SINGLE NUCLEOTIDE POLYMORPHISM ASSOCIATION STUDIES OF *ABCA13* AND *ABHD11* GENES AND THE BIOINFORMATICS ANALYSIS OF THE AUTISM CANDIDATE GENES LOCALIZED ON CHROMOSOME 7

Muneera Davids

UNIVERSITY of the
WESTERN CAPE

2431489

A thesis submitted in fulfilment of the requirements for the degree of MSc in the Department of Biotechnology, University of the Western Cape

Supervisor: Dr Z Arieff

Co-supervisor: Dr M Kaur

Co-supervisor: Dr A Pretorius

DECLARATION

By submitting this thesis/dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (safe to the extent explicitly otherwise stated), that reproduction and publication thereof by the University of the Western Cape will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Muneera Davids **November 2015**

Signed…….

Summary

It has been over 50 years since Dr Leo Kanner first wrote and published his paper applying the term autism to a group of children who were self-absorbed and who had severe social, communication and behavioural problems. Since then, research into autism and subsequently, ASDs, has increased by significant leaps and bounds. The vaccination argument has been debunked, but not completely eradicated from societal thinking, and scientists and psychologists alike agree that both "nature and nurture" affect an individual's predisposition to having autism. This work will look at some of the advances made in autism research, demonstrating this by using traditional genetic techniques, as well as more modern, computerized procedures. Furthermore, the use of investigative cohort studies look at the prevalence of autism in the South African perspective, with the aim of increasing genetic data relevant to this region.

Abstract

Autism, Aspergers Syndrome and Pervasive Developmental Delay-Not Otherwise Specified (PDD-NOS), among others, fall under an umbrella of disorders known as Autism Spectrum Disorder. Twin studies show that autism is a highly heritable disorder. More than 100 genes have been implicated in the aetiology of autism, each of which is involved in numerous biological processes and a variety of molecular interactions. William-Beuren syndrome is a multisystem developmental disorder caused by the deletion of contiguous genes at the 7q11.23 position. The aims of this study were (i) to genotype three SNPs (rs10279013, rs2293484 and rs17060) in the *ABHD11* and *ABCA13* genes, respectively, using Taqman® SNP Genotyping assays to detect association with autism in three distinct South African (SA) ethnic groups (Black, Caucasian and Mixed), and (ii) to ascertain common pathways or regulating transcription factors for genes on chromosome 7 that may attribute to it being an "autism hotspot". Chapter 3 objectives were to identify potential candidate genes using STRING analysis and the GeneCards database. The Taqman® study indicated significant association for SNP rs2293484 in the South African Caucasian group, as well as for the G allele in the South African Mixed group, where $p<0.001$. STRING analysis yielded 2 new candidate genes, *FZD1* and *FZD9*. It was also found that the *Wnt* pathway in mammals plays a significant role in both ASDs and cancer, and there is a definite link between genes regulating cancer, and genes implicated in autism. The study provides evidence for not only the association of the investigated SNP in a South African population, but also provides evidence for the co-morbidity of several neurological and psychological disorders such as depression and bipolar disorder with autism

Keywords: *autism, bioinformatics, candidate genes, chromosome 7, ABCA13, ABHD11, FZD1, FZD9, SNPs, genotyping , William-Beuren Syndrome, transcription factors*

ACKNOWLEDGEMENTS

Firstly, I would like to thank my Creator for granting me the opportunity, the strength and the knowledge to carry out my research to the best of my ability. To my parents, brother and extended family for their endless support and encouragement during this time, thank you.

A special thanks and appreciation goes to the principal, staff, learners and parents associated with the following schools:

Alpha School – Cape Town

Vera School for Learner's with Autism – Cape Town

Noluthando School for the Deaf – Cape Town

UNICA School for Autism – Pretoria

Browns' School – Kwa Zulu Natal

Kenridge Primary School – Cape Town

Labiance Primary School – Cape Town

Walter Teka High School – Cape Town

Thank you to Dr Craig Kinnear and Liezl Bloem at the Stellenbosch University Medical Campus for the use of their resources and equipment.

I would like to thank Dr Jyoti Rajan-Sharma for her support during this project. Her feedback and input provided excellent support.

Thank you to Dr Ashley Pretorius and Dr Mandeep Kaur, who have taught me so many new things about science. Your vast knowledge base and valuable input provided a refreshing new take on genetics that I would not have discovered otherwise.

A HUGE thank you to the new friends I have made, and the old friends who have remained by my side while writing this thesis; you have provided me with countless hours of support, laughter, procrastination, snacks and encouragement. To my URL friends from Twitter, thank you for the encouragement and the laughs; some of you I have never met, but I have found an online community of support and motivation on my timeline.

Last, but certainly not least, I would like to thank my supervisor, Dr Zainu Arieff, for her endless support, encouragement and patience. She has been understanding during a very difficult time for me, mentored me and reminded me to remain curious about my world while always caring for my fellow human being. Dr A., you are an amazing woman. Your experience of many years, your laid back but proactive attitude, and your always positive outlook has been an inspiration to me for all the years I have been your student. Thank you, thank you, thank you

TABLE OF CONTENTS

LIST OF TABLES & FIGURES

LIST OF FIGURES

LIST OF TABLES

UNIVERSITY of the
WESTERN CAPE

CHAPTER 1: INDEX

Chapter 1: LITERATURE REVIEW

CHAPTER 1: LITERATURE REVIEW

1.1. A history of autism

Eugene Bleuler first used the term autism around 1911 to describe patients with schizophrenia. The autistic disorder, or classic autism was first described over sixty years ago by Dr Leo Kanner of the Johns Hopkins University in the U.S.A. Around the same a time, a Swiss psychiatrist, Dr Hans Asperger described a milder form of the disorder, characterised by higher cognitive abilities and better language function (Wolff, 2004).

Autism and schizophrenia were linked in many researchers' minds until the 1960's. It was only during this time, after Dr. Kanner's descriptions that medical professionals would make the differentiation between the two disorders, especially in children (Schanen, 2006). Early treatment for autism was largely focused on medication such as LSD, as well as electric shock therapy and behavioural change techniques that were based on pain and punishment. Early intervention and unique, specialized treatments are the basis for modern-day treatment of autism and Autism Spectrum Disorders (ASDs) (Autism Speaks, 2013). Depending on the severity of the disorder in the child, treatment is behaviour-based, medication-based, or both. Many individuals with autism have additional conditions and disorders such as sleeping disorders, aggression, seizures, epilepsy and gastrointestinal distress, as well as Down's syndrome, Tourette's syndrome, Fragile-X syndrome (Piven and Palmer, 1999 and Schanen 2006). The latter disorders mentioned all have clearly defined genetic links as their causes, and this has lead researchers to look at autism as a genetic disorder, in order to find a causative factor.

While there is now considerable evidence to support genetic factors in the aetiology of autism, there is still no definitive causative factor. Since the landmark twin study by Folstein and

Rutter (1977) indicating the strong genetic link between the inheritance of autism between identical and fraternal twins, researchers have made leaps and bounds in the study of genetic factors influencing autism, as well as co-morbid disorders in individuals. Several 'candidate genes' have been identified, most of them being found to have a positive association to autism in specific population groups. Extensive single nucleotide polymorphism (SNP) studies have been done and are underway (Prasad *et al.*, 2013; Polumbo *et al.*, 2013; Arieff *et al.*, 2010) in order to glean the extent of DNA mutations present in specific population groups with regards to autism (Sharma *et al.*, 2013), as well as microarray studies (Gameeldien 2009) to research the prevalence of over- and under-expressed genes in a specific population.

Computational analysis and bioinformatics in genetic studies has also proven to be an invaluable research tool. Data-intensive research has been made simpler in terms of analysis and database maintenance (Perez-Iratxeta *et al.*, 2007). One of the important and extremely useful roles these types of programs play in biological studies is that they are able to organize genes and proteins within their larger biological context. This includes identifying common pathways and protein interactions of various genes and proteins in different biological functions (Romero *et al.*, 2004).

The study of the genetics of autism worldwide has yielded an overwhelming amount of information, particularly pertaining to candidate genes being identified for their various roles in the vastly differentiated symptoms of autism and ASDs (ASDs) (Lord, 2011). Bioinformatics and computational analysis allows researchers to streamline all of this information into a single database of their choice and further investigate the interactions of proteins and genes at the molecular pathway level. Pretorius *et al*. (2011) describes a detailed methodology for the functional analysis and characterization of the human Retinoblastoma Binding Protein 6 (*RBBP6)* promoters using a combination of molecular biology as well as computational analysis. Their study found that the *RBBP6* gene, although already linked to

apoptosis and cancer, also showed links with diabetes and Alzheimer's disease.

Zeidán-Chuliá *et al*. (2013) used a systematic computer-based approach to identify the major biological processes and cellular components molecular functions that are crucial to autism, while at the same time ascertaining the most essential contributions (genetic and/or environmental) *in silico* (Zeidán-Chuliá *et al*., 2013). Their research identified that the genes *Ras-Related C3 Botulinum Toxin Substrate 1 (RAC1)* and the *Rhodopsin (RHO)* family of GTPases could play a critical role in the neuropathological events associated with autism.

The structure of this work will be based on both a molecular and computational approach for looking at implicated genes in autism and ASDs in the South African population. Two genes have been identified by a previous study (Gameeldien, 2009) as important under- and over-expressed in the population, and this study will aim to look at the allelic frequency of SNPs in the population as well as the computational analysis of why chromosome 7 has been implicated as an 'autism hotspot' by looking at the molecular pathways within genes implicated in autism.

1.2. Autism

Autism, Aspergers Syndrome and Pervasive Developmental Delay-Not Otherwise Specified (PDD-NOS), among others, fall under an umbrella of disorders known as Autism Spectrum Disorder (Volkmar and Rutter, 1995; Volkmar *et al*., 1997). ASDs are neurodevelopmental disorders that are characterized and diagnosed by very specific behaviours - impaired social interaction, impaired communication and restrictive or repetitive interests or behaviours (Benayed *et al*., 2004). The autistic phenotype is a behaviour-based syndrome described by the DSM-V (American Psychiatric Association, 2014). The DSM-V (Diagnostic and Statistical Manual of Mental Disorders) provides a classification of mental disorder, written

in common language with standard criteria. It is widely used in the United States and also around the world. It is utilized by a number of various health professionals, as well as persons involved in the health industry (pharmaceutics, researchers, health insurance companies, policy makers) for diagnosis and decisions regarding mental health issues.

Autistic disorder is diagnosed with the presence of six symptoms in total; these include at least two symptoms of impairment in social interaction, at least one symptom of impaired communication skills, and at least one symptom of restrictive and repetitive behaviour, illustrated in figure 1.1. (Lamb *et al*., 2000). Symptoms are typically displayed from 3 years of age, and may persist into adulthood.

Figure 1.1: An example of the typical obsessive behaviour as a symptom of autism. (http://sandrarose.com/2012/02/medical-minute-autism/)

Autism has been classified as the most frequently occurring of all childhood neurological disorders (Baio, 2012). The number of people known to have autism has increased dramatically since the 1980's. Whether this is due to changes and advances in diagnostic practices or whether actual prevalence has increased is still a topic of intense scrutiny. In South Africa, there is currently no accessible database system that holds information as to the number of children affected in the population. One study indicates that the current global

incidence for autism and ASDs is 1 in 68 (Centre for Disease Control, 2015); affecting four times more males than females (Kopetz and Endowed, 2012).

There is no conclusive evidence to point to the male-female ratio discrepancy, but several theories have been put forward and active investigative studies are underway to find answers. One school of thought suggests that it is the inherent behavioural differences in males and females that may contribute to the high male:female ratio. Females are largely, by nature, sociable. So it is a theory that because of this, female children will seem better adjusted; socially, than their male peers. Girls are often diagnosed with ASD at much later ages than boys. A more complex theory is that autism is X-linked. Evidence is emerging from studies in both humans and mice for a general influence on intelligence. There are a large number of mental disorders that are X-linked. In addition to this, studies also show that X-linked genes have specific effects on social-cognition and emotional development. It is proposed that because females have an extra X chromosome, it serves as a genetic 'protection' against the effects that have been attributed to X-linked disorders (Skuse 2005).

In another study by Ptaff *et al*. (2011) they postulated that due to the higher testosterone levels (TST) in boys, higher amounts of arousal-related inputs to the amygdala sensitizes genetically vulnerable infants to very early stresses. Arousal-related inputs are known to prime amygdaloidal mechanisms for fear and anxiety, which may result in social avoidance. In contrast to boys, girls do not possess high levels of TST-facilitated arousal-causing inputs to the amygdala, and in addition to this, they have protection provided by oxytocin and its receptors.

According to a recent study (Yang. *et al*., 2012) there may be two different neural pathways involved in boys and girls. They tested the genes contained in *denovo* copy number variations (CNVs) from boys and girls with ASD by a web database. Significantly different neural pathways were observed in both boys and girls, which was indicative of the role that different

combinations of genes play in these pathways, and are implicated in the number of girl affected, versus the number of boys.

Though there has been an increasing incidence of autism, the causes of ASDs are still unknown; general consensus is that it is thought to be caused by a combination of both environmental and genetic factors. The study of identical and fraternal twins supports the argument for the heritability factor as aetiology for ASD. Epidemiological studies of autism report difficult pregnancies, viral infections, metabolic conditions and the ever-popular vaccine argument as causes of autistic disorder (St-Hilaire *et al*., 2012). It is therefore safe to draw the conclusion that both environment and genetics may contribute to the manifestation of autism in children (Rodier and Hyman, 1998).

1.3. Environmental factors

During the 1950s and 1960s it was well established that many infectious diseases were detrimental to the developing brain. Many of the causative factors of other diseases that caused mental retardation were then investigated as causative agents for autism. Pre-natal exposure to the rubella virus (Libbey *et al*., 2005) can cause birth defects including mental retardation, blindness, deafness and autism. Studies of large samples have shown the increased rate of autism in children who had been exposed to the rubella virus *in utero* (Sigman *et al*., 2006). A significantly controversial topic in the autism community is whether or not vaccines, particularly the measles, mumps and rubella (MMR) vaccine, can be blamed as a cause for autism. Up until 2001, sodium ethylmercurithiosalicylate, commonly known as thiomersal, was a mercury-based preservative that was frequently used in MMR vaccines (Kaye *et al*., 2001). Concern regarding the use of thiomersal was based on the fact that organic mercury has a negative effect on the developing brain and that children receiving

vaccinations were thus affected. There is currently no conclusive scientific evidence that positively links the effects of thiomersal to autism (Nelson and Bauman 2003; Peltola *et al*., 1998). Despite the scientific evidence for MMR vaccinations not being a cause for autism, there are still an alarming number of people who refuse to have their children vaccinated. This "anti-vaxxer" movement is so large that in 2014, a measles outbreak occurred in the United States, thought to have originated from a Disneyland theme park (Centre for Disease Control, 2014). In 2014, a meta-analysis study (a cohort study involving 1 256 407 experimental subjects and 9 920 control subjects) was published that showed no evidence for a link between autism and vaccines. The study showed statistical significance for no relationship between vaccines and autism, no relationship between autism/ASDs MMR, no relationship between autism/ASD and thiomersal and no relationship between autism/ASDs and mercury (Taylor *et al*., 2014).

Other environmental factors that may influence the development of autism in an individual include the use of addictive substances such as alcohol and drugs during early gestation. Autism has been linked to foetal alcohol syndrome and disorders caused by addictive drugs such as cocaine and methamphetamines (Davis *et al*., 1992) and have consequently shown to affect brain development and function (Baird *et al*., 2003). Mercury exposure from sources such as diet (particularly in fish) and dental amalgam fillings have been implicated as causes of autism (Lawler *et al*., 2004). Methylmercury exposure in humans and the role of selenium as a factor which counteracts methylmercury toxicity and protects against some neurological effects of exposure has yielded information on its potential etiological factors in ASD (Mania *et al*., 2012).

1.4. Genetic factors

It is clear that autism is a highly heritable disorder, but the precise molecular agents responsible for the heritability of the disorder are still unknown. Evidence for genetics as a

contributing agent of the cause of autism is supplied by studies of autistic twins (Folstein and Rutter, 1977). Twin studies provide information on complex genetic interactions in individuals by comparing the genetic similarity of monozygotic (MZ) twins (identical), who share 100 percent of their genes, to dizygotic (DZ) twins (fraternal), who share only 50 percent of their genes. The likelihood that the identical twin of an autistic child would also be autistic is approximately 82 percent, whereas the equivalent rate for DZ twins is around only 10% (Bailey *et al*., 1995, Wall *et al*., 2008). These findings suggest that autism is under a high influence of genetic manipulation, with the involvement of multiple loci.

It remains unclear whether the genetic component is a combination of a few common variants, or rare variants. More than 100 genes have been implicated in the aetiology of autism, each of which is involved in numerous biological processes and a variety of molecular interactions (Newschaffer *et al*., 2002; Phillippe *et al*., 2002). These processes include anything from brain development to speech (Wall *et al*., 2009). More than the specific genes, several chromosomal "hotspots" have been implicated in the aetiolgy of autism, with chromosome 7 (Ashley-Koch *et al*., 1999) currently housing the largest number of genes found to be linked to autism (*table 1*). All other pervasive developmental disorders (PDDs), with the exception of Rett's syndrome, have no monogenic link. Family studies and clinical reports indicate that there is significant overlap among the three most common PDDs; that is, Asperger's disorder, autistic disorder and PDD-NOS. However, the magnitude of the overlap is inconclusive (Constantino and Todd, 2003).

1.5. Autism as a neurological disorder

Studies of brain structure have implicated several aspects of brain development as causative effects in the characteristic symptoms of autism and ASDs. These neuronal disturbances may

lead to a range of abnormalities, including an increase in brain volume. Several studies have shown that the brain volume in children with autism is significantly larger than unaffected individuals (Hardan *et al*., 2000). Neuropathological studies have shown and reported an increase in brain weight in children affected with autism.

Figure 1.2 shows an fMRI scan of the differences activation patterns in the brain in unaffected versus autistic children. Morphometric analysis of the brain has demonstrated enlargement of the hippocampus, and amygdala, as well as excessive growth in the cerebellum during early development (Kellerher and Bear, 2008). However, Harden *et al*. (2000) had reported these brain volumes and weight differences in childhood only, not in adults with autism. It is also well-documented that children with autism have been found to have larger heads than the general population (Aylward *et al*., 2002). Some studies have suggested that the larger brain includes a higher white matter volume, with disrupted grey matter cellular columns, which may contribute to the difficulties with information integration experienced by people with autism (Schultz *et al*., 2000). A study by neuroscientists at Columbia University Medical Centre suggests that children with autism have a surplus of synapses. The researchers say that the excess is due to a slowdown in the normal pruning process that occurs during brain development (Tang, *et al*., 2014). The study further showed that when applied to mouse-models, the pruning defect was found to be linked to a protein called mTOR. The researchers restored normal autophagy and synaptic pruning in the mice by administering rapamycin – a drug that inhibits mTOR. Treatment eliminated the mice's autism-like behaviors. The treatment remained effective even when administered to older mice that had fully developed the autism-like behaviours. The researchers cite this as hopeful evidence that similar treatments might aid in treating autism symptoms.

Figure 1.2: fMRI of autism vs. healthy control activation pattern in the brain (http://www.apa.org/education/k12/brain-function.aspx?item=4#)

1.6. Single Nucleotide Polymorphisms

1.6.1. What are SNPs?

Single nucleotide polymorphisms (SNPs) are the most frequently occurring type of variation in the human genome. This variation occurs when there is a significant frequency (>1%) of a single nucleotide difference in a DNA sequence between members of a species or between paired chromosomes in an individual (*fig.1.3).* SNPs provide geneticists with a powerful tool for the study of human medical genetics (Wang *et al*., 2008)

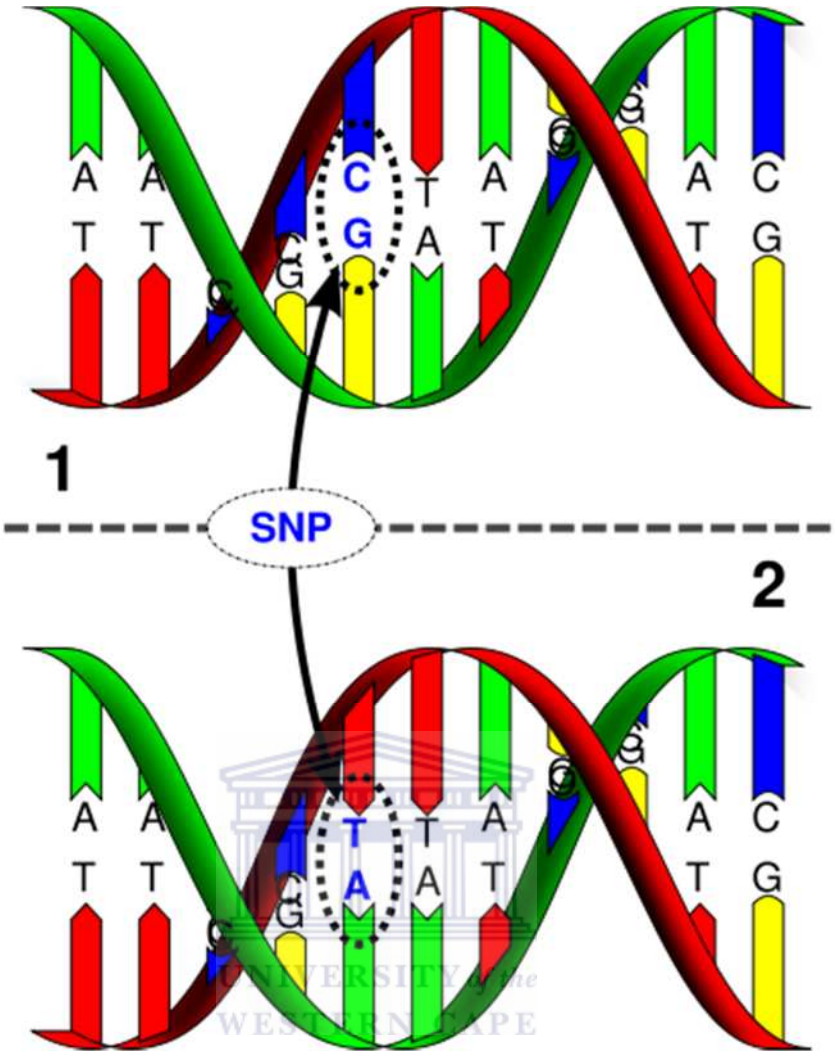

Figure 1.3: The most frequently occurring of all genetic variation, the single nucleotide polymorphism. (http://science.marshall.edu/murraye/341/Images/416px-Dna-SNP_svg.png)

With the complete sequencing of the human genome in 2000, information about SNPs has become more accessible for public and research use (The International SNP Map Working Group, 2001). Databases have been set up on the internet such as the NCBI SNP database, the HapMap database and GeneCards, which offers valuable information such as SNP sequences, whether the SNP is intronic or exonic, what function and disease or disorder it is implicated or involved in and the gene it is found on. Various biotechnology companies offer services such as custom or ready-made SNPs (www.appliedbiosystems.com). To date, more than 1.42 million SNPs have been identified in the human genome, the majority of which are publicly

available. SNPs have helped to identify genes that have been shown to be involved in a myriad of diseases and disorders, ranging from lifestyle diseases such as hypertension and diabetes, to debilitating diseases such as Alzheimer's and autism (Stambrook, 2004).

1.7. Autism and candidate genes

1.7.1. Chromosome 7 as an autism 'hotspot'

Chromosome 7 spans more than 158 million base pairs in the human genome and represents between 5 and 5.5 total DNA in cells. It is likely to contain between 1000 and 1400 genes (Schanen 2006, Berg *et al*., 2007). Many genes located on chromosome 7 have been implicated as causative genes for several disorders and diseases (Lord, 2011). These include non-syndromic deafness, schizophrenia, William-Beuren Syndrome, colour blindness, neurodevelopmental disorders and Tourette syndrome (fig 1.4). Other changes in the structure of chromosome 7 can cause delayed growth and development, mental disorders, characteristic facial features, skeletal abnormalities, delayed speech and medical problems. Several genes on chromosome 7 have been implicated in autism spectrum disorders. Some of these genes have been observed to have common transcription binding sites and are involved in similar molecular functions and pathways. It is not known why so many of these genes are situated on chromosome 7 (*table 1.1*).

Table 1.1: Genes situated on chromosome 7 that have been implicated in Autism

Gene ID (NCBI)	Gene Symbol	Gene Name	Location	Function	Reference
154664	ABCA13*	ATP-binding cassette, sub-family A (ABC1), member 13	7p12.3	In human, the ATP-binding cassette (ABC) family of transmembrane transporters has at least 48 genes and 7 gene subfamilies. This gene is a member of ABC gene subfamily A (ABCA). It is predicted to have two large extracellular domains and two nucleotide binding domains as is typical for ABCA proteins. Alternative splice variants have been described but their biological validity has not been demonstrated.	Knight *et al.*, 2009
83451	ABHD11*	abhydrolase domain containing 11	7q11.23	This gene encodes a protein containing an alpha/beta hydrolase fold domain. This gene is deleted in Williams syndrome, a multisystem developmental disorder caused by the deletion of contiguous genes at 7q11.23. Alternatively spliced transcript variants have been described, but their biological validity has not been determined	Merla *et al.*, 2010
26047	CNTNAP2	contactin associated protein-like 2	7q35	This gene encodes a member of the neurexin family which functions in the vertebrate nervous system as cell adhesion molecules and receptors. May play a role in the formation of functional distinct domains critical for saltatory conduction of nerve impulses in myelinated nerve fibers. Seems to demarcate the juxtaparanodal region of the axo-glial junction. This gene has been implicated in multiple neurodevelopmental disorders, including Gilles de la Tourette syndrome, schizophrenia, epilepsy, autism, ADHD and mental retardation.	Penagarikano *et al*., 2011
2020	EN2	engrailed homeobox 2	7q36	Homeobox-containing genes are thought to have a role in controlling development. The human engrailed homologs 1 and 2 encode homeodomain-containing proteins and have been implicated in the control of pattern formation during development of the central nervous system	Wang *et al*., 2007 Choi *et al*., 2011

93986	FOXP2	forkhead box P2	7q31	This gene encodes a member of the forkhead/winged-helix (FOX) family of transcription factors. It is expressed in fetal and adult brain as well as in several other organs such as the lung and gut. The protein product contains a FOX DNA-binding domain and a large polyglutamine tract and is an evolutionarily conserved transcription factor, which may bind directly to approximately 300 to 400 gene promoters in the human genome to regulate the expression of a variety of genes. This gene is required for proper development of speech and language regions of the brain during embryogenesis, and may be involved in a variety of biological pathways and cascades that may ultimately influence language development. Mutations in this gene cause speech-language disorder 1 (SPCH1), also known as autosomal dominant speech and language disorder with orofacial dyspraxia. Multiple alternative transcripts encoding different isoforms have been identified in this gene.	Gong *et al.*, 2004
3198	HOXA1	homeobox A1	7p15.3	In vertebrates, the genes encoding the class of transcription factors called homeobox genes are found in clusters named A, B, C, and D on four separate chromosomes. Expression of these proteins is spatially and temporally regulated during embryonic development. This gene is part of the A cluster on chromosome 7 and encodes a DNA-binding transcription factor which may regulate gene expression, morphogenesis, and differentiation. The encoded protein may be involved in the placement of hindbrain segments in the proper location along the anterior-posterior axis during development. Two transcript variants encoding two different isoforms have been found for this gene, with only one of the isoforms containing the homeodomain region.	Ingram *et al.*, 2000
83943	IMMP2L	IMP2 inner mitochondrial membrane	7q31	Catalyzes the removal of transit peptides required for the targeting of proteins from the mitochondrial matrix, across the inner	Petek *et al.*, 2007

		peptidase-like		membrane, into the inter-membrane space. Known to process the nuclear encoded protein DIABLO.	
5649	RELN	Reelin	7q22	This gene encodes a large secreted extracellular matrix protein thought to control cell-cell interactions critical for cell positioning and neuronal migration during brain development. This protein may be involved in schizophrenia, autism, bipolar disorder, major depression and in migration defects associated with temporal lobe epilepsy. Mutations of this gene are associated with autosomal recessive lissencephaly with cerebellar hypoplasia. Two transcript variants encoding distinct isoforms have been identified for this gene. Other transcript variants have been described but their full length nature has not been determined.	Bonora *et al.*, 2003 Fatemi *et al.*, 2005 Sharma *et al.*, 2012
7328	UBE2H	ubiquitin-conjugating enzyme E2H	7q32	The modification of proteins with ubiquitin is an important cellular mechanism for targeting abnormal or short-lived proteins for degradation. Three alternatively spliced transcript variants have been found for this gene and they encode distinct isoforms.	Vourc'h *et al.*, 2003
7472	WNT2	wingless-type MMTV integration site family member 2	7q31.2	This gene is a member of the **WNT** gene family. The **WNT** gene family consists of structurally related genes which encode secreted signaling proteins. These proteins have been implicated in oncogenesis and in several developmental processes, including regulation of cell fate and patterning during embryogenesis. Alternatively spliced transcript variants have been identified for this gene.	Marui *et al.*, 2010 Lin *et al*., 2012

*genes have not been implicated in Autism, but have been implicated as genes of interest as a result of microarray studies (and the subsequent bioinformatics analysis) that were done for high- and low-functioning affected individuals in a previous study.

1.7.2. *ABCA13* gene

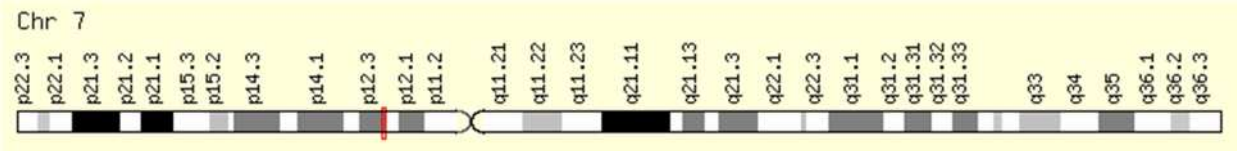

Figure 1.4: Red line indicates the position of the ABCA13 gene on chromosome 7. (http://www.genecards.org/cgi-bin/carddisp.pl?gene=ABCA13)

The adenosine phosphate (ATP)-binding casette, sub-family A (*ABC1*), member 13 gene is the largest known protein of the ABC family. Knight, *et al*. (2009) reported a cytogenic abnormality and rare coding variants in *ABCA13* that may contribute significantly to the risk of schizophrenia, bipolar disorder and major depression. *ABCA13* maps to chromosome 7p12.3, in which multiple rare coding variants were observed *(fig 1.4)*. For this one gene, 1 nonsense and 9 missense mutations, as well as compound heterozygosity/homozygosity was detected in 6 cases in a study sample group (Ovsanna *et. al*., 2006).

Although no significant linkage for *ABCA13* has been found with Autism Spectrum Disorders, there is strong evidence for comorbidity of ASDs with other psychiatric disorders associated with the gene. A study was conducted in by Stahlberg *et al*. (2003) to observe comorbidity of ASD and/or ADHD with schizophrenia, bipolar disorder and major depression in adults with childhood onset ASD and/ADHD. Of those with ASD, 7% tested for bipolar disorder and 7.8% for schizophrenia and other psychiatric disorders. It was also reported that rates of psychiatric disorders in ASDs are high and associated with functional impairment (Stahlberg *et al*., 2003).

1.7.3. *ABHD11* gene

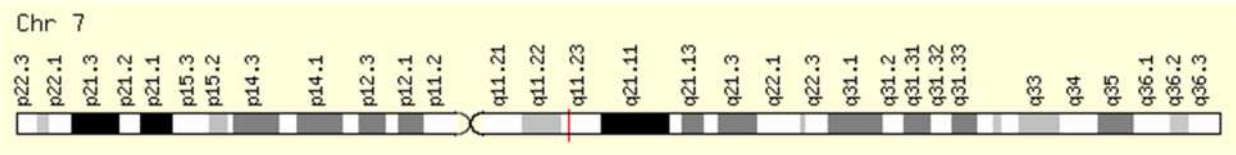

Figure 1.5: Red line indicates the position of the ABHD11 gene on chromosome 7.

(http://www.genecards.org/cgi-bin/carddisp.pl?gene=ABHD11&search=abhd11)

The anhydrolase domain containing 11 gene (*ABDH11*), previously known as the William Beuren Syndrome (WBS) chromosome region 21 gene is located 7q11.23, a region on chromosome 7 that has been implicated in many genetic disorders and diseases.

WBS is a multisystem developmental disorder caused by the deletion of contiguous genes at the 7q11.23 position. Some of these genes include elastin (*ELN)* and *GTF21* implicated as causative genes in spatial visual disorders and CAP-GLY Domain Containing Linker Protein 2 *(CLIP2)*, implicated in behavioural disorders (Chin-See-Chong, *et al*., 2015). WBS is a rare neurodevelopmental disorder that is characterised by an elfin facial appearance, a low nasal bridge, an unusually cheerful demeanour and ease with strangers, as well as developmental delays, strong language skills and cardiovascular problems (Berg *et al*., 2007). The estimated prevalence for WBS is approximately 1:7 500 to 1 in 20 000. The behavioural phenotype of WBS has long been considered to be the polar opposite ASD. New evidence has emerged to show that the co morbidity for autism with WBS is more frequent than previously thought. The research suggests that common WBS deletions can result in a continuum of social communication impairments like excessive talkativeness and over friendliness, as well as the absence of verbal language and poor social relationships (Tordjman *et al*., 2012). Using clinical and molecular examination methods, it was reported that nine French individuals were found to have co-morbidity for ASD and WBS (Tordjman *et al*., 2012). Of the individuals studied, female children were more prone to excessive talkativeness, whereas the

male children in the group were non-verbal. All the children were found to have the WBS deletion, and were placed on the autism spectrum.

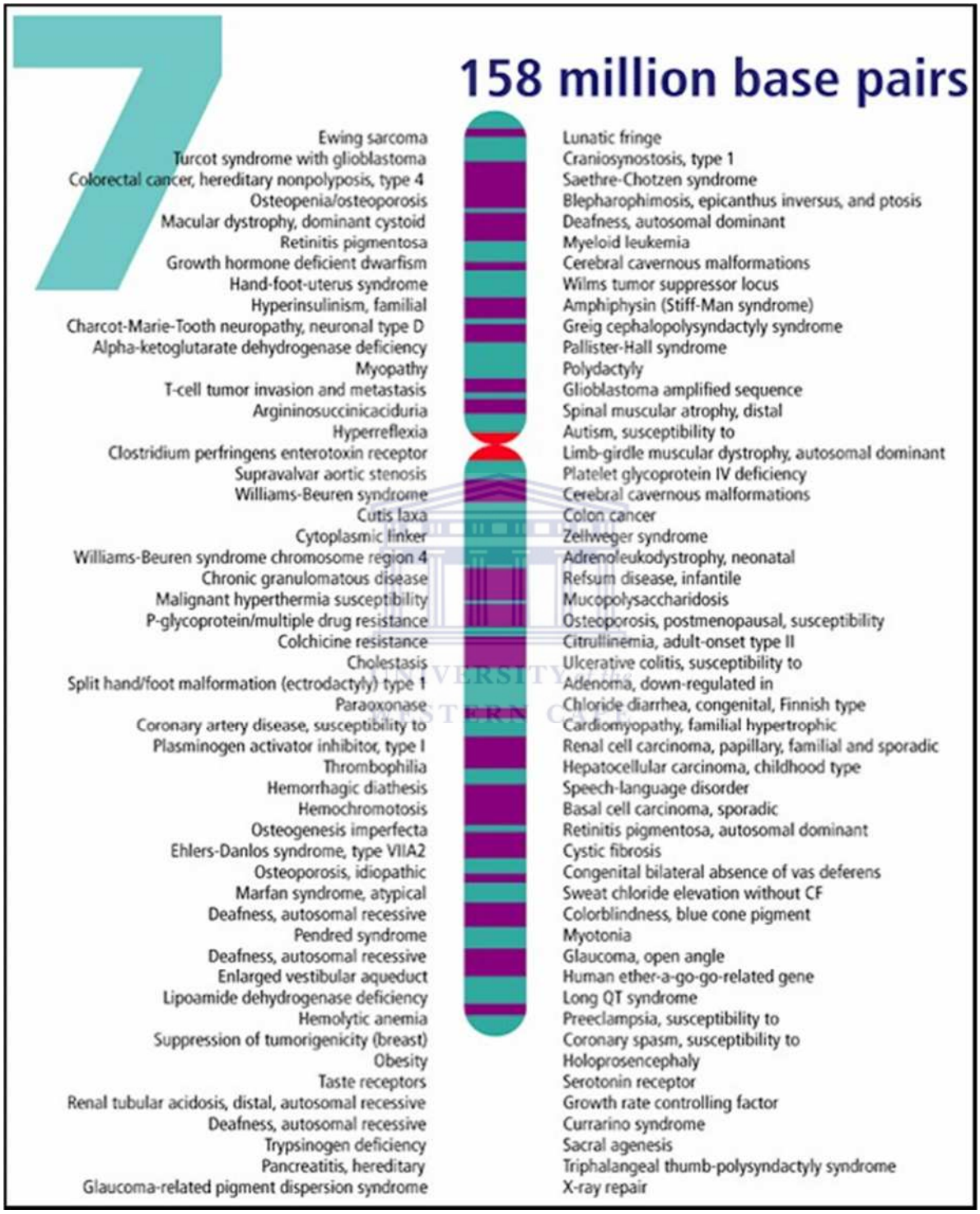

Figure 1.6: Diagram depicting gene-linked diseases and disorders on chromosome 7 (http://www.dnarss.com/Chromosome_7.html)

Hypothesis:

I hypothesised that three SNPs from two genes on chromosome 7 is associated with autism in a South African population.

Aim:

The main aim of this body of work was to genotype SNPs in a South African healthy and affected population

Objectives:

- Genotype rs10279013 from *ABHD11* and compare the allelic distribution in a healthy and affected South African population
- Genotype rs2293484 from *ABHD11* and compare the allelic distribution in a healthy and affected South African population
- Genotype rs17060 from *ABCA13* and compare the allelic distribution in a healthy and affected South African population

CHAPTER 2: INDEX

Chapter 2: MATERIALS & METHODS

CHAPTER 2: MATERIALS & METHODS

2.1. Methods of genotyping

The process of determining differences in the genetic make-up (genotype) of an individual by examining the individual's DNA sequence using biological assays and comparing it to another individual's sequence or a reference sequence is known as genotyping. It reveals the alleles an individual has inherited from their parents. Traditionally genotyping is the use of DNA sequences to define biological populations by use of molecular tools. It does not usually involve defining the genes of an individual.

Traditional genotyping methods include restriction fragment length polymorphism (RFLP) and single-stranded conformational polymorphism (SSCP) (Itoga *et al.*, 2004). Chip-based genotyping assays, combined with knowledge of the patterns of coinheritance of markers (linkage disequilibrium [LD]), have stimulated genome-wide association studies (GWAS) of complex diseases. Recent successes of GWAS in identifying specific genes that affect risk for common diseases are dramatic illustrations of how improved technology can lead to scientific breakthroughs. A key issue in high-throughput genotyping is to choose the appropriate technology for goals and for the stage of the experiment, being cognizant of sample numbers and resources (Edenberg and Liu, 2009).

2.2. Ethical Clearance

The project was approved by the Western Cape Ethics Committee (SR: 5/9/33) of the University of the Western Cape and the Western Cape Education Department. Consent was also obtained from principals of schools and parents for collection of samples for DNA analysis (Appendices 1 – 3) in accordance with the Helsinki Declaration.

Subjects for the study were chosen from three different ethnic groups in the South African population; namely Caucasian (White), African (Black) and Mixed population (Coloured).

The present study group was comprised of 81 autistic black, 82 autistic white and 56 autistic mixed individuals. The number of control individuals was 83, 82 and 56, respectively.

Subjects were between the ages of 4-16, from schools in Cape Town, Kwa-Zulu Natal and Pretoria, both male and female. Samples were collected from children who were diagnosed with Autism Spectrum Disorder(s), as per a psychologist and paediatrician, as well as from healthy children of the same ethnic groups and age group.

2.3. Sample collection

DNA samples were collected by rubbing a sterile swab (Puritan Sterile Polyester Tip Applicators, Manta Forensics, South Africa) on the inside of the cheek for 1 minute and placing it into a labeled sterile 15ml falcon tube (Boeco, Germany). The same procedure was repeated for the other cheek. Tubes were labeled carefully with a sample number, the child's name, date of birth and ethnicity. In addition to this, clinical information was provided by the psychologist or parent/guardian with regards to any complications during the mothers' pregnancy as well as ongoing neurological and physical health issues.

2.4. Isolation of DNA

DNA was extracted from the swabs using the BuccalAmpTM Extraction Kit (Epicentre, USA). The swab tip was placed into a labelled 2 ml eppendorf tube, with the excess of the swab snapped off. 250 ml of the QuickExtract DNA solution 1.0 was added to the tube and incubated at 65°C for 1 minute. They were then mixed by vortexing for 15 seconds and incubated at 98°C for 4 minutes. Thereafter they were again vortexed for 15 seconds. Tubes were stored at -20°C.

2.5. Quantitation of DNA

The concentration of each DNA sample was determined using the NanoDrop® ND-1000 spectrophotometer (NanoDrop Technologies, USA).

2.6. Taqman Genotyping

2.6.1. Methods of SNP detection

With all the information available about the molecular mechanism of SNPs, researchers have developed a variety of detection methods for analysing the presence of SNPs in the human genome. The majority of these techniques are based on basic PCR protocols, with added technology to enhance the efficacy of looking for and analysing. New technology looks at developing genotyping methods that high throughput, accurate and inexpensive (Syvanen, 2001).

2.6.2. Taqman® SNP Genotyping Assay

The Taqman® SNP Genotyping Assay is a rapid fluorophore-based real-time PCR method. It measures the accumulation of amplified product during the exponential phase of the PCR cycle. Two fluorescent probes, VIC and FAM, detect the presence of alleles 1 and 2, respectively. When both probes are fluoresced, this indicates heterozygosity for both alleles (figure 2.1.).

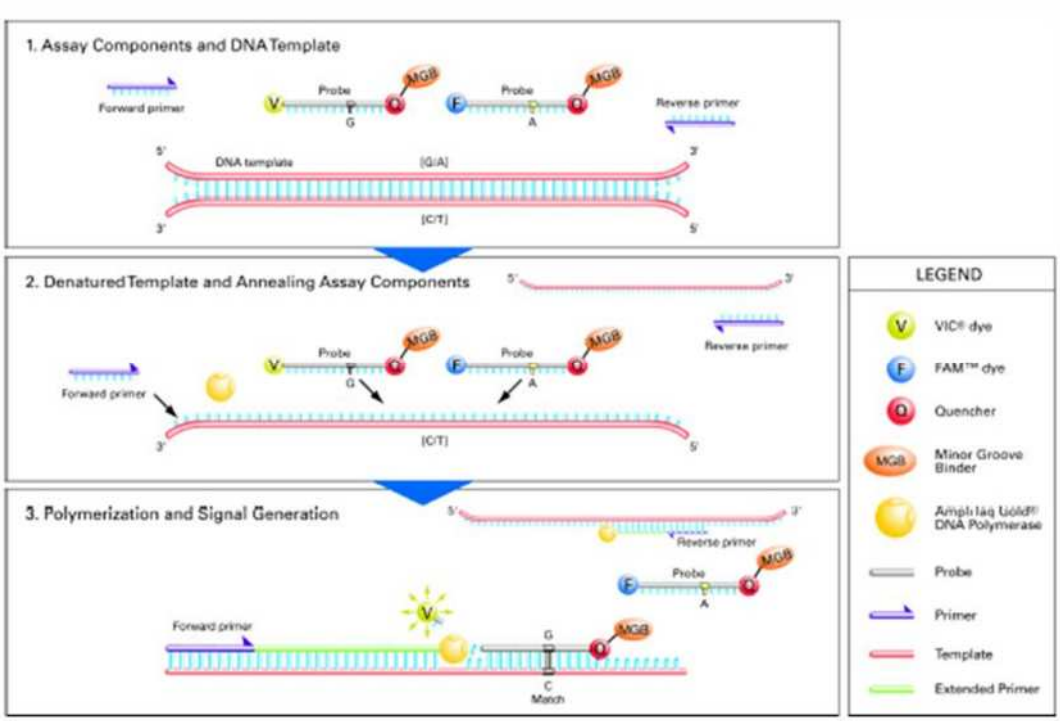

Figure 2.1: The active mechanism of the Taqman SNP Genotyping method (http://www.dnavision.com/taqman-genotyping-assays.php)

SNP genotyping using the Taqman® method has been shown to be wholly successful in the detection of SNPs. It has been used to characterize the APOE haplotypes in Alzheimer's disease (Livak, 1999) and has been used to identify an association between two SNPs and autism in the Chinese Han population (Wang *et al.*, 2008).

2.7. Gene and SNP selection

Two genes from chromosome 7 that have not been previously implicated in Autism but have been linked to other disorders co-morbid to Autism were selected; namely *ABHD11* and *ABCA13* and from these genes, two SNPs were selected per gene for further investigation. These SNPs were rs2293484 (intronic) and rs10279013 (intronic) from *ABHD11* and rs17060 (intronic) from *ABCA13*. These genes were also implicated as genes of interest by microarray

analysis in a previous study (Gameeldien *et al*., 2009). The SNP sequences were blasted against the *Homo sapiens* genome using the BLAST tool at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Primers were then manufactured by Applied Biosystems according to the specifications as indicated by the File Builder application. The Taqman® Custom Genotyping Assay probes (40x) were labelled VIC and FAM to the respective sequences for each SNP:

rs2293484: VIC: 5' – AGGGTCCCCAGGCTGGGGCCTGGGC – 3'

FAM: 5' – AGTAAAGCATGCCCCTCCTACTACC – 3'

rs10279013: VIC: 5' – GAGGAGGCCCCAGACATGGGCGGCT – 3'

FAM: 5' – ACCCTCTGTAGTGCCAGCAGCATGG – 3'

rs17060: VIC: 5' – ATATGGACTCATTTCTTCTTTGCATC – 3'

FAM: 5' – TTATTTCTAGGTTATTTGCAGCCCCG – 3'

2.8. PCR preparation

DNA samples were prepared by diluting samples to a final concentration of 20ng/µl with double SABAX water in 96-deepwell plates (Applied Biosystems, USA), based on the concentrations obtained. 50µl of SABAX water was aliquoted into the first row (row A) of each of the 96-deepwell plates to serve as the negative control. A PCR mastermix for 650 samples was prepared as follows:

1650 µl of Taqman® Universal Genotyping Mastermix (Applied Biosystems, USA) was added to a tube. 165µl of 20x Taqman® SNP Custom Genotyping Assay (Applied Biosystems, USA) was added to the tube, along with 825 µl of SABAX water. The solution

was vortexed for 10 seconds to mix. Samples and the mastermix were then transferred into a 384-well PCR plate (Applied Biosystems, USA) to a final volume of 5 µl, using the Eppendorf epMotion 5070 automated pipetting system (Applied Biosystems, USA).

The dilution of samples for this experiment was the most important step. The Taqman® Genotyping Assay is highly sensitive and is able to detect minute amounts of contamination, which affects the end result of the experiment negatively. Samples were thus prepared under highly sterile conditions. The use of the Epimotion Automated Pipetting system was proven to be time efficient and reduced the risk of human error. Non-template samples were used in regular intervals throughout the experiment to indicate if contamination had occurred. Due to the time lapse in between running some of the SNP assays, fresh dilutions were prepared to ensure accurate results.

2.9. PCR amplification

The Perkin Elmer 9700 PCR System (Applied Biosystems, USA) was used for PCR amplification. PCR parameters were carried out according to the standard PCR protocol provided by the manufacturer. An initial holding step was applied for 10 minutes at 95°C, followed by 40 cycles of the following:

- Initial denaturation for 10 min at 95°C
- Denaturation for 15 sec at 92°C
- Annealing for 1 min at 60°C 40 cycles
- Extension for 1 min at 60°C

40 cycles

2.10. Statistical sample analysis

Allele frequencies for the A and G variants were calculated using the Hardy-Weinberg

equation. Chi square testing was performed to compare the number of alleles of the South African autistic group to the South African control group at a confidence interval of 95%, to compare and determine the significance between different populations (Esau *et al*., 2008). Chi square testing was done using the standard 2-by-2 and 3-by-2 method and validation of these results was done using http://www.vasserstats.net and http://www.danielspoer.com.

Genotype percentages and chi square testing was applied as per the autistic sample group. It is to be noted that these methods of analyses were applied for the genotyping of all subsequent SNPs.

CHAPTER 3: INDEX

CHAPTER 3: RESULTS

3.1. Genotyping results for autistic and healthy populations

Genotyping results were obtained following PCR amplification and subsequent viewing with the Applied Biosystems 7900HT Real-Time PCR system. The raw genotyping data was displayed in a graphical output file (*fig. 3.1*). The data represents autistic and control samples, as well as negative controls (NTCs). Cluster A represents the wild-type homozygous genotype; cluster C represents the mutant homozygous genotype, while cluster B shows the heterozygous alleles. Cluster D represents the negative controls.

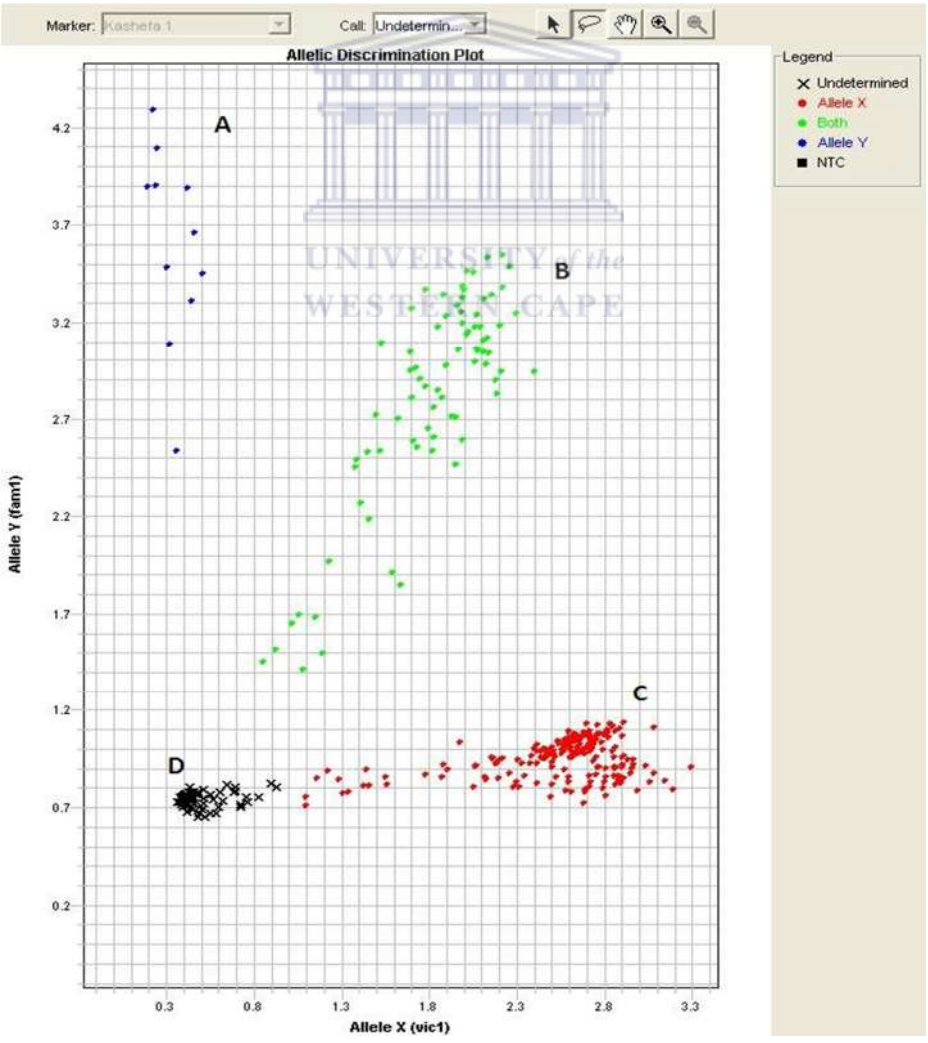

Figure 3.1: Graphical data output for Taqman ® SNP Genotyping Assay

3.2. SNP genotyping analysis

3.2.1. Genotyping analysis for rs10279013

For the autistic sample group, 219 samples were genotyped, with the majority of these displaying the homozygous C allele (*table 3.1*). For the control sample group, 218 samples were genotyped, with the majority of individuals again displaying the CC genotype as dominant. There was no representation of the homozygous T allele for this sample group. The CC genotype in the autistic and the healthy groups in the total SA population was 93% and 73%, respectively. The CT genotype in the autistic and the healthy groups in the total SA population was 7% and 27%, respectively (*figures 3.2 and 3.3*). The comparison of the genotypes for the autistic and healthy groups showed statistical significance, with P=0.00093.

Table 3.1. Genotype and allelic distributions of rs10279013 on *ABHD11* in 3 distinct ethnic groups in the South African autistic and healthy population

Ethnic Group	Number	Genotype (%)			P-value	Allele (%)		P-value
		CC	TT	CT		C Allele	T Allele	
Total South African								
Autistic	219	203 (93)	0	16 (7)	0.00093	422 (96)	16 (4)	0.00014
Control	218	158 (73)	0	60 (27)		376 (86)	60 (14)	
SA Black								
Autistic	81	70 (86)	0	11 (14)	0.70828	151 (93)	11 (7)	0.71843
Control	83	70 (84)	0	13 (16)		153 (92)	13 (8)	
SA Caucasian								
Autistic	82	80 (98)	0	2 (2)	0.000191	162 (99)	2 (1)	0.00016
Control	60	46 (77)	0	14 (23)		106 (88)	14 (12)	
SA Mixed								
Autistic	56	53 (95)	0	3 (5)	0.00016	109 (97)	3 (3)	0.00012
Control	75	42 (56)	0	33 (44)		117 (78)	33 (22)	

*values in red are highly significant, where p-value <0.0001.

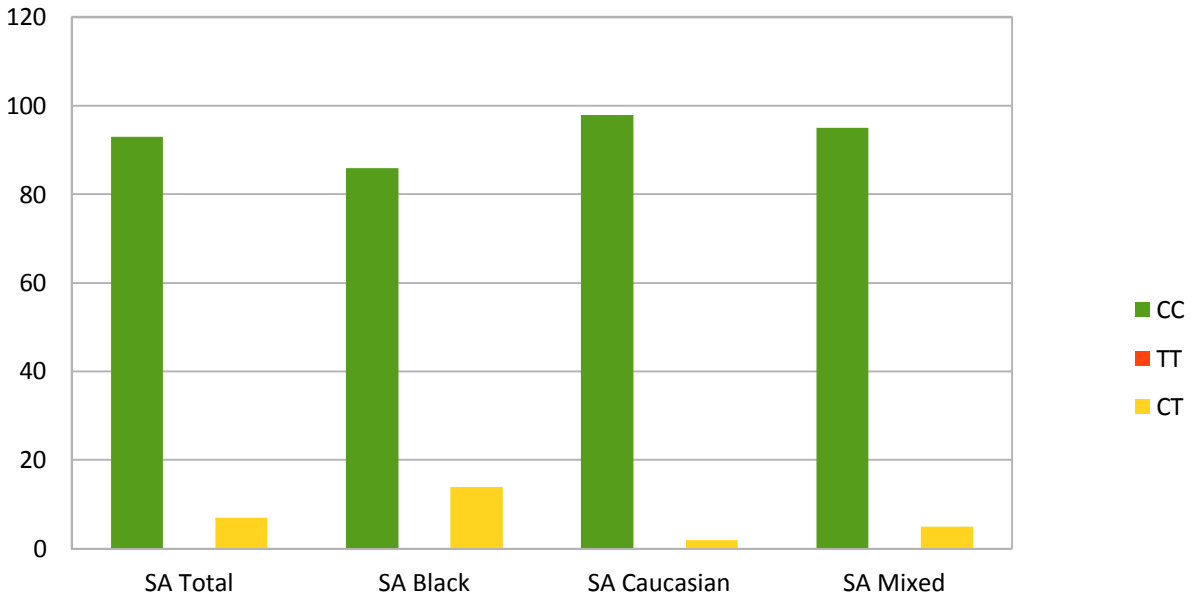

Figure 3.2: Percentage of genotype distributions for rs10279013 on the *ABHD11* gene in an autistic South African population

Figure 3.3: Percentage of genotype distributions for rs10279013 on the *ABHD11* gene in a healthy South African population

The SA Black group displayed high numbers of individuals with the homozygous CC genotype, with the autistic group at 86% and the control group at 84%. The heterozygous CT genotype had a low prevalence of 14% in the autistic group and 16% in the controls. The homozygous TT genotype was not found in any individuals in this group. There was no statistical significance for this allele, at P=0.70828.

In the SA Caucasian group, 98% autistic individuals displayed the CC genotype, with 77% in the control group. The control group showed 23% and 2% in the autistic group for the heterozygous CT genotype. Again, the homozygous TT genotype was not represented in this sample group. The difference in genotypic frequency between the autistic and healthy groups was statistically significant, with **P=0.000191**.

The SA Mixed group showed a high number of individuals in the autistic group displaying the CC genotype, at 95%. In the control group, that number is just over half of the autistic, at 56%. The CT heterozygous genotype is at a very low 5% in the autistic group, with a much higher number of 44% in the controls. Once again, the TT homozygous allele was not represented. With **P=0.00016** and **0.00012** respectively, statistical significance is indicated by the difference between the genotype frequencies for the autistic and healthy group.

3.2.2. Allele frequency analysis for rs10279013

For the C allele, the autistic group displayed 96% of the total South African group having this allele and the control number was at 86%. In contrast, numbers were very low for the T allele, with 4% and 14% in the autistic and control samples, respectively (*table 3.1 and figures 3.4 and 3.5*). However, the differences between the autistic groups compared to the healthy groups was highly significant, where **P=0.00014**.

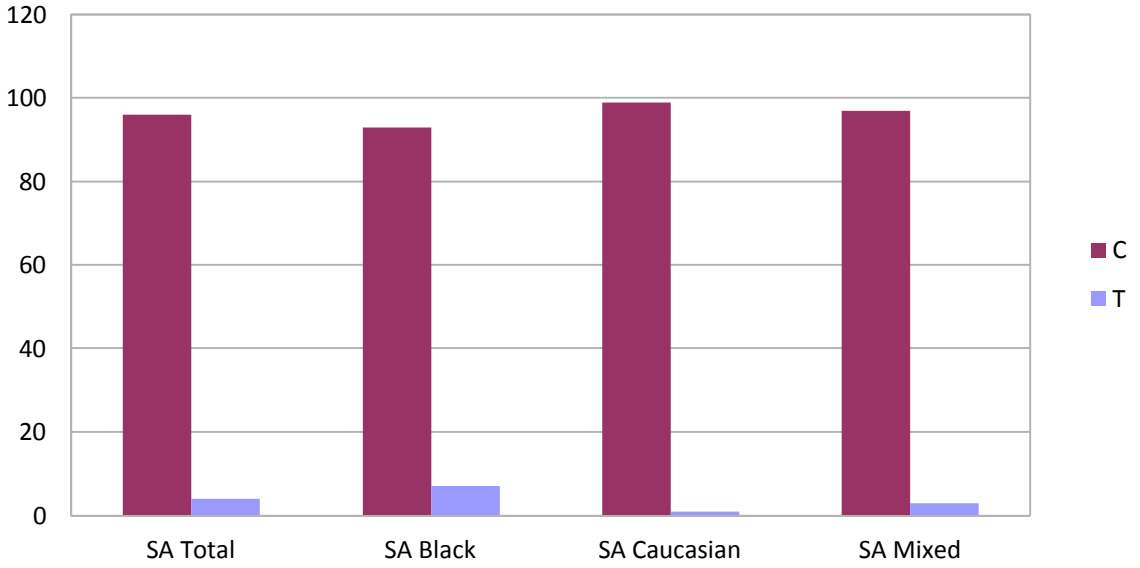

Figure 3.4: Percentage of allelic distributions for rs10279013 on the *ABDH11* gene in an autistic South African population.

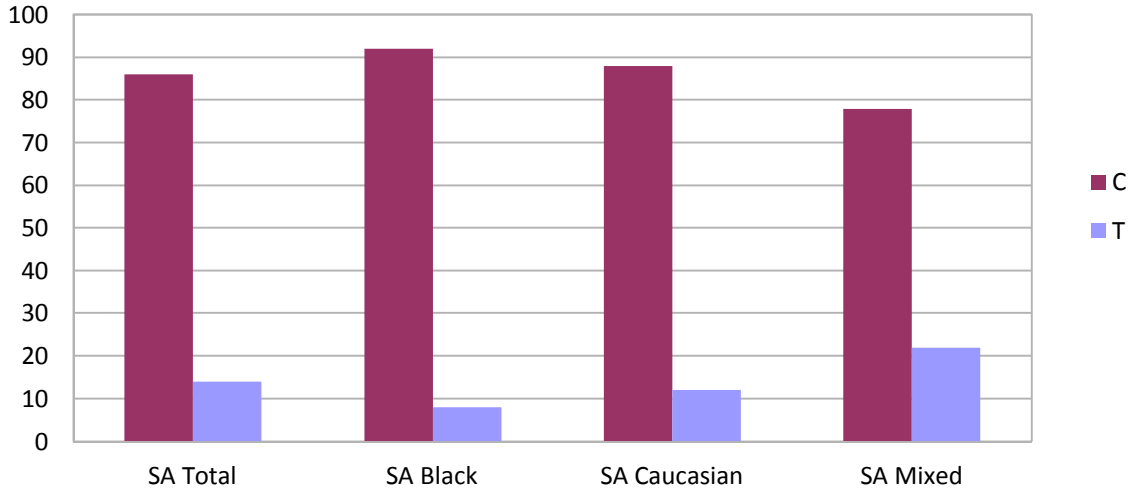

Figure 3.5: Percentage of allelic distributions for rs10279013 on the *ABHD11* gene in a healthy South African population

Overall, the autistic group showed high numbers of individuals displaying the C allele, with all ethnic groups above 90%. In the SA African group, 93% displayed the C allele in the autistic group, with 92% in the controls. For the T allele, numbers were very low, with 7% and 8% in the autistic and control groups, respectively. Allelic composition of the healthy and autistic groups showed a slight difference, but this difference was not statistically significant, with P=0.71843.

The SA Caucasian group displayed high numbers of individuals displaying the C allele, with 99% and 88% in the autistic and controls, respectively. The T allele was represented by low numbers in both groups with only 1% in the autistic sample group, and 12% in the controls. These percentages indicate a high statistical significance between the allele frequencies for the autistic and healthy groups, with **P=0.00016**.

In the SA mixed group, 97% of the autistic samples were found to represent the C allele and 78% for this allele was displayed by the control group. For the T allele, 3% was displayed by the autistic group, with the control group at a higher 22%. There is a highly significant statistical difference between the allelic numbers for the autistic and healthy group, with **P=0.00012**.

3.2.3. Genotyping analysis for rs2293484

In the total SA population (autistic and healthy), over 50% of the individuals genotyped were homozygous for the GG genotype. The CC genotype was distributed in low numbers in both autistic and healthy groups, at 10% and 7% respectively. The GG genotype was 53% and 55% for the autistic and healthy groups and the CG genotype was 37% and 38% respectively for the autistic and healthy populations. No significance was found when comparing the genotype frequencies for the autistic and healthy groups, with P=0.47237 (*table 3.2*).

Table 3.2: Genotype and allelic distributions of rs2293484 on *ABHD11* in 3 distinct ethnic groups in the South African autistic and healthy populations

Ethnic Group	Number	Genotype (%)			P-value	Allele (%)		P-value
		CC	CG	GG		C Allele	G Allele	
Total South African								
Autistic	197	20 (10)	72 (37)	105 (53)	0.47237	183 (39)	281 (61)	0.00018
Control	220	15 (7)	83 (38)	122 (55)		122 (28)	318 (72)	
SA Black								
Autistic	67	7 (11)	25 (37)	35 (52)	0.20413	39 (29)	95 (71)	0.76418
Control	62	2 (3)	30 (48.5)	30 (48.5)		34 (27)	90 (73)	
SA Caucasian								
Autistic	74	7 (9)	26 (35)	41 (56)	0.96856	41 (28)	107 (72)	0.64677
Control	71	6 (8)	24 (34)	41 (58)		36 (25)	106 (75)	
SA Mixed								
Autistic	56	7 (8)	38 (44)	42 (48)	0.66739	32 (29)	79 (71)	0.84148
Control	87	6 (11)	21 (37)	29 (52)		52 (30)	122 (70)	

*values in red are highly significant, where p-value <0.0001.

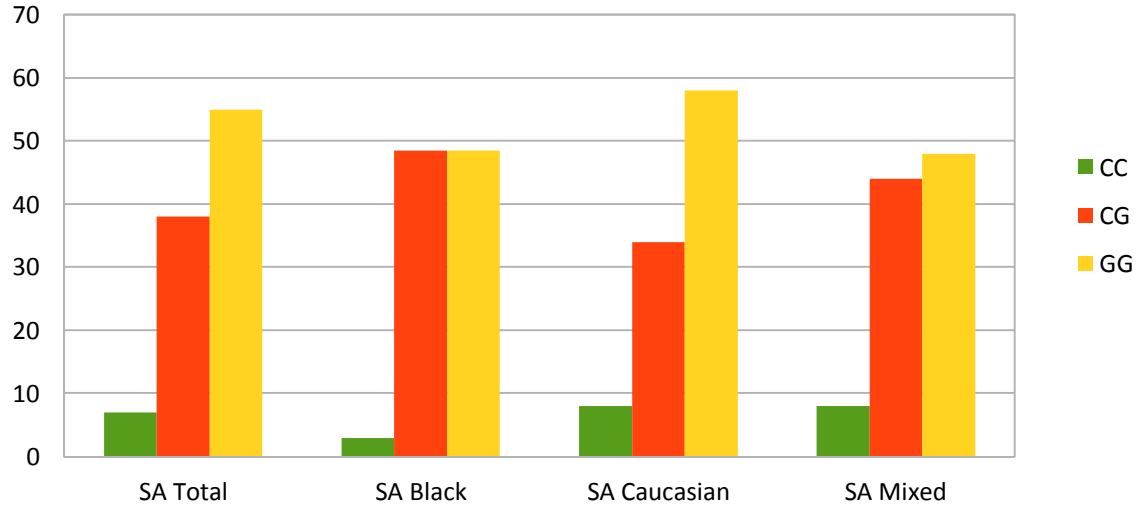

Figure 3.6: Percentage of genotype distributions for rs2293484 on the *ABHD11* gene in an autistic South African population

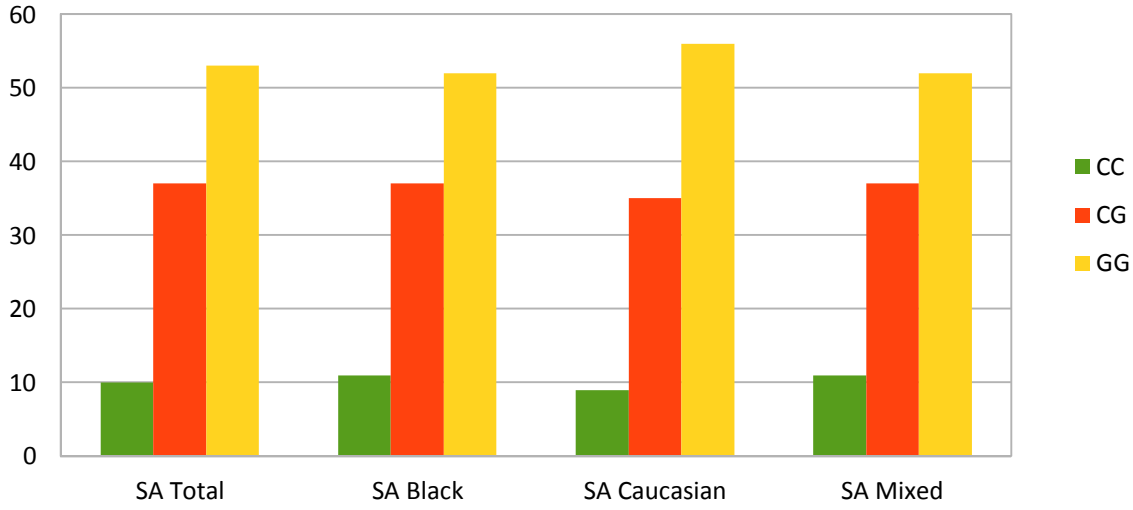

Figure 3.7: Percentage of genotype distributions for rs2293484 on the *ABHD11* gene in a healthy South African population.

For the SA Black population, the frequencies of the CC genotype was 11% and 3%, among the autistic and healthy groups, respectively. The GG genotype was dominantly expressed in both groups, but was somewhat higher in the autistic group at 52% as opposed to the 48% in the healthy group. The heterozygous CG genotype was 37% and 48.5% in the autistic and healthy groups, respectively. No statistical significance was found with P=0.20413 (*fig. 3.6 and 3.7*).

The SA Caucasian population had similar genotype distributions to the SA Black group. The homozygous GG genotype was expressed in the majority of individuals, with 56% and 58% in the autistic and healthy groups, respectively. The homozygous CC genotype was displayed at low vales, 9% and 8% in autistic and healthy groups, respectively. The heterozygous CG was expressed in 35% and 34% in the autistic and healthy populations, respectively. No statistical significance was found with P=0.096856.

In the SA Mixed population, the GG genotype was once again dominantly expressed in both autistic and healthy groups with a value of 48% and 52% respectively. The homozygous CC had the lowest values for both groups at 11% and 8% in the autistic and healthy populations, respectively. The CG heterozygous genotype was expressed at 44% in the autistic group and slightly lower in the control group at 33%. However, when comparing the autistic and the control samples, no statistical significance was found.

3.2.4. Allelic frequency distributions for rs2293484

Sixty-one percent of the individuals genotyped in the total South African autistic group expressed the G allele. In the control group, that number was at 72%. The C allele was displayed at lower levels, with 39% and 28% in the autistic and healthy groups respectively. Statistical significance was found when comparing the allelic frequencies of the autistic and

healthy groups, with **P=0.00018.**

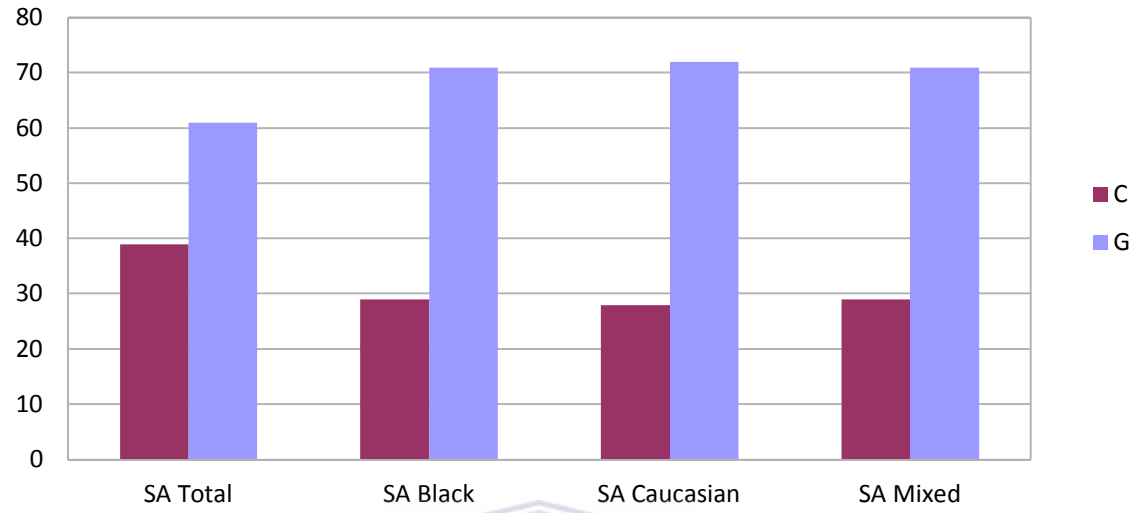

Figure 3.8: Percentage of G and C alleles displayed for rs2293484 on the *ABHD11* gene in an autistic South African population

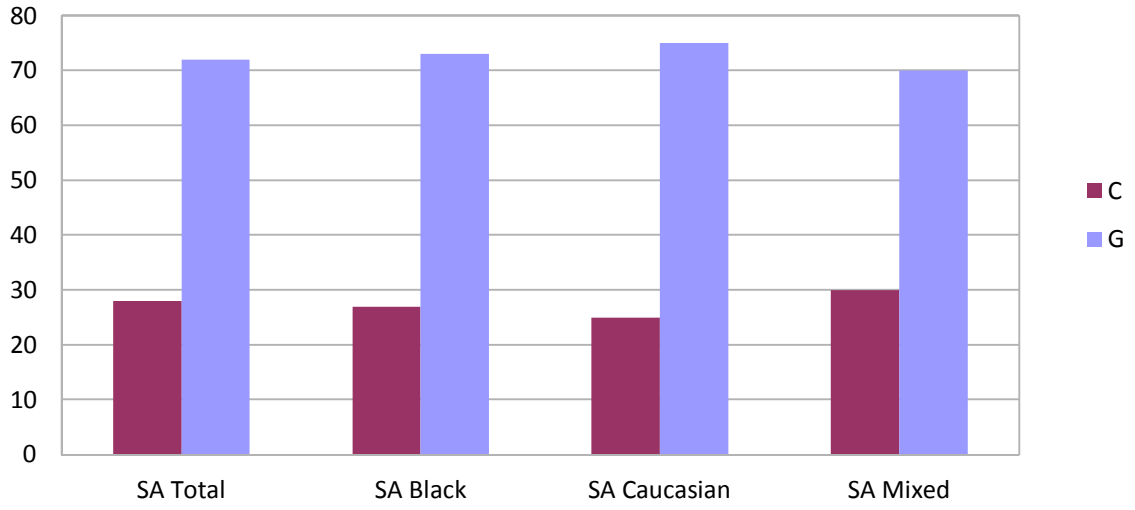

Figure 3.9: Percentage of G and C alleles displayed for rs2293484 on the *ABHD11* gene in a healthy South African population

In the SA Black population, the G allele was distributed at 71% in the autistic population and 73% in the control group. The C allele was expressed at less than that amount, 29% and 27% in the autistic and healthy groups, respectively. No statistical significance was found, with P=0.76418 (*fig 3.8 and 3.9*).

The healthy SA Caucasian group had the highest expression levels for the G allele, across the population for all ethnic groups, at 75%. The autistic group displayed this allele in 72% of individuals. The C allele was expressed at 25% and 28% in the control and autistic groups, respectively. No statistical significance was found when comparing the allelic frequencies for the autistic and healthy populations, with a P-value of 0.64677.

The SA Mixed population showed similar frequencies for the C and G alleles as the other ethnicity groups, with the G allele again being highly expressed in both the autistic and healthy groups, both in the region of and the C allele in the region of 30%70%. No statistical significance was found when comparing the association of the G and C allelic frequencies in the autistic and healthy populations, with P=0.84148.

3.2.5. Genotyping analysis for rs17060

In the total SA population, the homozygous CC genotype was dominant at 53% in the autistic group and 60% in the control group. The heterozygous CT genotype was similar for both groups, with 39% in the autistic group and 35% in the controls. The TT homozygous genotype was very low for both groups at 8% and 5% respectively (*table 3.3*). There was no statistical significance found when comparing the total autistic and healthy groups, with P=0.41066.

Table 3.3: Genotype and allelic distributions for rs17060 from the *ABCA13* gene in South African healthy and autistic groups

*values in red are highly significant, where p-value <0.0001.

Ethnic Group	Number	Genotype (%)			P-value	Allele (%)		P-value
		CC	CT	TT		C Allele	T Allele	
Total South African								
Autistic	206	112 (53)	80 (39)	14 (8)	0.41066	331 (78)	95 (22)	0.20032
Control	213	128 (60)	75 (35)	10 (5)		331 (78)	95 (22)	
SA Black								
Autistic	75	31 (41)	35 (47)	9 (12)	0.72885	97 (65)	53 (35)	0.43479
Control	74	35 (47)	32 (43)	7 (10)		102 (69)	46 (31)	
SA Caucasian								
Autistic	80	50 (63)	26 (32)	4 (5)	0.05558	125 (79)	34 (21)	0.01991
Control	70	54 (77)	16 (33)	0		125 (89)	16 (11)	
SA Mixed								
Autistic	51	31 (61)	19 (37)	1 (2)	0.79399	81 (79)	21 (21)	0.54300
Control	69	39 (56)	27 (39)	3 (5)		105 (76)	33 (24)	

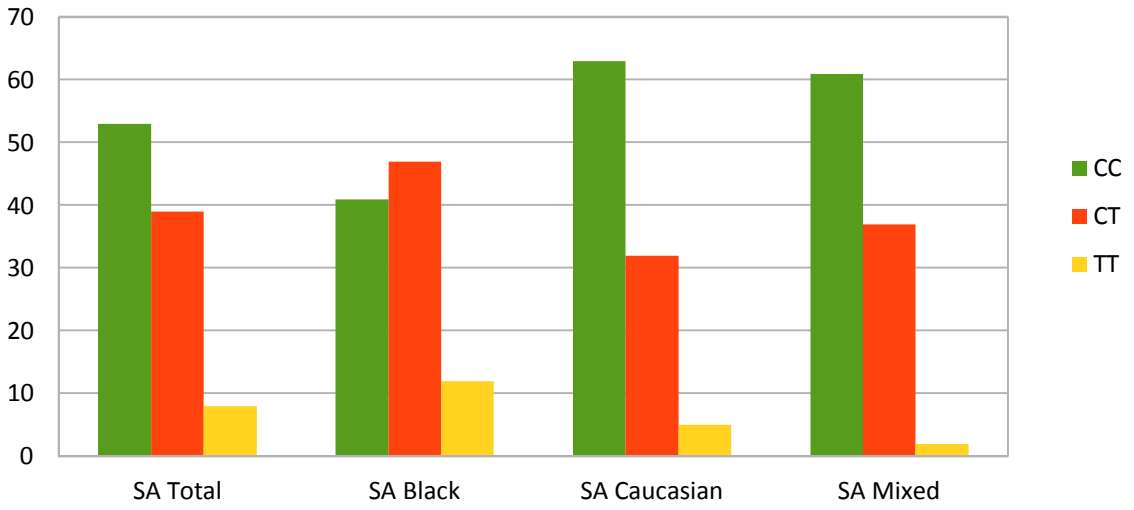

Figure 3.10: Percentage of genotype distributions for rs17060 on the ABCA13 gene in an autistic South African population.

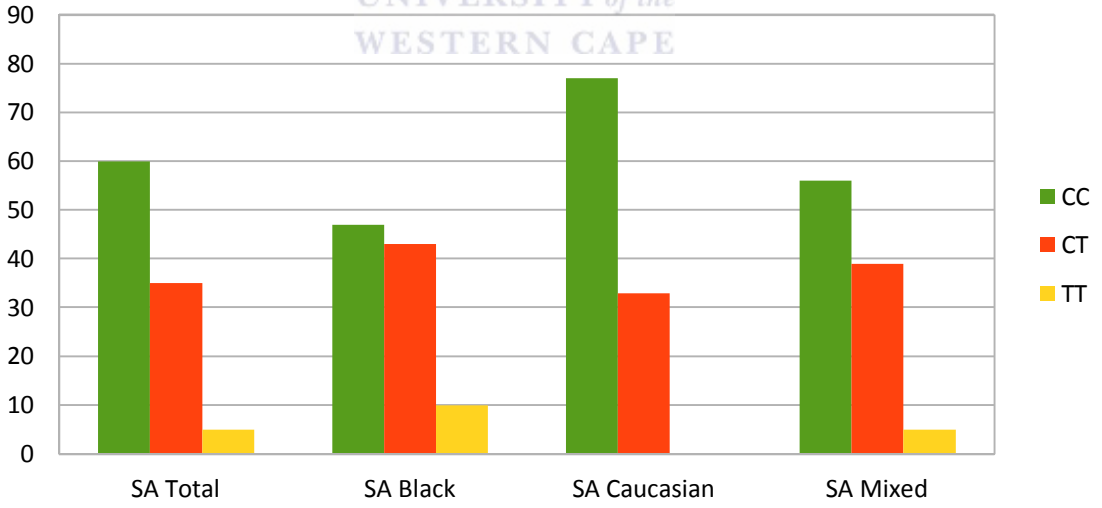

Figure 3.11: Percentage of genotype distributions for rs17060 on the ABCA13 gene in a healthy South African population.

The SA Black population showed that 41% of individuals in the autistic group were homozygous for the CC genotype, while the respective figure in the control group was 47%. The heterozygous CT genotype was present at 47% and 43% in the autistic and control groups, respectively. The respective values for the TT genotype was 12% and 10% in both autistic and control groups. When comparing the different genotypes, no significant difference was found between the autistic and control groups, with P=0.72885.

In the SA Caucasian group, the frequency of the CC genotypes among the autistic and control groups was 63% and 77% respectively. The heterozygous CT genotype was present in 32% of individuals in the autistic group and 33% in individuals in the control group. The TT homozygous genotype was under-represented in both groups, with 5% in the autistic group and none of the individuals in the control group tested for the TT genotype (*fig 3.10 and 3.11*). When comparing the different genotypes, no significant difference was found between the autistic and control groups, with P=0.0.5558.

For the SA mixed group, genotype frequency of the CC, CT and TT genotype among the healthy group was 56%, 39% and 5%, whereas in the autistic group the same was found to be 61%, 37% and 2%, respectively. When comparing the different genotypes, no significant difference was found between the autistic and control groups, with P=0.79399.

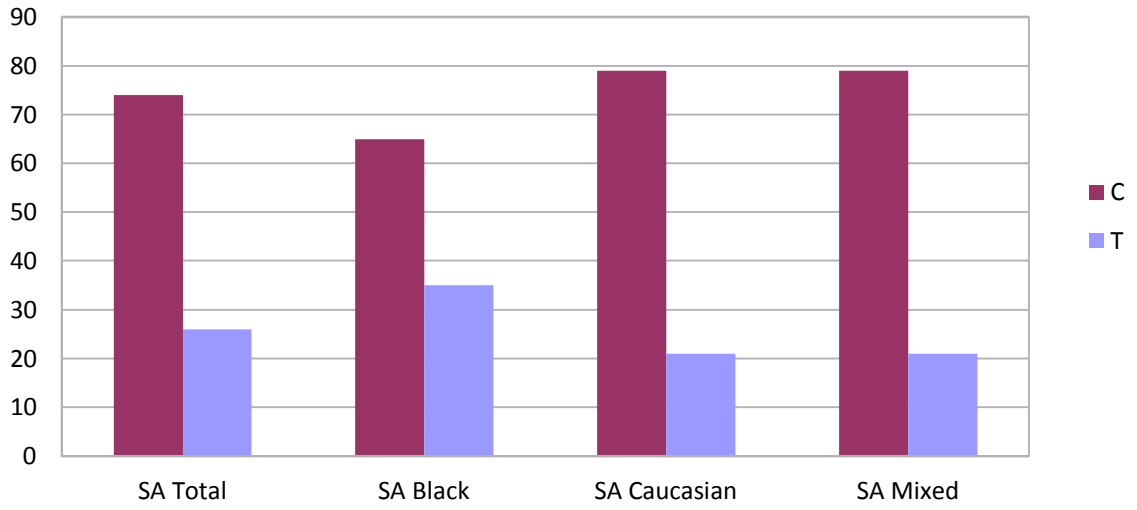

Figure 3.12: Percentage of A and C alleles displayed for rs17060 on the *ABCA13* gene in an autistic South African population

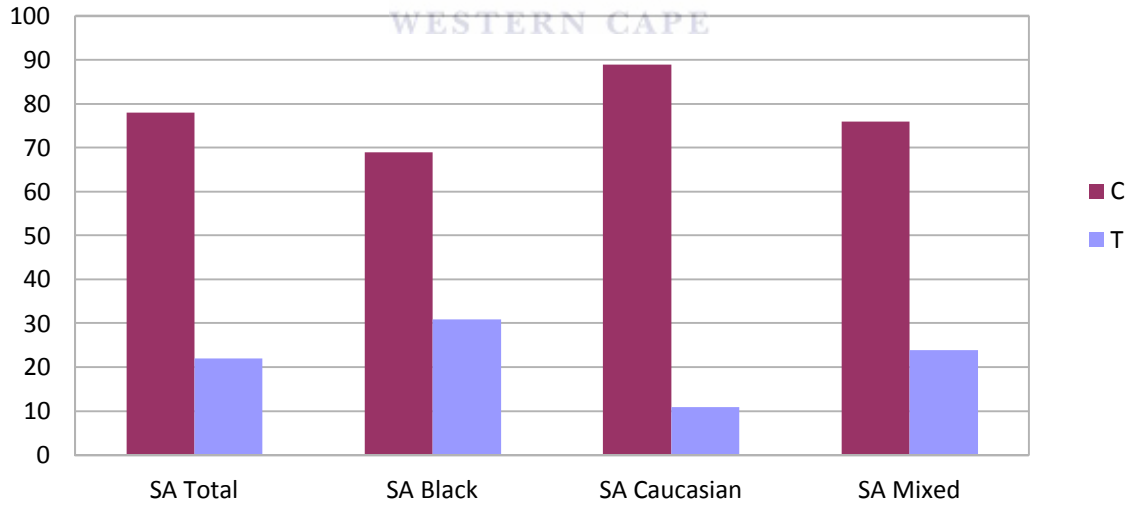

Figure 3.13: Percentage of A and C alleles displayed for rs17060 on the *ABCA13* gene in a healthy South African population

3.2.6. Allelic analysis for rs17060

In the total SA population, the frequency of the dominant C allele, was 78% for in both healthy and autistic groups. The T allele frequency was 22% in both the healthy and autistic groups (*fig. 3.12 and fig. 3.13*). No statistical significance was found when comparing the autistic and healthy populations, with P=0.20032.

In the SA Black population, the frequency of the C allele was the highest, with 65% in the autistic group and 69% in the controls, respectively. The T allele was displayed at 35% and 31% for autistic and control groups, respectively. No significance was found when comparing the autistic and control groups, with P=0.43479.

The SA Caucasian group had the highest values for the C allele across all the ethnic groups, with 79% in the autistic group and 89% in the control group. The T allele was lower at 21% and 11% respectively. The difference of allelic frequency of autistic and healthy groups was found to be statistically significant, with P=0.01991.

In the SA Mixed group, the frequency for the C allele was 79% in the autistic group and 76% in the control group. The T allele was displayed at 21% and 24% respectively, with no statistical significance found when the allelic frequency for the autistic and healthy groups was compared, where the P-value=0.54300.

CHAPTER 4: INDEX

Chapter 4: A BIOINFORMATICS APPROACH TO IDENTIFYING CANDIDATE GENES INVOLVED IN AUTISM

4.1. Introduction

Over the past two decades, more than 1200 genes causing human diseases or traits have been identified, largely by a process that is called 'positional cloning' (Bostein and Risch, 2003). Through positional cloning, genes controlling Mendelian traits or diseases are identified and isolated with only the prior information that the phenotype is inherited. One of the most successful and widely-used methods of candidate gene discovery is by the use of single nucleotide polymorphisms (SNPs). Linkage analysis is dependent on the availability of 300-500 highly informative genetic markers spanning the entire genome and this research is applied on families (Collins *et al*., 1998). However it is considerably harder to look for genes contributing to the risk of heart disease, cancers and psychiatric disorders as these phenotypes are affected by multiple genes, each with a small effect in combination with environmental factors. While these classic methods of candidate gene discovery have been wholly useful in laying the foundation for effective genetic research, these methods are limited by its reliance on previous knowledge of the physiological, biochemical and functional aspects of possible candidates (Zhu and Zhao, 2007).

Bioinformatic resources prove invaluable in facilitating gene identification. Genomic information is cumulative and the availability of the collected knowledge of human genetics has proven to be the central success of the Human Genome Project. Processes that used to be very tedious and labour intensive such as physical mapping has become a simple process of PCR with data being sent to a website and results sent to the requestor via email (Bostein and Risch, 2003).

Computational analysis and bioinformatics in genetic studies has proven to be an invaluable

research tool. Data-intensive research has been made simpler in terms of analysis and database maintenance (Perez-Iratxeta et al., 2007). One of the important and extremely useful roles these types of databases play in biological studies is that they are able to organize genes and proteins within their larger biological context. This includes identifying common pathways and protein interactions of various genes and proteins in different biological functions (Romero et al., 2004).

4.2. STRING database (http://www.bork.embl-heidelberg.de/STRING/)

Direct physical binding is not the limitation of protein-protein interactions. They may also interact indirectly – by sharing a substrate in a metabolic pathway, regulating each other transcriptionally, or during participation in larger multi-protein assemblies (von Meiring et al., 2003). The prediction of these functional associations (for direct and indirect binding) is made simpler by the current rate and growth of genomic information becoming available. This information offers unique opportunities of inference methods through 'genomic context' or 'nonhomology-based' information (Galperin and Koonin, 2000; Marcotte, 2000; Huynen et al., 2000).

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a wide-ranging visual database that assimilates information on known and predicted protein interactions (Snel *et al.*, 2000). This database offers a unique visual representation of very specific protein-protein interaction networks, which can be manipulated by adding more proteins, taking them away or changing their positioning within the network (Jensen *et al.*, 2009; Sklarczyk *et al.*, 2011). STRING can also be used to identify new/unknown proteins in a specific network of genes. The latest version of STRING covers more than 1100 completely sequenced organisms. Interactions include direct and indirect associations and these are

derived from four sources: genomic context, high-throughput experiments, conserved co-expression and previous knowledge (http://string-db.org).

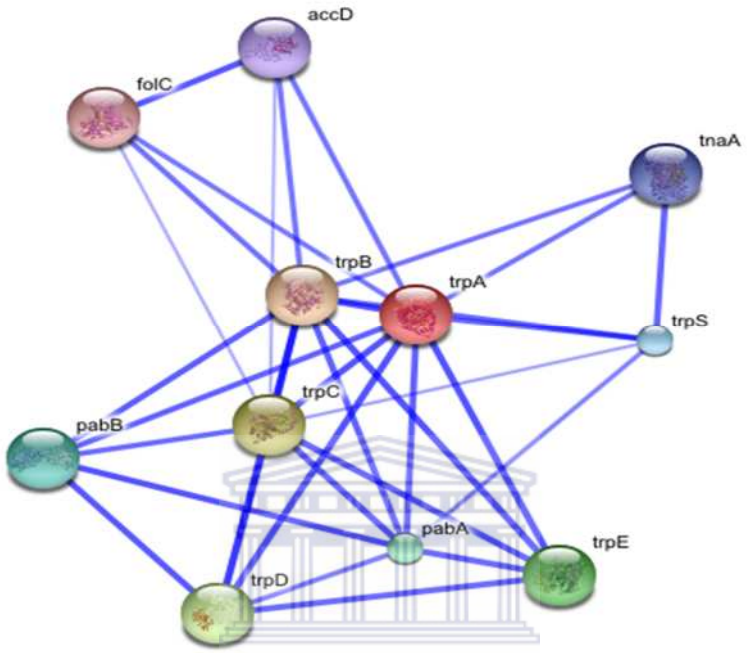

Fig. 4.1: An example of the visual representation of a STRING network (http://string-db.org/version_9_05/newstring_cgi/show_network_section.pl?limit=0&targetmode=proteins&caller_identity=gene_cards&network_flavor=evidence&identifiers)

4.3. GeneCards® (http://www.genecards.org)

The Weizmann Institute of Science in collaboration with Life Map Sciences Incorporated established the GeneCards ® database in 1997 with the aim to integrate the massive amounts of genomic data being generated by the Human Genome Project and subsequently spread over various specialized databases (Reghan *et al*., 1997). The current version automatically mines over 90 sources in an offline process and constructs a combined gene list as follows:

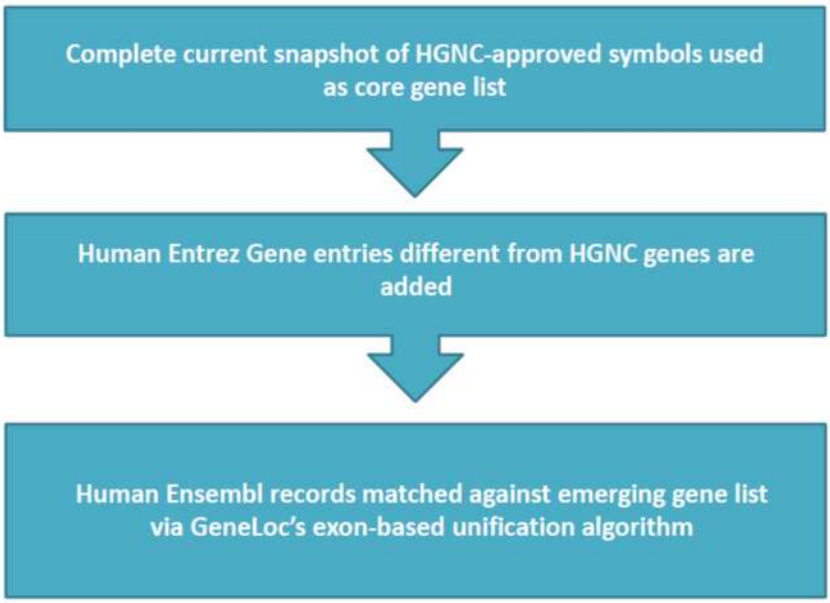

Fig 4.2: The methodology used by the GeneCards ® database to mine and store information on all known genes

These primary sources of information provide annotations for aliases, descriptions, previous symbols, gene category, location, summaries, paralogues and non-coding RNA details (Stelzer *et al*., 2011). The GeneCards database provides the user with a comprehensive, multifaceted tool that allows for the characterization of genes from the most basic point, the nucleotide, right up to the associated pathways, protein interactions and modifications, protein domain information, transcriptional information and gene expression while providing researchers a summarized look at complicated genomic data for use in molecular research (Safran *et al*., 2010).

4.4. Transcription factors and binding sites

The most common identifier for transcription factors (TFs) is that they are proteins that bind to specific DNA sequences, thereby controlling the flow of genetic information from DNA to mRNA. TFs either perform this function alone or with other proteins in a complex, by activating or blocking the recruitment of RNA polymerase to specific genes.

A defining factor for TFs is that they contain one or more DNA-binding domains which attach specific sequences of DNA adjacent to the genes they regulate. The efficiency of TFs is responsible for the regulation and control of all biological functions in eukaryotic organisms. Muscle differentiation in embryonic development (MYOD), helping kidneys reclaim water at times of dehydration (NR3C2) and even the instigation of oncogenesis (MYC) are just some of the examples of the functions that TFs serve (Yusuf *et al.*, 2012).

Many TFs are implicated in disease. Out of a growing list of 1321 human TFs, 197 are currently linked to one or more diseases in the OMIM Morbid Map (Yusuf *et al*., 2012; Fulton *et al*., 2009 and Vaquerizas *et al*., 2009). The Transcription Factor Encyclopedia is a web-based compendium of literature and content regarding information on transcription factors. Much in the style of GeneCards®, the TF Encyclopedia offers a comprehensive look at identified TFs from human, rat and mouse. However, the information output is available in pdf format, in the form of mini reviews written and vetted by experts in the field (http://www.dimasyusuf.com/transcription-factor-encyclopedia-website/).

Aims and objectives of this chapter

Aim: The main aim of this chapter is to ascertain whether the genes on chromosome 7 have any common pathways and/or regulating transcription factors that may attribute to the fact that chromosome 7 is viewed as an "autism hotspot'.

Objectives: Further objectives of the study are:

- The identification of additional potential candidate genes, using the STRING database
- The bioinformatics analysis of the candidate genes by summarizing genomic information provided by the freely-available GeneCards® database
- The identification of common transcription factors regulating the identified candidate genes

4.5. Materials and methods

4.5.1. Identification of potential candidate genes

A list of already identified candidate genes from chromosome 7 was identified from literature (*table 1).* A search was then performed in the STRING database, using the genes as the input query to determine further candidate genes on chromosome 7 interacting with the existing list of genes from the literature. The default parameters were used. The top results from the STRING search - only genes with a confidence score of 0.9 and above were included in the network, and no more than 20 genes were included in the network - as well as the existing list of candidate genes was included. The full list of genes then further investigated by submitting them to a GeneCards® database search. This was done to obtain specific information regarding the molecular function(s) of the genes, interacting transcription factors and any involvement in autism or ASDs. Characteristics of interest included cellular location of the gene, molecular function, participation in a molecular process, possible involvement in cellular processes, protein-protein interactions, domains present in the gene and the presence of transcription factor binding sites in the promoter region of the gene.

4.6. Results

9 candidate genes, already positively linked to autism and ASDs by previous studies (*table 1*) were subjected to STRING analysis. A further 19 genes were identified as co-expressed genes with the list of previously identified candidate genes from chromosome 7. Of the 19 identified genes, two were found to be located on chromosome 7.

4.6.1. GeneCards® results

A summary of the GeneCards® results assigning basic functional annotations for the 9 previously identified genes, as well as the two genes from the STRING analysis can be seen in table 4.2.

While there was no direct link to Autism and ASDs in the criteria that was investigated, the GeneCards® results for most of the candidate genes yielded information pertaining to transcriptional activities as important molecular processes that the genes are involved in. GeneCards® also substantiated the results as per table 1, showing the associated disorders and diseases for the different candidate genes. For the genes identified by the previous microarray study carried out by this research group (Gameeldien 2010), *ABCA13* and the *FZD9* and *FZD1* genes identified by the STRING interactions, all 3 showed associated diseases or disorders that have been proven to be co-morbid with Autism and ASDs

Table 4.1: Summary of Genecards ® results

Gene name and symbol
WNT2
wingless-type MMTV integration site family member 2
Alternative names
Int-1-Like Protein, Secreted Growth Factor
Cellular location of protein
extracellular region, proteinaceous extracellular matrix, extracellular space, colocalizes with plasma membrane

Molecular function
receptor binding, frizzled binding, frizzled-2 binding, cytokine activity, receport agonist activity

Molecular process
positive regulation of endothelial cell proliferation, positive regulation of mesenchymal cell proliferation, lens development in camera-type eye, signal transduction, cell-cell signalling

Cellular pathways
DNA damage response, Wnt signaling pathway, Wnt signaling network

Protein domains
Belongs to the Wnt family
Protein-protein interactions
FZD1, FZD9, SFRP1, WNT3A, WNT2B

Gene name and symbol	EN2 **Engrailed Homeobox 2**
Alternative names	Homeobox Protein Engrailed-2, Engrailed Homolog 2, Engrailed-2
Cellular location of protein	nucleus, membrane
Molecular function	sequence-specific DNA binding transcription factor activity, sequence-specific DNA binding
Molecular process	multicellular organismal development, neuron differentiation, midbrain development, hindbrain development, positive regulation of transcription from RNA polymerase II promoter
Cellular pathways	-
Protein domains	Belongs to the engrailed homeobox family, Contains 1 homeobox DNA-binding domain
Protein-protein interactions	FOXA2

Table 4.1 continued

Gene name and symbol	*IMMP2L* **IMP2 inner mitochondrial membrane peptidase-like (S. cerevisiae)**
Alternative names	Mitochondrial Inner Membrane Protease Subunit 2, Inner Mitochondrial Membrane Peptidase 2 Like
Cellular location of protein	nucleus, mitochondrion, integral to membrane, mitochondrial inner membrane peptidase complex
Molecular function	peptidase activity, serine-type peptidase activity
Molecular process	ovarian follicle development, proteolysis, protein processing involved in protein targeting to mitochondrion, spermatogenesis, ovulation
Cellular pathways	Protein export
Protein domains	Belongs to the peptidase S26 family. IMP2 subfamily
Protein-protein interactions	ALDH1B1, GPX4, GUF1, HACL1, IDO2

Gene name and symbol	FOXP2 **Forkhead Box P2**
Alternative names	CAG Repeat Protein , Forkhead/Winged-Helix Transcription Factor, Forkhead Box Protein P2, Trinucleotide Repeat Containing 10
Cellular location of protein	nucleus, transcription factor complex, cytoplasm
Molecular function	DNA binding, chromatin binding, double-stranded DNA binding, sequence-specific DNA binding transcription factor activity, RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity
Molecular process	positive regulation of mesenchymal cell proliferation, pattern specification process, skeletal muscle tissue development, embryo development, post-embryonic development
Cellular pathways	**Wnt / Hedgehog / Notch** The leucine-zipper is required for dimerization and transcriptional repression
Protein domains	(By similarity), Contains 1 C2H2-type zinc finger, Contains 1 fork-head DNA-binding domain
Protein-protein interactions	FOXP1, FOXP4, GATAD2B, CTBP1, NFATC2

Table 4.1 continued

Gene name and symbol	RELN **Reelin**
Alternative names	RL, PRO1598
Cellular location of protein	proteinaceous extracellualr matrix, extracellular space, cytoplasm, dendrite
Molecular function	protein serine/threonine/tyrosine kinase activity, serine-type peptidase activity, metal ion binding, lipoprotein particle receptor binding, very-low-density lipoprotein particle receptor binding
Molecular process	cell morphegenisi involved in diferentiation, neuron migration, proteolysis, cell adhesion, axon guidance
Cellular pathways	Focal adhesion, ECM-receptor interaction
Protein domains	Belongs to the reelin family, 16BNR repeats, 8 EGF-like domains, 1 reelin domain
Protein-protein interactions	LRP8, VLDLR, DAB1, PCDHA4, PCDHA6

Gene name and symbol	ABCA13 **ATP-binding cassette, sub-family A (ABC1), member 13**
Alternative names	ATP-Binding Cassette, Sub Family (ABC1), ATP Binding Cassette Transporter
Cellular location of protein	integral to membrane
Molecular function	ATP binding, ATPase activity
Molecular process	Transport
Cellular pathways	ABC transporters
Protein domains	Belongs to the ABC transporter superfamily, 2 ABC domains
Protein-protein interactions	ACD, GABARAP, APOA1, APOA2

Table 4.1 continued

Gene name and symbol	HOXA1 **Homeobox A1**
Alternative names	Homeobox Protein Hox-A1, Homeobox Protein Hox-1F, HOX A1 Homeodomain Protein
Cellular location of protein	Nucleus
Molecular function	sequence-specific DNA binding transcription factor activity, protein binding, sequence-specific DNA binding
Molecular process	transcription, DNA-dependent, regulation of transcription from RNA polymerase II promoter, multicellular organismal development, nervous system development, sensory perception of sound
Cellular pathways	Neural Crest Differentiation
Protein domains	Belongs to the Antp homeobox family. Labial subfamily, Contains 1 homeobox DNA-binding domain
Protein-protein interactions	TRAPPC6A, EFEMP2, LPXN, PRNP, CHIC2

Gene name and symbol	UBE2H **ubiquitin-conjugating enzyme E2H**
Alternative names	UBE2H, Ubiquitin-Conjugating Enzyme E2H
Cellular location of protein	
Molecular function	ubiquitin-protein ligase activity, protein binding, ATP binding, acid-amino acid ligase activity
Molecular process	ubiquitin-dependent protein catabolic process, protein ubiquitination, protein K48-linked ubiquitination, protein K11-linked ubiquitination
Cellular pathways	Ubiquitin mediated proteolysis/protein modification, protein ubiquination
Protein domains	Belongs to the ubiquitin-conjugating enzyme family
Protein-protein interactions	HIST2H2AA3, CD99, DTX3, MARCH2, MKRN3

Table 4.1 continued

Gene name and symbol	CNTNAP2 **contactin associated protein-like 2**
Alternative names	Contactin-Associated Protein-Like, Homolog Of Drosophila Neurexin, NRXN4, AUTS15
Cellular location of protein	early endosome, Golgi apparatus, voltage-gated potassium channel complex, cell surface, membrane
Molecular function	receptor binding, protein binding, enzyme binding
Molecular process	cell adhesion, signal transduction, brain development, behaviour, neuron recognition
Cellular pathways	cell adhesion molecules
Protein domains	Belongs to the neurexin family, Contains 2 EGF-like domains, Contains 1 F5/8 type C domain, Contains 1 fibrinogen C-terminal domain, Contains 4 laminin G-like domains
Protein-protein interactions	CASK, CNTN2, CTR9, ZMIZ1, KCNA2

Gene name and symbol	FZD9 **frizzled family receptor 9**
Alternative names	Frizzled Homolog, Frizzled 9, Seven Transmembrane Spanning Receptor
Cellular location of protein	cytoplasm, plasma membrane, cell surface, membrane, integral to membrane, filopodium membrane, perinuclear region of cytoplasm signal transducer activity, receptor activity, transmembrane signaling
Molecular function	receptor activity, G-protein coupled receptor activity, protein binding, Wnt-protein binding, PDZ domain binding, protein homodimerization activity, Wnt-activated receptor activity, protein heterodimerization activity
Molecular process	vasculature development, cell surface receptor signaling pathway, nervous system development, neuroblast proliferation, brain development, learning or memory, gonad development, embryo development, B cell differentiation
Cellular pathways	Wnt signaling pathway, Melanogenesis, Pathways in Cancer, Basal cell in carcinoma
Protein domains	G-protein coupled receptor Fz/Smo family, 1 FZ (frizzled) domain
Protein-protein interactions	MDFI, WNT2, WNT1, WNT7A, LRP6

Table 4.1 continued

Gene name and symbol	FZD1 **frizzled family receptor 1**
Alternative names	Frizzled 1, Seven Transmembrane Spanning Receptor, Wnt Receptor
Cellular location of protein	cytoplasm, plasma membrane, cell surface, integral to membrane, synaptosome
Molecular function	receptor binding, protein binding, Wnt-protein binding
Molecular process	positive regulation of protein phosphorylation, vasculature development, membranous septum morphogenesis, muscular septum morphogenesis, outflow tract morphogenesis
Cellular pathways	Wnt signaling pathway, Melanogenesis, pathways in cancer, basal cell carcinoma
Protein domains	Lys-Thr-X-X-X-Trp motif interacts with the PDZ doman of Dvl (Disheveled) family members and is involved in the activation of the Wnt/beta-catenin signaling pathway, The FZ domain is involved in binding with Wnt ligands
Protein-protein interactions	DLG4, WNT4, GNAO1, WNT1, WNT2

4.6.2 Common transcription factors for the identified candidate genes on chromosome 7

A list of transcription factors for each candidate gene was obtained via the Genecards ® database. The identified TFs were ranked based on the number of genes they regulate above 75% of the genes. Finally, a list of ranked TFs was generated.

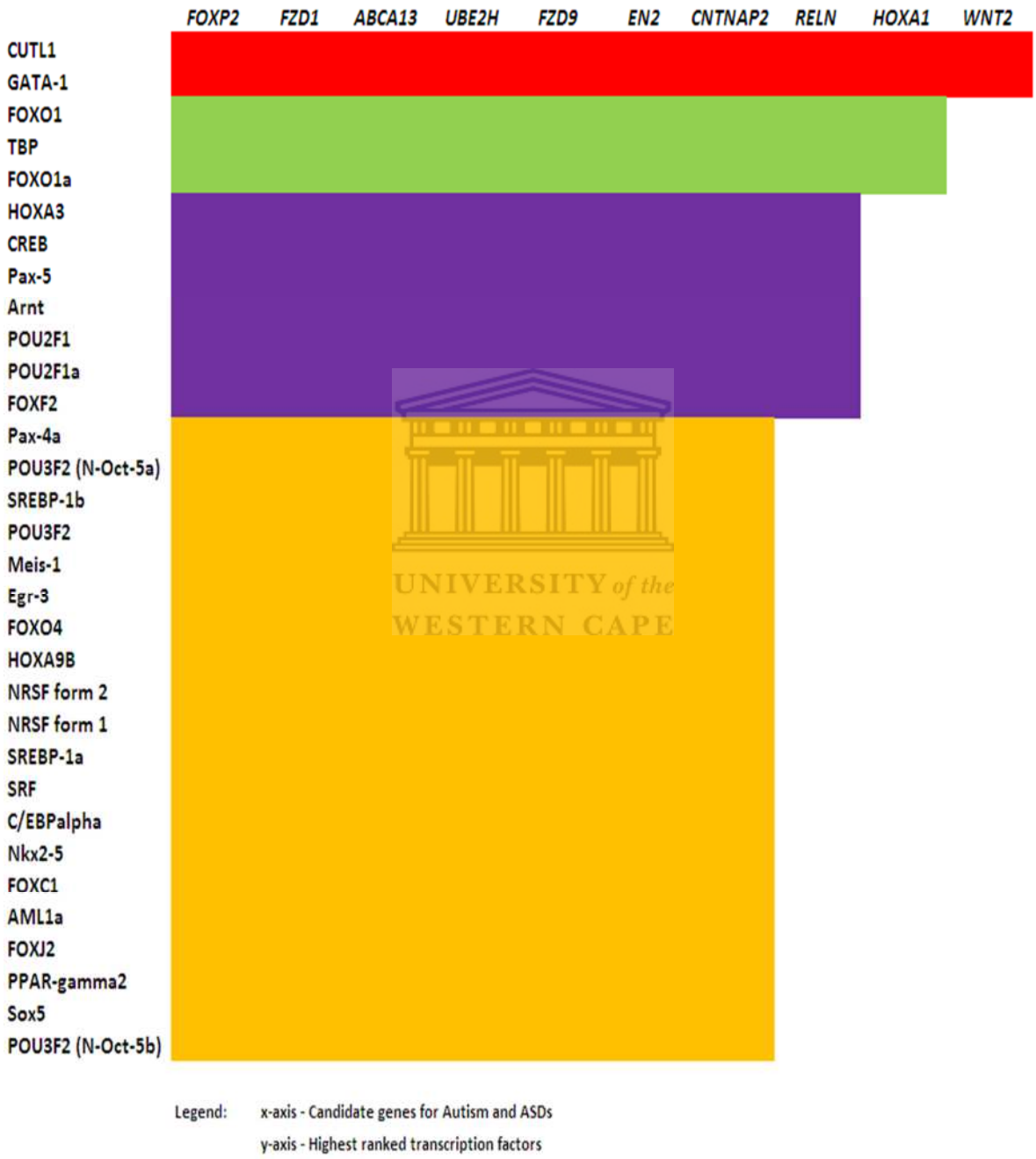

Figure 4.3: The highest ranking transcription factors regulating candidate genes for Autism and ASDs

CHAPTER 5: INDEX

Chapter 5: DISCUSSIONS AND CONCLUSIONS

Chapter 5: DISCUSSIONS AND CONCLUSIONS

5.1. SNP analysis

Recent research has shown that hundreds of genetic variants – big, small, rare, common, inherited and spontaneous – along with environmental factors contribute to autism risk (Hughes 2013). In 2012, in a massive genome-wide association study, researchers identified 57 SNPs associated with autism. The findings were presented at the International Meeting for Autism Research in 2012. The identified 57 SNPs predict with a high level of certainty that siblings of children with ASD have a higher risk of also developing the condition (Schellenberg *et al*., 2012). This study comprised of more than 1 000 multiplex families of which one or more children were diagnosed with ASD. These findings support an earlier study conducted by Carayol *et al*., (2011), where 8 autism-related SNPs were identified (Carayol *et al*., 2011).

5.2. Genes *ABCA13* and *ABHD11*

While there has been no conclusive association for either *ABHD11* or *ABCA13* to autism and ASDs, strong association has been found for both for co-morbid disorders to autism and ASDs (Tordjmann *et. al*., 2012 and Stahlberg *et. al*., 2003). Co-morbid disorders for *ABCA13* include schizophrenia, bipolar disorder and major depression. *ABHD11* has been implicated as a causative genetic factor for WBS. The development of this disorder is caused by the deletion of contiguous genes at the 7q11.23 position (Berg *et al*., 2007). The behavioural phenotype of WBS has been considered to be the polar opposite of ASD, but research suggests that common WBS deletions can result in a myriad of social communication impairments, such as excessive talkativeness and over friendliness, but also the absence of verbal language and poor social skills, which is a common phenotype displayed by individuals on the ASD spectrum (Tordjmann *et al*., 2012). The most common types of

structural variation in the human genome, CNVs (copy number variants) have been attributed to many genetic disorders associated with WBS (Merle *et al*., 2010). Some of these include cardiovascular and connective tissue abnormalities, growth and endocrine problems, neurological problems and cognitive and behavioural problems.

The present study is the first to investigate the prevalence of *ABHD11* and *ABCA13* SNPs in a South African population across different ethnic groups.

5.3. SNP rs10279013 and rs2293484 on *ABHD11*

5.3.1. rs10279013

The CC genotype was displayed at the highest frequency in the total South African population, across all ethnic groups. 98% of the autistic Caucasian group exhibited this genotype, with similarly high numbers in the other two ethnic groups (*table 2.1*). The TT genotype was not displayed in any of the groups, across the ethnic groups for healthy or autistic. The heterozygous CT genotype was found to be present in varying values; the autistic population exhibited low values between 2% and 14% whereas the healthy population had higher frequencies of this genotype, between 16% and 44%. The autistic Caucasian population showed a low frequency for the CT genotype at 2%. The highest frequency for this genotype was found in the healthy Mixed population, at 44%. The minor C allele for this SNP has been found in the East Asian, North African, Central African and West African populations (www.genecards.org). The statistical significance indicates that the CC genotype and C allele predisposes individuals in the Caucasian and Mixed groups to autism (*table 2.1*).

5.3.2. rs2293484

In both the autistic and control groups, the G allele was displayed at the highest frequency by individuals in the South African population, across all the ethnic groups. The SA Caucasian groups displayed the highest number of individuals with the G allele, with the SA Mixed group showing the lowest number. Similar results have been shown in the Han Chinese and Japanese populations. The C allele was displayed in low numbers across all groups in the population (Knight *et al.*, 2009) No significance for this SNP was found in any cohort studies with regards to autism and its' associated disorders.

5.4. SNP rs17060 on *ABCA13*

5.4.1. rs17060

In the SA total population, for rs17060, the C allele was dominant, displayed in more than 60% of individuals in the autistic group and 70% in the healthy group. The T allele was displayed in very low numbers in both groups, with only 11% of individuals in the SA Caucasian group of the controls displaying it. The genotype and allelic frequencies has not yet been investigated in human populations for autism, and the statistical significance indicated in this study has found that in a South African healthy and autistic population, males with the T allele for rs17060 are more susceptible to autism.

The present study provides evidence for not only the association of the investigated SNPs in a South African population, but could also allude to the co-morbidity of autism and WBS, as the *ABCA13* gene is involved in the development of WBS.

5.5. Bioinformatics

During the course of this study, a bioinformatics approach was used to identify and characterise novel candidate genes on chromosome 7 possibly involved in Autism and ASDs. Candidate genes were identified from a previous study and two additional new candidate genes were identified using the STRING database by using existing candidate genes and searching for available networks and interactions with other genes.

5.5.1. New candidate genes *FZD1* and *FZD9*

Both new, additional *FZD1* and *FZD2* genes that were identified using the STRING database are members of the "frizzled"gene family. Members of this gene family encode 7-transmembrane domain proteins that are receptors for Wnt signaling proteins. Of the 19 new potential candidate genes that were identified from STRING analysis, *FZD1* and *FZD9* were the only two genes located on chromosome 7, which made them suitable candidates for further investigation as to their properties and possible involvement in ASD and co-morbid disorders.

FZD1 is located at chromosomal position 7q21, and contains a single peptide; a cysteine-rich domain in the N-terminal extracellular region, 7 transmembrane domains, and a C-terminal PDZ domain-binding motif. The *FZD1* transcript is expressed in various tissues.

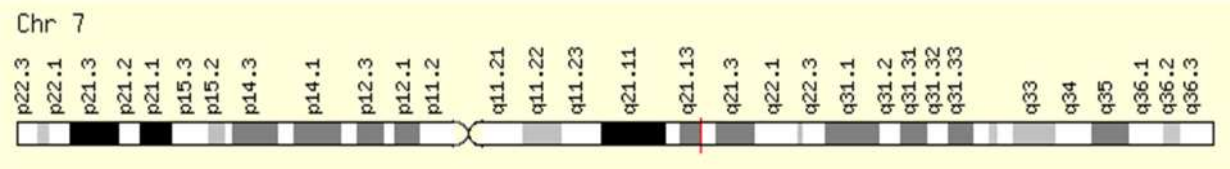

Figure 5.1: Red line indicates the position of the *FZD1* gene on chromosome 7 (http://www.genecards.org/cgi-bin/carddisp.pl?gene=FZD1)

The *FZD1* gene has been implicated in regulating chemo-resistance for cancer treatment (Flahaut *et al*., 2009). In the 2009 study, FZD1 upregulation in resistant variants was shown to mediate and sustain activation of the Wnt/ß-catenin pathway. During this process, a strong decrease was seen in the expression of *MDR1* – another ß-catenin target gene – revealing a complex interaction, implicating *FZD1* in clinical chemoresistanc via the Wnt/ß-catenin mediation. Wnt receptors have been shown to play a role in the development of human cancer (Sagara *et al*., 1998) by the regulation of ß-catenin pathways and *FZD1* mRNA is found to populate a significant amount of human tissue, including the adult heart, placenta, lung, kidney and fetal kidney. ß-catenin mutation is also detected in some cases of gastric cancer, coclorectal cancer, ovarian cancer and endometrial uterus cancer (Kawanishi *et al*., 1995, Rubinfeld *et al*., 1994, Palacios and Gamallo 1998 and Fukuchi *et al*., 1998). Another study implies that the Wnt/Fzd signaling pathway specificity is reliant on the subcellular localization of the *FZD* genes, through association with other membrane proteins (Jun et al., 2004). They have found that the subcellular localization difference contributes directly to the signalling specificity outcome. The *FZD1* gene has not previously been implicated in ASD or related disorders.

The *FZD9* gene is located within the Williams syndrome common deletion region of chromosome 7 and heterozygous deletion of the *FZD9* gene may contribute to the Williams syndrome phenotype. *FZD9* is expressed predominantly in brain, testis, eye, skeletal muscle, and kidney (http://www.ncbi.nlm.nih.gov/gene/8326). As previously mentioned in this work, WBS is a rare neurodevelopmental disorder that is characterised by an elfin facial appearance, a low nasal bridge, an unusually cheerful demeanour and ease with strangers, as well as developmental delays, strong language skills and cardiovascular problems (Berg *et al*., 2007).

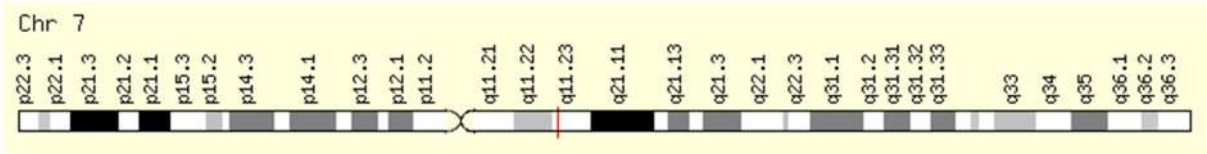

Figure 5.2: Red line indicates the position of the *FZD9* gene on chromosome 7 (http://www.genecards.org/cgi-bin/carddisp.pl?gene=FZD9&keywords=FZD9)

FZD9 gene is selectively expressed in the hippocampus throughout life with a strong expression in the neuroepithelium. *Fzd9* null heterozygous mice have increased apoptotic cell death and increased precursor proliferation during hippocampus development. Moreover, *Fzd9* mutants exhibit defects in learning and memory, reflecting hippocampus functional deficits.30 Overall, these evidences point out that *Fzd9* has an important role in hippocampus development, and therefore, *Fzd9* may be an intriguing candidate for the neurodevelopmental and behavioral phenotype of WBS individuals. Overall, these observations suggest that *BAZ1b* and *FZD9* remain two good candidate genes for the facial features and neurodevelopmental phenotype of WBS patients, respectively (Fusco *et al*., 2014).

5.5.2. Frizzled genes and the Wnt2 pathway

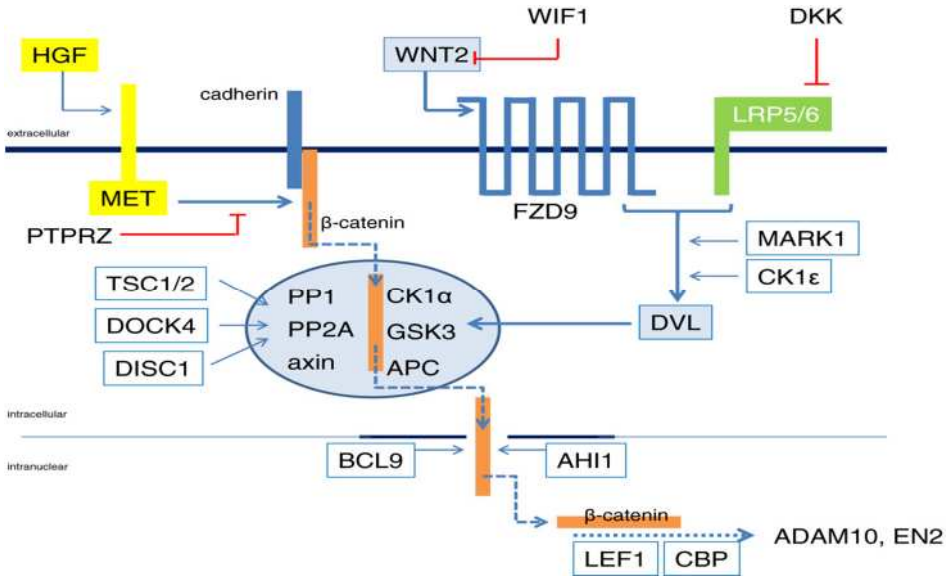

Figure 5.3: Schematic representation of the canonical Wnt2 pathway (Nelson and Nusse, 2004)

Figure 5.3 shows the Wnt signal transduction cascade. An estimated 19 Wnt members exist in mammals, and these act as sort range ligands to activate receptor-mediated signalling cascades (Nelson and Nusse, 2004). Proteins which act as cell surface receptors for Wnts are called 'frizzled' and the FZD family of genes is directly involved in this process. Activated frizzled receptors connect to several downstream pathways.

Wnt2 activates the 7-transmembrane-spanning Fzd9 receptor, which together with the co-receptor LRP5/6 activates dishevelled (DVL). Activated DVL inhibits the activity of the β-catenin "destruction complex" (indicated as a light blue ellipse). β-catenin is released from its complex with cadherin by the activity of the HGF receptor MET. When β-catenin is protected against destruction, it can enter the nucleus, bind the transcription factor LEF1 and co-factors to promote transcription of target genes like, for example, engrailed 2 (EN2). The functional consequence is an increase in cell growth and motility (Kalkman, 2012). The canonical Wnt pathway plays an important role in brain development (Galceran *et al*., Zhou *et al*., 2004, Lie *et al*., 2005 and Rosso *et al*., 2005) and synaptic function (Krylova *et al*., 2002, Packard *et al*., 2002 and Chen *et al*., 2006). It is for these reasons that mutations in the Wnt pathway are implicated in ASDs and indeed psychiatric disorders in general.

5.6. Transcription factors and candidate genes

Because transcription factors essentially initiate and control the regulation of genes, we thought it important to look at the commonality of transcription factors (TFs) across the candidate genes on chromosome 7. This might lead to clues about pathways that are common or might overlap between genes that confer phenotypical attributes, but may not necessarily be associated with autism and ASDs. For example, mutations in exon 2 of the GATA1 TF are present in almost all cases of Down-Syndrome associated transient myeloproliferative disorder or transient leukemia (Wechsler *et al*., 2003). We see from our Genecards ® data-

gathering that GATA1 TF is involved in the regulation of all the candidate genes from our STRING network, and we know that Down-Syndrome is present as a co-morbid disorder for any ASD cases. The CUTL1 TF, which is a homeodomain protein that is involved in gene expression, morphogenesis, and differentiation, also plays a role in cell-cycle progression (Nepveu, 2000). Genetic data from over 7,600 cancer patients shows that over 1% has the deactivated CUX1 which links to progression of tumour growth (Wong *et al.*, 2014).

In recent years, more and more genetic evidence has become available that points to a link between cancer and autism. According to a 2013 study at the University of Washington, 10% of children with mutations in a gene called *PTEN* –associated with breast, colon and thyroid cancer, among others-have autism. As with all genetic disorders, not everyone with the mutation develops autism or cancer, or any other neurological diseases or disorders. The findings have only been seen in a small test population thus far. It was Dr. Charis Eng, a cancer geneticist at the Cleveland Clinic, who first noticed a surprising incidence of autism in children whose parents had the *PTEN* mutation (pronounced p-10). Eventually, investigators discovered that the rate of autism was 10 percent, about 10 times what would normally be expected. At the same time, researchers found that another genetic disorder was even more likely to result in autism. That disorder, tuberous sclerosis, increases the risk for kidney cancer and a type of brain cancer; half of tuberous sclerosis patients had autism.

Although PTEN and tuberous sclerosis genes are not the same, they are part of the same network of genes that halt or hamper cell growth. Disabling PTEN or one of the tuberous sclerosis genes reactivates cell growth. One result can be cancer or tumours. Another can be abnormal wiring of nerve fibres in the brain and autism. In families with no genetic history of autism, it was found that autistic children had two to three times as many mutations that disabled a gene than their parents or siblings. These mutated genes were often involved in pathways that controlled cell growth. As seen in the results of this study, the FDZ genes are

significantly involved in Wnt pathway, which is in turn involved in cellular growth. The other candidate genes that have been investigated and summarized via the GeneCards database shows significant involvement in cell growth, pathways in cancer and influences on neurological function at a molecular level.

The bioinformatics methods used here have effectively shown the interconnectivity of genes on chromosome 7; indicating that common pathways and common transcription factors, as well as close proximity on the chromosome plays a role in the implication of these candidate genes in autism and related disorders. An even more surprising genetic link was found between cancer and autism, and as indicated in the literature review by Crespi in 2011, the genetic overlap between cancer and autism allows for further research into looking at more well-known cancer genes. Pharmalogical therapies that help to alleviate cancer symptoms might also be employed in the treatment of autism (Crespi, 2011).

5.7. Future studies

While the causes of autism are unclear, it is the general consensus that the causative factors of the disorder are linked to both environment and genetics. While the MMR vaccine theory has been completely ruled out now, other factors such as mercury levels in food, difficult pregnancies and viral infections have been reported as causative factors for autism and ASDs. These factors, along with strong evidence for a clear genetic link have illuminated the causes for the disorder. By researching the genetic factors of a disorder, it allows health care practitioners, teachers, parents and guardians to make better, informed decisions as to the treatment and care for individuals with autism and ASDs. Epigenetic studies will assist in the eventual development of specialized medication tailored to an individual's unique genetic makeup. One of the biggest problems experienced by children with ASD is 'leaky gut syndrome'; in which the gastrointestinal system and thus the diet is compromised by

inflammation in the bowel linings. A 2013 study has shown that incorporating certain doses of a human gut microbe in mice with autism-like symptoms helped reverse behavioural problems in these mice (Hsiao *et al.*, 2013).This treatment also reduced gastrointestinal problems in the mice that were similar to those seen in humans with autism. Future studies may make use of different methods of high throughput SNP detection such as DHPLC, sequencing by hybridization and UNG-mediated T-sequencing (Kwok and Chen, 2003).

Further studies can look at family cohort studies to investigate the hereditary pattern of autism across generations and may be able to predict the possibility of predisposition for autism in future generations. With the major advances being made in medical science and development of the Human Genome Project, the possibility of personalized medicine becomes an option for individuals with specific needs who are on the Autism spectrum.

Bionformatics results have yielded in two new candidate genes to look at for this study and can thus be further employed to continue to search pools of genes for more candidate genes. Employing computational methods in an era when overwhelming amounts of data is gleaned from traditional lab practices will only serve to aid in a better understanding – and organization – of said data.

REFERENCES

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

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Baio, J. (2012). Prevalence of Autism Spectrum Disorders — Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. *Morbidity and Mortality Weekly Report, 61*(3), 1-24.

Benayed, R., Gharani, N., Rossman, I., Mancuso, V., Lazar, G., & Kamdar , S. (2005). Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. *American Journal of Human Genetics, 77*, 851-868.

Chen, J., Park, C., & Tang, S. (2006). Activity-dependent synaptic Wnt release regulates hippocampal long term potentiation. *Journal of Biological Chemistry*, 281:11910-11916.

Chin-See-Chonga, T., Polaka, J., Veldscholtea, K., van den Worma, J., Mirzac, S., & Pabbap, I. (2015). The relation between GTF21 gene and positive emotonality in preschool children: A candidate gene study. *Erasmus Journal of Medicine, 4*(2), 43-46.

Crespi, B. (2011). Autism and Cancer Risk. *Autism Resarch,* 4:302-310.

Folstein, S., & Rutter, M. (1977). Infantile Autism: a genetic study of 21 twin pairs. *Journal of Child Psychology and Psychiatry, 18*(4), 297-321.

Galceran, J., Miyashita-Lin, E., Devaney, E., Rubinstein, J., & Grosschedl, R. (2000). Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. *Development*, 127:469-482.

Gameeldien, H. (2009). *Thesis: The Identification of Candidate Genes Using CDNA Microarray and the Analysis of Two SNPs of the Reelin Gene in a South African Austistic Population.* Cape Town: University of the Western Cape.

Itoga, S., Nanmoku, T., Uchimoto, T., Sunaga, M., Nezu, M., & Tomonaga, T. (2004). Comparative Analyses of Four Different Methods of Genotyping ALDH2. *Alcoholism: Clinical and Exprimental Research, 28*(Supplement s2), 117S-122S.

Kalkman, H. (2012). A review of the evidence for the canonical Wnt pathway in autism spectrum disorders. *Molecular Autism,* 3:10.

Kaye, J., del Mar Melero-Montes, M., & Jick, H. (2001). Mumps, measles and rubella vaccine and the incidence of autism recorded by general practitioners: a time trend analysis. *British Medical Journal , 322*(7248), 460-463.

Kopetz, P. B., & Endowed, E. (2012). Autism Worldwide: Prevalence, Perceptions, Acceptance, Action. *Journal of Social Sciences, 8*(2), 196-201.

Krylova, O., Herreros, J., Cleverley, K., Ehler, E., Henriquez, J., Hughes, S., et al. (2002). WNT-3, expressed by motoneurons, regulates terminal arborization of neurotrophin-3-responsive spinal sensory neurons. *Neuron*, 35:1043-1056.

Lamb, J. A., Moore, J., Bailey, A., & Monaco, A. P. (2000). Autism: recent molecular genetic advances. *Human Molecular Genetics, 9*(6), 861-868.

Libbey, J., Sweeten , T., McMahon, W., & Fujinami, R. (2005). Autistic disorder and viral infections. *Journal of NeuroVirology, 11*, 1-10.

Lie, D., Colamarino, S., Song, H., Desire, L., Hira, H., Lein, E., et al. (2005). Wnt signalling regulates adult hippocampal neurogenesis. *Nature*, 437:1370-1375.

Lord, C. (2011). Unweaving the Autism Spectrum. *Cell, 147*, 24-25.

Nelson, K. B. (2003). Thimerosal and Autism? *Pediatrics, 111*(3), 674-679.

Nelson, W. N. (2004). Convergence of Wnt, β-catenin, and cadherin. *Science*, 303:1483-1487.

Nepveu, A. (2001). Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene*, 270:1-15.

Packard, M., Koo, E., Gorczyca, M., Sharpe, J., Cumberledge, S., & Budnik, V. (2002). The drosophila Wnt, Wingless, provides an essential signal for pre- and postsynaptic differentiation. *Cell*, 111:319-330.

Peltola, H. (1998). No Evidence for Measles, Mumps, and Rubella Vaccine-Associated Inflammatory Bowel Disease or Autism in a 14-year Prospective Study. *Lancet, 351*, 1327-1328.

Perez-Iratxeta, C., Plaidwor, G., & Andrade-Navarro, M. (2007). Towards completion of the Earth's proteome. *The European Molecular Biology Organization (EMBO), 8*(12), 1135-1141.

Pfaff, D. W., Rapin, I., & Goldman , S. (2011). Male predominance in autism: neuroendocrine influences on arousal and social anxiety. *Autism Research, 4*(3), 163-176.

Piven, J., & Palmer, P. (1999). Psychiatric Disorder and the Broad Autism Phenotype: Evidence From a Family Study of Multiple-Incidence Autism Families. *American Journal of Psychiatry, 156*(4), 557-563.

Prasad, C., Prasad, A. N., Chodirker, B. N., Lee, C., Dawson, A. K., Jocelyn, L. J., et al. (2013). Genetic evaluation of pervasive developmental disorders: the terminal 22q13 deletion syndrome may represent a recognizable phenotype. *Clinical Genetics, 57*(2), 103-109.

Rodier, P., & Hyman, S. (1998). Early environmental factors in autism . *Mental Retardation*

and Developmental Disabilities Research Reviews , 4(2), 121-128.

Romero, P., Wagg, J., Green, M. L., Kaiser, D., Krummenacker, M., & Karp, P. D. (2004). Computational prediction of human metabolic pathways from the complete human genome. *Genome Biology, 6*, R2.

Rosso, S. S., Wynshaw-Boris, A., & Salinas, P. (2005). Wnt signaling through dishevelled, Rac and JNK regulates dendritic development. *Nature Neuroscience*, 8:34-42.

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

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*(2), 93-98.

Sigman, M., & McGovern , C. (2006). Improvement in cognitive and language skills from preschool to adolescnece in autism. *Journal of Autism and Developmental Disorders, 35*, 15-23.

Skuse, D. H. (2005). X-linked genes and mental functioning. *Human Molecular Genetics, 14*(Review Issue 1), R27-R32.

St-Hilaire, S., Ezike, V. O., Stryhn, H., & Thomas, M. (2012). An ecological study on

childhood autism. *International Journal of Health Geographics, 11*(44), 2-8.

Taylor, L. S. (2014). Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine, 32*(29), 2623-2629.

Various. (2013). *Autism Speaks*. Retrieved August 15, 2013, from www.autismspeaks.org

Various. (2014). *American Psychiatric Association*. Retrieved February 18, 2013, from http://www.psych.org/

Volkmar, F. (2005). Autism in Infancy and Early Childhood. *Annual Review of Psychology, 56*, 315-356.

Volkmar, F. R., & Rutter, M. (1995). Childhood disintegrative disorder: Results of the DSM IV autism field trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 34*, 1092-1095.

Wechsler, J., Greene, M., McDevitt, M., Anastasi, J., Karp, J., Le Beau, M., et al. (2002). Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nature Genetics*, 32(1):148-152.

Wolff, S. (2004). The history of Autism. *European Child and Adolescent Psychiatry, 13*(4), 201-208.

Wong, C., Martincorena, I., Rust, A., Rashid, M., Alifrangis, C. A., Tiffen, J., et al. (2014).

Inactivating CUX1 mutations promote tumorigenesis. *Nature Genetics*, 46:33-38.

Zeidán-Chuliá, F., & Moreira, J. (2013). Clostridium Bacteria and its Impact in Autism Research: Thinking "Outside. *Communication Disorders, Deaf Studies & Hearing Aids*, *1*(1), 1-2.

Zhou, C., Zhao, C., & Pleasure, S. (2004). Wnt signaling mutants have decreased dentate granule cell production and radial glial scaffolding abnormalities. *Journal of Neuroscience*, 24:121-126.

http://www.genecards.org

http://www.bork.embl-heidelberg.de/STRING/

GLOSSARY

Allele frequency – *Allele frequency is the number of copies of a particular allele divided by the number of copies of all alleles at the genetic place (locus) in a population. It can be expressed for example as a percentage. In population genetics, allele frequencies are used to depict the amount of genetic diversity at the individual, population, and species level.*

Autism Spectrum Disorder - *Autism spectrum disorder (ASD) and autism are both general terms for a group of complex disorders of brain development. These disorders are characterized, in varying degrees, by difficulties in social interaction, verbal and non-verbal communication and repetitive behaviours. With the May 2013 publication of the DSM-5 diagnostic manual, all autism disorders were merged into one umbrella diagnosis of ASD.*

Bioinformatics - *an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data.*

Candidate genes - *A candidate gene is a gene located in a chromosome region suspected of being involved in the expression of a trait such as a disease, whose protein product suggests that it could be the gene in question.*

Chromosomal hotspot – *chromosomes in the genome that carry a significant amount of candidate genes that have been positively linked to a particular disease or disorder.*

Chromosome 7 - *spans about 159 million base pairs and represents between 5 and 5.5 percent of the total DNA in cells.*

Epigenetics - *the study, in the field of genetics, of cellular and physiological phenotypic trait variations that are caused by external or environmental factors that switch genes on and off and affect how cells read genes instead of being caused by changes in the DNA sequence.*

Genotype - *The genotype is the genetic constitution of a cell, an organism, or an individual (i.e. the specific allele makeup of the individual) usually with reference to a specific character under consideration.*

Hardy-Weinberg equation - *The Hardy–Weinberg principle states that both allele and genotype frequencies in a population remain constant—that is, they are in equilibrium—from generation to generation unless specific disturbing influences are introduced. Those disturbing influences include non-random mating, mutations, selection, limited population size, "overlapping generations", random genetic drift and gene flow.*

Heterozygous - *Having different alleles at one or more corresponding chromosomal loci.*

Homozygous - *Having the same alleles at a particular gene locus on homologous chromosomes.*

Human Genome Project - *an international scientific research project with the goal of determining the sequence of chemical base pairs which make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and functional standpoint.*

Neurodevelopmental disorders – *Severe disabling conditions often associated with life-long impairment and are recognized to be the result of abnormalities in brain development due to both genetic and environmental/biological causes.*

Polymerase Chain Reaction - *In molecular biology, the polymerase chain reaction (PCR) is a technique to amplify a single or few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.*

Single nucleotide polymorphism – *frequently called SNPs, these are the most common type of genetic variation among people (e.g. 1%) in which a single nucleotide in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes.*

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) - *a wide-ranging visual database that assimilates information on known and predicted protein interaction. It offers a unique visual representation of very specific protein-protein interaction networks, which can be manipulated by adding more proteins, taking them away or changing their positioning within the network.*

Transcription factor - *a protein that binds to specific DNA sequences, thereby controlling the rate of transcription of genetic information from DNA to messenger RNA.*

Williams Syndrome - *a multi-system developmental disorder caused by the deletion of contiguous genes at the 7q11.23 position. It is a rare neurodevelopmental disorder that is characterised by an elfin facial appearance, a low nasal bridge, an unusually cheerful demeanour and ease with strangers, as well as developmental delays, strong language skills and cardiovascular problem.*

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

Appendix I: Cover letter

Department of Biotechnology

June/July 2011

Dear parents/ guardians

I am lecturing in the Biotechnology Department at the University of the Western Cape. I am currently heading a project which examines the genetics of Autism in a South African population. A study of this type has not been undertaken in South Africa but has been extensively done in other countries. We anticipate the results from this study will aid the understanding of genetics as a cause of Autism.

The project was started in 2005 and is ongoing. Some parents have already consented and have participated in the study but we require new and more samples because we are doing additional genetic investigations. Our research has generated one scientific paper in 2008 and we have presented our research at the Autism Conference in 2006 at the Convention Centre Cape Town. Our more recent research was presented at the joint Africa-South Africa Human Genetics Society Conference in 2011.

We welcome additional participants to the project and any families that are interested in family studies are most welcome.

In order to do this study, we need your consent to obtain information and biological material from you and your children. The biological material will be obtained from the scholars by collecting cheek cells from the inside of the mouth, using a swab. It is a painless and non-intrusive procedure. The whole procedure will only take a few minutes and will be done by properly trained individuals under hygienic conditions in their respective classroom in the presence of their teacher.

The project has been approved by the principal and the UWC Ethics Committee. I am most willing to meet with you and discuss the project. **Please fill in the attached consent form and return the form to the school promptly; if you have consented previously, we need you to complete the form again for ethical reasons.**

Yours in anticipation

……

Dr. Zainunisha Arieff

021 959 2214

zarieff@uwc.ac.za

Appendix II: Consent form

Clearly print: Full name of child……..

Date of birth ……

Grade ……..

School…….

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 childs' genetic material in the study outlined above.
2. I understand that the genetic material for analysis is to be obtained from blood and urine samples that I will 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:**

Dr Zainunisha Arieff

Senior Lecturer

Dept of Biotechnology

Email: zarieff@uwc.ac.za

Tel: (021) 9592214/5

Parent's/Guardian's signature: _____

Contact number(s):_____

Appendix III: Background of Project

Title of project: Genetics of Autism in South Africa

Autism, also known as the autistic spectrum disorder (ASD) or pervasive developmental disorder (PDD) is a complex, behavioral defined disorder of the immature brain.

There are 5 different sub-types of autism (Classic autism, Asperger disorder, disintegrated disorder, PDD – pervasive developmental disorder and Rett syndrome). It was shown using genetic tools that each sub-type is linked to one or more chromosomal regions (Muhle et al. 2004). To date, a minimum of 15 genes have been linked to this disorder (Yonan et al. 2003). Since autism has such a broad clinical spectrum, a genetic study will aid the understanding of the genes involved in the cause of the disorder. Thus far genetic research has been shown that specific genes and/or chromosomal regions have been associated with these disorders (Muhle et al. 2004, Wassink et al. 2004).

Although numerous clinical investigations have been performed on autism in South Africa, no articles have been published on genetic studies. For this reason the genetics group at UWC would like to investigate the possible causative genes, or to identify biomarkers involved in autism. To perform this type of research, we will require biological material such as cheek swabs and/or blood/urine from the children at the school, the unaffected children and the parents. This cannot be done without the written consent of the parents or guardians and the committee. In addition; the consent of the afore-mentioned bodies will support our application to the UWC Ethics Committee. The application is registered with the UWC Ethics committee and research funding has been acquired for this type of research.

The reasons for undertaking this type of research project are as follows:

(a) **Genetic diagnosis and treatment** – earlier diagnosis would lead to earlier treatment/therapy of affected children. The affected individuals will therefore be able to develop to their full potential with the proper treatment and therapy. Genetic studies will also allow us to establish the pathways which are affected in the disorder.

(b) **Education and awareness of the genetic causes of the disorder**. Education of disorder will remove the stigmatism associated with the disorder and more awareness could lead to greater funding being steered in the direction of this type of research. The awareness will also allow us to estimate the real incidence of the disorder in the various communities in South Africa.

(c) **South African research in a global context** – many genetic studies has been performed in the USA and UK but none on affected South African individuals. Our team will publish and present the research findings at both local and international conferences.

References:

Muhle R, Trentacoste V and Rapin I (2004). The Genetics of Autism. Pediatrics. 113:472-486.

Venter PA, Op't Hof J and Coetzee (1986). The Martin-Bell syndrome in South Africa**.** Am J Med Genet.;23(1-2):597-610

Wassink TH, Brzustowicz, Bartlett CW and Szatmari P (2004). The search for autism genes. MRDD Research Reviews 10:272-283.

Yonan AL, Palmer A, Smith KC, Feldman I, Lee HK, et al. (2003). Bioinformatic analysis of autism positional candidate genes using biological databases and computational gene network prediction. Genes, Brain and Behavior 2:303-320.

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