nervous tissue. *Cell Regulation,* 2**,** 261.
SUZUKI, S. C., INOUE, T., KIMURA, Y., TANAKA, T. & TAKEICHI, M. 1997. Neuronal circuits are subdivided by differential expression of type-II classic cadherins in postnatal mouse brains. *Molecular and Cellular Neuroscience,* 9**,** 433-447.
TAKEICHI, M. 1990. Cadherins: a molecular family important in selective cell-cell adhesion. *Annual Review of Biochemistry,* 59**,** 237-252.
TAKEICHI, M. 1995. MORPHOGENETIC ROLES OF CLASSIC CADHERINS. *Current Opinion in Cell Biology,* 7**,** 619-627.
TAKEICHI, M. 2007. The cadherin superfamily in neuronal connections and interactions. *Nature Reviews of Neuroscience,* 8**,** 11-20.
TANAKA, H., SHAN, W. S., PHILLIPS, G. R., ARNDT, K., BOZDAGI, O., SHAPIRO, L., HUNTLEY, G. W., BENSON, D. L. & COLMAN, D. R. 2000. Molecular modification of N-cadherin in response to synaptic activity. *Neuron,* 25**,** 93-107.
TANG, L. X., HUNG, C. P. & SCHUMAN, E. M. 1998. A role for the cadherin family of cell adhesion molecules in hippocampal long term potentiation. *Neuron,* 20**,** 1165-1175.
TEPASS, U. 1999. Genetic analysis of cadherin function in animal morphogenesis. *Current Opinion in Cell Biology,* 11**,** 540-548.
THEODORESCU, D. & MISCHAK, H. 2007. Mass spectrometry based proteomics in urine biomarker discovery. *World Journal of Urology,* 25**,** 435-443.

THONGBOONKERD, V., MCLEISH, K. R., ARTHUR, J. M. & KLEIN, J. B. 2002. Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *Kidney International,* 62**,** 1461-1469.
TODOROV, S. & ARNAOUDOVA, M. 2012. On Differential Diagnosis between Autistic Disorder and Asperger's Syndrome. *Journal of IMAB - Annual Proceeding (Scientific Papers),* 18, 3**,** 334-336.
TREFFERT, D. 1999. The savant syndrome and autistic disorder. *CNS Spectrums,* 4**,** 57.
TURNER, J. J. O. 2003. UROMODULIN Mutations Cause Familial Juvenile Hyperuricemic Nephropathy. *Journal of Clinical Endocrinology & Metabolism,* 88**,** 1398-1401.
UCHIDA, N., HONJO, Y., JOHNSON, K. R., WHEELOCK, M. J. & TAKEICHI, M. 1996. The catenin/cadherin adhesion system is localized in synaptic junctions bordering transmitter release zones. *The Journal of Cell Biology,* 135**,** 767-779.
UEMURA, T. 1998. The cadherin superfamily at the synapse: more members, more missions. *Cell,* 93**,** 1095-1098.
VAN ROOIJEN, J. J., VOSKAMP, A. F., KAMERLING, J. P. & VLIEGENTHART, J. F. 1999. Glycosylation sites and site-specific glycosylation in human Tamm-Horsfall glycoprotein. *Glycobiology,* 9**,** 21-30.
VATTER, G. 1998. Diagnosis of Autism in Children with Down syndrome.
VERRILLS, N. M. 2006. Clinical proteomics: present and future prospects. *Clinical Biochemist Reviews,* 27**,** 99.
VIDAL, B. X. A. C., BONVENTRE, J. X. A. V. & HSU, S. 2005. Towards the application of proteomics in renal disease diagnosis. *Clinical Science,* 109**,** 421-430.
VOURC'H, P., MARTIN, I., BONNET-BRILHAULT, F., MAROUILLAT, S., BARTHÉLÉMY, C., MÜH, J. P. & ANDRES, C. 2003. Mutation screening and association study of the UBE2H gene on chromosome 7q32 in autistic disorder. *Psychiatric Genetics,* 13**,** 221-225.
WALTES, R., DUKETIS, E., KNAPP, M., ANNEY, R. J., HUGUET, G., SCHLITT, S., JARCZOK, T. A., SACHSE, M., KÄMPFER, L. M. & KLEINBÖCK, T. 2014. Common variants in genes of the postsynaptic FMRP signalling pathway are risk factors for autism spectrum disorders. *Human genetics***,** 1-12.
WANG, D. G., FAN, J.-B., SIAO, C.-J., BERNO, A., YOUNG, P., SAPOLSKY, R., GHANDOUR, G., PERKINS, N., WINCHESTER, E. & SPENCER, J. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science,* 280**,** 1077-1082.
WANG, K., ZHANG, H., MA, D., BUCAN, M., GLESSNER, J. T., ABRAHAMS, B. S., SALYAKINA, D., IMIELINSKI, M., BRADFIELD, J. P., SLEIMAN, P. M., KIM, C. E., HOU, C., FRACKELTON, E., CHIAVACCI, R., TAKAHASHI, N., SAKURAI, T., RAPPAPORT, E., LAJONCHERE, C. M., MUNSON, J., ESTES, A., KORVATSKA, O., PIVEN, J., SONNENBLICK, L. I., ALVAREZ RETUERTO, A. I., HERMAN, E. I., DONG, H., HUTMAN, T., SIGMAN, M., OZONOFF, S., KLIN, A., OWLEY, T., SWEENEY, J. A., BRUNE, C. W., CANTOR, R. M., BERNIER, R., GILBERT, J. R., CUCCARO, M. L., MCMAHON, W. M., MILLER, J., STATE, M. W., WASSINK, T. H., COON, H., LEVY, S. E., SCHULTZ, R. T., NURNBERGER, J. I., HAINES, J. L., SUTCLIFFE, J. S., COOK, E. H., MINSHEW, N. J., BUXBAUM, J. D., DAWSON, G., GRANT, S. F., GESCHWIND, D. H., PERICAK-VANCE, M. A., SCHELLENBERG, G. D. & HAKONARSON, H. 2009. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature,* 459**,** 528-533.
WANG, L., JIA, M., YUE, W., TANG, F., QU, M., RUAN, Y., LU, T., ZHANG, H., YAN, H. & LIU, J. 2008. Association of the ENGRAILED 2 (EN2) gene with autism in

Chinese Han population. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* 147**,** 434-438.
WANG, L., LI, F., SUN, W., WU, S., WANG, X., ZHANG, L., ZHENG, D., WANG, J. & GAO, Y. 2006. Concanavalin A-captured Glycoproteins in Healthy Human Urine. *Molecular & Cellular Proteomics,* 5**,** 560-562.
WARKANY, J. & HUBBARD, D. M. 1953. Acrodynia and mercury. *The Journal of Pediatrics,* 42**,** 365-386.
WASSINK, T., PIVEN, J., VIELAND, V., PIETILA, J., GOEDKEN, R., FOLSTEIN, S. & SHEFFIELD, V. 2004. Examination of AVPR1a as an autism susceptibility gene. *Molecular Psychiatry,* 9**,** 968-972.
WASSINK, T. H., PIVEN, J., VIELAND, V. J., HUANG, J., SWIDERSKI, R. E., PIETILA, J., BRAUN, T., BECK, G., FOLSTEIN, S. E. & HAINES, J. L. 2001. Evidence supporting WNT2 as an autism susceptibility gene. *American Journal of Medical Genetics,* 105**,** 406-413.
WEAVING, L. S., ELLAWAY, C. J., GÉCZ, J. & CHRISTODOULOU, J. 2005. Rett syndrome: clinical review and genetic update. *Journal of Medical Genetics,* 42**,** 1-7.
WEIGEL, M. T. & DOWSETT, M. 2010. Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocrine-Related Cancer,* 17**,** R245-R262.
WU, J., CHEN, Y. D. & GU, W. 2010a. Urinary proteomics as a novel tool for biomarker discovery in kidney diseases. *Journal of Zhejiang University Science B,* 11**,** 227-237.
WU, Q. & MANIATIS, T. 1999. A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell,* 97**,** 779-790.
WU, Q. & MANIATIS, T. 2000. Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes. *Proceedings of the National Academy of Sciences,* 97**,** 3124-3129.
WU, Y., LI, Y., LANGE, E. M., CROTEAU-CHONKA, D. C., KUZAWA, C. W., MCDADE, T. W., QIN, L., CUROCICHIN, G., BORJA, J. B., LANGE, L. A., ADAIR, L. S. & MOHLKE, K. L. 2010b. Genome-wide association study for adiponectin levels in Filipino women identifies CDH13 and a novel uncommon haplotype at KNG1–ADIPOQ. *Human Molecular Genetics,* 19**,** 4955-4964.
YAGI, T. & TAKEICHI, M. 2000. Cadherin superfamily genes: functions, genomic organization, and neurologic diversity. *Genes & Development,* 14**,** 1169-1180.
YAP, I. K., ANGLEY, M., VESELKOV, K. A., HOLMES, E., LINDON, J. C. & NICHOLSON, J. K. 2010. Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. *Journal of Proteome Research,* 9**,** 2996-3004.
YASUDA, Y., HASHIMOTO, R., YAMAMORI, H., OHI, K., FUKUMOTO, M., UMEDA-YANO, S., MOHRI, I., ITO, A., TANIIKE, M. & TAKEDA, M. 2011. Gene expression analysis in lymphoblasts derived from patients with autism spectrum disorder. *Molecular Autism,* 2**,** 9.
YU, K. K., CHEUNG, C., CHUA, S. E. & MCALONAN, G. M. 2011. Can Asperger syndrome be distinguished from autism? An anatomic likelihood meta-analysis of MRI studies. *Journal of Psychiatry and Neuroscience,* 36**,** 412-421.

Internet Resources

- Applied Biosystems. Available from: http://www.appliedbiosystems.com/filebuilder> [2010]
- Connecticut Autism Spectrum Resource center. Available from: http://www.ct-asrc.org> [2012]
- Ensemble. Available from: http://www.ensembl.org/Homo_sapiens/Gene/Compara_Tree?db=core;g=ENSG00000040731;r=5:24487209-24645087> [29 August 2011]
- Genetics Home reference, National Institutes of health. Available from: http://ghr.nlm.nih.gov> [12 July 2011]
- Graphpad, P value calculator software. 2011 available from: http://graphpad.com/quickcalcs/pValue2/\> [October 2011]
- Mass Spectrometry based proteomics, Proteomics Shared Research. Available from: http://www.ohsu.edu/xd/research/research-cores/proteomics/educational-links/upload/Mass_Spectrometry_Based_Proteomics.pdf.> [5 September 2011]
- Nextprot beta. Available from : http://www.nextprot.org/db/entry/NX_P01834> [14 December 2012]
- Smith, K., Genetic polymorphism and SNPS's. Available from: http://www.cs.mcgill.ca/~kaleigh/compbio/snp/snp_summary.html> [19 February 2002]
- The center for brainhealth, The University of Texas at Dallas News Center. Available from: http://www.brainhealth.utdallas.edu/blog_page/schizophrenia-the-healing-potential-of-knowing-how-to-navigate-social-relat> [28 November 2012]
- The Gene gateway workbook, Oak ridge national laboratory. Available from: http://web.ornl.gov/sci/techresources/Human_Genome/posters/chromosome/ggworkbook1.pdf> [February 2011]
- The gene ontology, Amigo. Available from: http://amigo.geneontology.org/cgi-bin/amigo/search.cgi?gptype=protein&speciesdb=all&taxid=9606&ont=all&evcode=all&query=cdh10&action=filter&search_constraint=gp&session_id=1369amigo1386718816> [1 November 2011]
- The human genome project: Sequencing the Future, 2003. Available from: http://www.ornl.gov/sci/techresources/Human_Genome/faq/SNPs.shtml
- Wilson, L., 'The autism epidemic and naturals solutions", The center for development. Available from: http://drlwilson.com/Articles/AUTISM.htm> [2011]

www.expasy.org
http://www.genome.gov/gwastudies
www.appliedbiosystems.com
www.ensemble.org
www.nextprot.org
www.swissprot.org
www.uniprot.org)
www.ncbi.nlm.nih.gov/snp

UNIVERSITY of the
WESTERN CAPE

Appendix I

Navrae
Enquiries
IMibuzo
Dr RS Cornelissen

Telefoon
Telephone
IFoni
(021) 467-2286

Faks
Fax
IFeksi
(021) 425-7445

Verwysing
Reference
ISalathiso
20060105-0011

Wes-Kaap Onderwysdepartement

Western Cape Education Department

ISebe leMfundo leNtshona Koloni

Dr Zainunisha Arieff
Department of Biotechnology
University of the Western Cape
Private Bag X17
BELLVILLE
7535

Dear Dr Z. Arieff

RESEARCH PROPOSAL: GENETICS OF AUTISM IN SOUTH AFRICA.

Your application to conduct the above-mentioned research in schools in the Western Cape has been approved subject to the following conditions:

1. Principals, educators and learners are under no obligation to assist you in your investigation.
2. Principals, educators, learners and schools should not be identifiable in any way from the results of the investigation.
3. You make all the arrangements concerning your investigation.
4. Educators' programmes are not to be interrupted.
5. The Study is to be conducted from **1st July 2010 to 30th September 2012**.
6. No research can be conducted during the fourth term as schools are preparing and finalizing syllabi for examinations (October to December).
7. Should you wish to extend the period of your survey, please contact Dr R. Cornelissen at the contact numbers above quoting the reference number.
8. A photocopy of this letter is submitted to the Principal where the intended research is to be conducted.
9. Your research will be limited to the list of schools as submitted to the Western Cape Education Department.
10. A brief summary of the content, findings and recommendations is provided to the Director: Education Research.
11. The Department receives a copy of the completed report/dissertation/thesis addressed to:

    The Director: Education Research
    Western Cape Education Department
    Private Bag X9114
    CAPE TOWN
    8000

We wish you success in your research.

Kind regards.

Signed: Ronald S. Cornelissen
for: **HEAD: EDUCATION**
DATE: 10th August 2007

UNIVERSITY of the
WESTERN CAPE

OFFICE OF THE DEAN
DEPARTMENT OF RESEARCH DEVELOPMENT

08 May 2013

To Whom It May Concern

I hereby certify that the Senate Research Committee of the University of the Western Cape, at their meeting held in July 2010, has approved the methodology and ethics for the following research by: Dr Z Arieff (Biotechnology)

Research Project: Genetics of Autism (*ongoing project*).

Registration no: 05/9/33

Any amendments, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

The Committee must be informed of any serious adverse event and/or termination of the study.

Ms Patricia Josias
Research Ethics Committee Officer
University of the Western Cape

Private Bag X17, Bellville 7535, South Africa
T: +27 21 959 2988/2948 . F: +27 21 959 3170
E: pjosias@uwc.ac.za
www.uwc.ac.za

UNIVERSITY of the
WESTERN CAPE

OFFICE OF THE DEAN
DEPARTMENT OF RESEARCH DEVELOPMENT

02 September 2013

To Whom It May Concern

I hereby certify that the Senate Research Committee of the University of the Western Cape approved the methodology and ethics of the following research project by Prof Z Arieff (Biotechnology)

Research Project: Genetics of Autism in South Africa.

Registration no: 13/6/29

Any amendments, extension or other modifications to the protocol must be submitted to the Ethics Committee for approval.

The Committee must be informed of any serious adverse event and/or termination of the study.

Ms Patricia Josias
Research Ethics Committee Officer
University of the Western Cape

Private Bag X17, Bellville 7535, South Africa
T: +27 21 959 2988/2948 . F: +27 21 959 3170
E: pjosias@uwc.ac.za
www.uwc.ac.za

Private Bag X17, Bellville, 7535
South Africa
Tel: +27 (0) 21 959 2215
Fax: +27 (0) 21 959 3505
Website: www.uwc.ac.za

FACULTY OF NATURAL SCIENCES

Department of Biotechnology

March 2012

Parents/ Guardians
Nolunthando School
Lansdowne Ext., Site B
Khayelitsha
Tel: (021) 361 1160

Dear Parents/ Guardians

I am a Masters student in the Biotechnology department at the University of the Western Cape. As part of my studies, I need clinical information and biological material. The biological material will be obtained from the scholars with the aid of a teacher by collecting urine in a plastic container during the morning. It is painless and non intrusive procedure and will only take a few minutes.

The project has been approved by the principal and the UWC Ethics Committee.

Two students from our research laboratory will be coming to the school for collections. Nolunthando School has previously participated in this study earlier last year, 2011 and we would like to take this opportunity to say thank you. Being the parent/guardian of the child, I seek your permission for the participation of your child in the study. The consent form follows.

I the undersigned parent/guardian of……in grade……..hereby

Give permission	Do NOT give permission

for my son/daughter to take part in the research project organized by the university of the Western Cape.

PLEASE MAKE SURE TO CROSS OUT ONE

Yours in anticipation

Dr. Z Arieff

Ms. Firzana October (M.Sc Student), Human Genetics Laboratory
Department of Biotechnology, University of the Western Cape
Cell 072 572 4872, email: firzanao@gmail.com

Private Bag X17, Bellville, 7535
South Africa
Tel: +27 (0) 21 959 2215
Fax: +27 (0) 21 959 3505
Website: www.uwc.ac.za

FACULTY OF NATURAL SCIENCES

Department of Biotechnology

March 2012

Parents/ Guardians
Lantana School
Lentegeur
Mitchell's plain
Tel: (021) 371 1142

Dear Parents/ Guardians

I am a Masters student in the Biotechnology department at the University of the Western Cape. As part of my studies, I need clinical information and biological material. The biological material will be obtained from the scholars with the aid of a teacher by collecting urine in a plastic container during the morning. It is painless and non intrusive procedure and will only take a few minutes.

The project has been approved by the principal and the UWC Ethics Committee.

Two students from our research laboratory will be coming to the school for collections. Lantana School has previously participated in this study earlier last year, 2011 and we would like to take this opportunity to say thank you. Being the parent/guardian of the child, I seek your permission for the participation of your child in the study. The consent form follows.

I the undersigned parent/guardian of……in grade……..hereby

Give permission	Do NOT give permission

for my son/daughter to take part in the research project organized by the university of the Western Cape.

PLEASE MAKE SURE TO CROSS OUT ONE

Yours in anticipation

Dr. Z Arieff

Ms. Firzana October (M.Sc Student), Human Genetics Laboratory
Department of Biotechnology, University of the Western Cape
Cell 072 572 4872, email: 2319597@uwc.ac.za

Private Bag X17, Bellville, 7535
South Africa
Tel: +27 (0) 21 959 2215
Fax: +27 (0) 21 959 3505
Website: www.uwc.ac.za

FACULTY OF NATURAL SCIENCES

DNA CONSENT FORM

Clearly print

Full name of child……..
Date of birth ……
Grade ……..
School…….
Sibling…….

Please indicate with a cross on the slip, the option which you have chosen.

Option 1: I hereby give permission for my child to be included in this research. ☐

Option 2: I would like to have more information. ☐

Option 3 I do not give permission for participation in this research. ☐

If you have chosen option 1, please read and fill in the form below:

1. I, _____, consent to the use of my/ genetic material in the study outlined above.

2. I understand that the genetic material for analysis is to be obtained from a sample of the cheek cells I'm about to donate.

3. I understand that the sample has been assigned a unique identification number and that there will be no link between my child's name and the unique identification number.

4 Since the sample has been collected anonymously it cannot be withdrawn from the study.

5 The sample will be stored indefinitely.

6 The results of the project will be published in a scientific journal.

7. The analysis procedure only provides information on variable genetic elements on specific chromosomes and cannot determine the complete genetic makeup of an individual.

8. At no stage will the sample provided be used for anything other than the analysis of biomolecules involved in autism.

9. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

   Parent's/Guardian's signature: _____

Dr Arieff
Dr Zainunisha Arieff
Senior Lecturer
Dept of Biotechnology
Email: zarieff@uwc.ac.za
Tel: (021) 9592214/5

Appendix II

Mass Spectrometry analysis

Method used By DR. Z. Macdonald (Molecular and Cell Biology Laboratory, University of Cape Town).

After in gel digestion extracted peptides were dried under vacuum. Peptides were re-solubilized in 5 % acetonitrile, 0.1% formic acid loading buffer. 5 µL was injected onto an LC/MS system consisting of an 1100 Series liquid chromatograph, HPLC-Chip Cube MS interface, and an Accurate Mas 6530 Q-Tof mass spectrometer (all AgilentTechnologies). The system was equipped with an HPLC-Chip (Agilent Technologies) that incorporated a 40 nL enrichment column and a 43 mm x 75 µm analytical column packed with ZORBAX 300SB-C18 5 µm particles. Peptides were loaded onto the enrichment column with 97% solvent A (water with 0.1% formic acid) and 3% B (90% acetonitrile with 0.1% formic acid) at 4 µL/min. They were then eluted with a gradient from 3% B to 45% B in 20 minutes, followed by a steep gradient to 80% B in 5 minutes at a flow rate of 0.3 µL/min. The total run time, including column reconditioning, was 35 minutes.

For protein ID the SwissProt database was searched using the Agilent Spectrum Mill software. Search parameters included a static modification on cysteine residues, a 60% minimum matched peak intensity, ±20 ppm mass tolerance on precursor ions and ±0.05 Da mass tolerance on product ions and two missed tryptic cleavages. A minimum of two high confidence unique peptide sequence matches were required for confirmed protein identification.