Université de Montréal

Finding genetic elements that lead to the autistic phenotype

Par

Robert Francis Fraser Gillis

Programmes de Biologie Moléculaire

Faculté de Médecine

Mémoire présenté à la Faculté de Médecine
en vue de l'obtention du grade de Maîtrise

Décembre, 2007

Université de Montréal
Faculté de Médecine

Ce mémoire intitulé :

Finding genetic elements that lead to the autistic phenotype

Présenté par :

Robert Francis Fraser Gillis

a été évalué(e) par un jury composé des personnes suivantes :

Dr Patrick Cossette
président-rapporteur

Dr Guy Rouleau
directeur de recherche

Dr Eric Shoubridge
membre du jury

Abstract

The identification of genes whose mutations lead to genetic disorders, is the most direct and efficient method to understand the molecular components of diseases with a genetic basis. In this thesis, an attempt is made at unraveling the mysterious nature of the genetic architecture of autism by comparing the traditional polygenic model for the genetics of autism with a model based on *de novo* mutations in multiple genes that is largely unaccepted by the scientific community, followed by a scientific write up of two experiments.

Under a polygenic pretense, a transmission disequilibrium test (TDT) was performed on 185 trios. Twenty six SNPs spanning 820Kb of the $GABA_A$ receptor subunit cluster at 15q12 were selected based on minor allele frequency >0.05, spacing, and coverage. This led to the identification of three SNP's in transmission disequilibrium: rs6606905, rs2078439, as well rs2873027 whose association has already been shown in previous studies.

Polyalanine trinucleotide repeats are under increasing scrutiny as they have been discovered to be involved in an increasing number of genetic disorders. In this study, 56 alanine tracts were screened and compared between 192 patients and 192 controls. This led to four distinct groups: one group which revealed no variation in cases or controls, one group in which four tracts revealed a high degree of polymorphism for which no identifiable wild type could be discerned, one group whose variants were matched in cases and controls, and one group that contained four alanine tracts from *ZIC2*, *TBLXR1*, *ARID1A*, *FOXD3* and *POU3F3* whose rare variants were unmatched in a larger number of controls and may warrant future functional studies.

Keywords: genetics of autistic disorder, *de novo* mutation, polygenic inheritance, association study, $GABA_A$ receptor subunit, alanine repeats/expansion/contraction

Résumé

L'identification de gènes dont les mutations mènent à de troubles génétiques, est la méthode la plus directe et efficace de déterminer et comprendre les composants moléculaires de ces affections. Cette thèse tente de dénouer la nature mystérieuse de l'architecture génétique de l'autisme, en comparant le modèle polygénique traditionnel avec un modèle basé sur des mutations *de novo* dans des gènes multiples, qui n'est pas largement accepté dans la communauté scientifique. Par la suite, deux expériences sont exécutées, l'une basée sur un modèle polygénique et l'autre basée sur un modèle de mutations *de novo*.

Une étude de déséquilibre de transmission a été effectuée sur 185 trios. Vingt-six SNPs comprenant 820 Kb de l'agrégat du récepteur sous-unité $GABA_A$ à 15q12 ont été sélectionnés basées sur leurs fréquences allèles mineures (>0,05) et leur espacement. Trois SNPs avec une association significative ont été identifiés : rs6606905, rs2078439, ainsi que rs2873027 dont l'association a déjà été démontrée dans des études précédentes.

Pour la deuxième expérience, 56 tractus d'alanine ont été examinés et les variants ont été comparés entre 192 patients et 192 contrôles. Quatre groupes distincts ont résulté : un groupe n'ayant démontré de variation ni dans les cas affectés, ni dans les contrôles, un groupe dans lequel quatre tractus ont révélé un degré élevé de polymorphisme pour lequel aucun allèle de type sauvage ne pouvait être discerné, un groupe dont la fréquence des variants rares était semblable dans les cas affectés et les contrôles, et un groupe dont les membres pourraient mériter de futures études fonctionnelles, comme quatre tractus alanine contenus dans *ZIC2, TBLXR1, ARID1A, FOXD3* et *POU3F3* possédaient des variants génétiques non-observés dans les contrôles.

Mots clés : désordre autistique, mutation *de novo*, héritage polygénique, études d'association, récepteur sous-unité $GABA_A$, expansions/contractions des tractus d'alanine

Table of Contents

List of Abbreviations

A	Adenine
AA	Amino Acid
AD	Autistic Disorder
ADI-R	Autism Diagnostic Interview Revised
ADOS-G	Autism Diagnostic Observed Schedule Generic
ARX	Aristaless X
AS	Asperger Syndrome
ASD	Autism Spectrum Disorder
C	Cytosine
CGH	Comparative Genomic Hybridization
CNV	Copy Number Variation
DNA	Deoxyribonucleic Acid
DNM	*De novo* mutation
DSM-IV	Diagnostic and Statistical Manual
DZ	Di-Zygotic
FC	French Canadian
G	Guanine
GABA	Gamma amino butyric acid
GABA(A)	GABA Receptor Type A
ICD-10	International Classification of Disease
Kb	Kilobase
LOD	Log Of Odds
MECP2	Methyl Cytosine Binding Protein 2
MMR	Measles, Mumps, Rubella
mRNA	Messenger Ribonucleic Acid
MZ	Mono-zygotic
NF1	Neurofibromatosis type 1
NLGN	Neuroligin
NRXN	Neurexin
OPMD	Ocular Pharyngeal Muscular Dystrophy
PABPN1	PolyA Binding Protein 1
PCR	Polymerase Chain Reaction
PDD-NOS	Pervasive Developmental Disorder Not Otherwise Specified
RNA	Ribonucleic Acid
SHH	Sonic Hedgehog
SNP	Single Nucleotide Polymorphism
T	Thymine
TDT	Transmission Disequilibrium Test
TRD	Trinucleotide repeat disorder
TSC	Tuberous Sclerosis
UTR	Untranslated Region
XLMR	X-Linked Mental Retardation

List of Tables and Figures

List of Tables

List of Figures

Acknowledgments

I would like to thank Dr. Guy Rouleau first for giving me the opportunity to work in his lab, and for providing me with key insight into problem solving that I will take with me wherever I go. I would also like to thank Claudia Gaspar who also played a key role in my supervision and helped all of my work and projects progress during my time at the lab.

Julie Gauthier also helped me throughout my studies and her hard work initiated this project in the lab and provided the samples upon which several other projects are based. Nathalie Girard provided me with great lab supervision upon my arrival. There are also several people that worked on these projects with me such as Judith St. Onge whose hard, meticulous work was integral to the progression of the alanine project, as well as Lan Xiong, and Nathalie Gauthier. Sandra Laurent taught me many lab skills and showed much patience, kindness and care during the process of the GABA project and Jean Baptiste Riviere was very helpful in teaching the UNPHASED software package, SNP selection and general advice.

Several people helped me at many points during my time at the lab such as Paul Valdmanis who was always able to provide insightful comments on the projects I was involved in, as well as current scientific literature in general as well as Dominique Verlaan who was very helpful as a project organizer and mentor. I would like to thank everyone in general at the Rouleau lab, all students and

Furthermore, I would like to thank all of my family and friends whose guiding hands and support helped me get through many of the steps that led to this point, particularly the four people that guided my growth and development: my parents as well as my sister Heather Gillis and her husband Darcy Trefiak. I would also like to thank Marina Amato who has provided me with motivation in everything I do.

1.0 General Introduction

In 1943, Leo Kanner published a case report describing 11 children with subtle phenotypic similarities he termed "autistic disturbances of affective contact"[1]. In this report, Kanner compared and contrasted many phenotypic and background commonalities among these children taking note of several features, many of which have developed into the modern diagnostic guidelines for autism.

One of Kanner's contemporaries, a pediatrician from Vienna named Hans Asperger, described children with similar developmental difficulties one year after Kanner, also using the term autistic to describe his patients[2]. This may initially seem like a startling coincidence, however autism or autistic had been a commonly used term in psychology dating back to 1910 to describe human behaviour in which the patient is in social withdrawal, a feature used primarily to describe a phenotypic symptom of schizophrenia[3]. The concurrent clinical characterization of autism within this short time frame should not be recognized as either a sudden outbreak of a new disorder, or that patients modernly diagnosed as autistic did not exist prior to this. It simply represented a shifting cultural climate in the psychological features that psychiatrists focused on to categorize and diagnose human mental disorders. These guidelines have experienced several revisions over the years with many more likely to come as continuing research in multiple fields will shed light on this mysterious condition and begin to homogenize patients currently diagnosed with autism into different clinical categories.

Due to the work of anthropologist Margaret Mead in the early 1900's[4], it was a commonly held belief that the human brain and associated personality was a blank slate, with little to no contribution from genetic factors[5]. Therefore, most psychiatric conditions had a large bias towards examining the parents of affected individuals to determine the cause of the disorder. When autism was initially described, the search began for the environmental component that led to this disease. Unfortunately, scientific rigor in the field of psychology was applied loosely during this period, therefore weak observations often led to dominating theories regarding the cause of autism. Observations of socially reticent mothers of autistic children led to the "refrigerator mother" theory of autism, a theory that posited that infants that do not get the proper care from the mother will experience developmental delays resulting in autism. In 1957, Leon Eisenberg reported that often the fathers of autistic children were socially isolated, which led to the idea that perhaps the behaviour of both parents could lead to the autistic phenotype in their children[6]. This idea was held as common belief for much of the mid 1900's and was not completely abolished until early twin studies in 1977 had definitively shown a genetic basis for the disorder[7], although the concept likely remained embedded within the general public for several years afterwards.

Autism has become the focus of much more publicity and research as data amassed between 1991-1997 resulted in a reported 556% increase in the prevalence[8], a dramatic increase that sparked public concern. A crisis developed in the late 1990s, as several individuals presented anecdotal evidence that pointed

to the measles, mumps, rubella (MMR) vaccine as the causal agent behind the outbreak. However, there is no epidemiological evidence to suggest that the MMR vaccine confers any risk to developing the autism phenotype[9]. Afterwards it was suggested that perhaps thimerosal, a mercury containing preservative used in vaccines, was associated with autism[10], although recent studies have refuted this hypothesis[11, 12]. Furthermore, no evidence at this point supports an increased incidence of the disease, as studies reporting an increased prevalence did not control for changes in diagnostic criteria for autism that took place in 1994[13].

Currently, the general term autism has been replaced with autistic spectrum disorder (ASD) and is diagnosed in accordance with the classifications set forth in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision 7* (DSM-IV) and the *International Classification of Diseases, Tenth Revision (ICD-10)* [14, 15]. According to these diagnostic manuals, ASD is characterized by impairments in three behavioral domains: 1) social interaction; 2) language, communication, and imaginative play; and 3) range of interests and activities. ASD has been further subdivided into five subtypes according to the varying clinical presentations of the above three phenotypic characteristics: 1) the prototypical Autistic Disorder (AD) 2) Asperger Syndrome (AS), which is generally recognized as a milder form of AD without the associated language delay 3) Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) 4) Childhood Disintegrative Disorder, a rarely diagnosed condition where a childhood experiences dramatic regression between two and ten

years of age after normal development and 5) Rett Syndrome (RS) a genetic disorder of postnatal brain development, caused by a single-gene defect predominantly affecting girls.

This thesis will focus on the discovery of genetic elements that lead to the phenotypes in categories 1-3, AD, AS and PDD-NOS. Although psychiatrists have developed these diagnostic criteria in an attempt to categorize and homogenize patients in a meaningful way, there is no evidence that these classifications homogenize patients based on the underlying genetic mutations involved. Due to the subtle differences between these categories the profound impact of differential diagnosis, as well as differing clinical presentations that depend on the age of the child during examination, the same mutation can lead to a variable diagnoses on the autistic spectrum. In fact, this has been shown for the known genetic mutations that lead to the AD phenotype. In one instance a sibling pair discovered to share a single gene defect in the *NLGN*3 gene, each received a different diagnosis; one with AD, the other with AS[16]. In another example, a large multiplex pedigree with 13 affected individuals shared a fully penetrant mutation in *NLGN4* with perfect x-linked segregation, which resulted in multiple diagnoses including AD, PDD-NOS as well as X-linked mental retardation (XLMR)[17]. Therefore, for the purpose of discovering genes that lead to the autistic phenotype, the three clinical categories that are focused on in this thesis can be treated as one, and will be referred to as autistic disorder (AD).

1.1 Pathology

Symptoms of AD generally begin to appear between 12-36 months after birth. This is usually preceded with no phenotypic symptoms or signs until the age of onset, which makes this disorder both mysterious and distressing. However, the age of onset reveals the importance of this molecular developmental time period of the nervous system, as it coincides with the appearance of spatial and temporal transcriptional cascades that lead to remodeling and elaboration of neuronal circuitry[18,19]. The discovery of genes that lead to the AD phenotype should provide insight into this process. To date, neuroscientists have discovered very few physical abnormalities in the brains of autistic children, none of which are hallmark signatures common to all patients diagnosed with AD. Increased hippocampal and amygdala volumes have been reported in some children and some adults with autism[20]. Furthermore, damage to the amygdala has been hypothesized to play a role in the onset of the disorder[21, 22], although this theory has received less attention in recent years. Hyperserotonemia has been observed in children as well as families of children with AD[23, 24], which has led to several examinations of polymorphisms within the serotonin transporter gene[25-28], although these studies have failed to conclusively determine any association with AD[28]. Macrocephaly (head circumference > 95th percentile) is observed in approximately 20% of patients with AD[29], as well as epilepsy in approximately 30%[30, 31].

1.2 Evidence that genetic factors lead to the AD phenotype

Familial aggregation studies have shown that the sibling recurrence rate of AD is approximately 2-8%, with the most commonly accepted value currently at 4.5%[32, 33], whereas the population prevalence is currently estimated at 0.5-0.6%[34]. This information reveals that there is a higher tendency for autism to cluster in families, however does not point exclusively to a genetic origin, as it does not rule out a shared environment unless combined with data from twin studies. To date, three twin studies of AD have been performed, each using small sample sizes[7, 35, 36]. Combining these three studies, the total number of twin pairs is 66, with 36 monozygotic (MZ) pairs and 30 di-zygotic (DZ) pairs. The average MZ concordance rate is 70%, with values ranging from 60-90%, compared with a DZ rate of 0%. The observed rate of 0% is a chance finding that likely results from the small number of DZ pairs in these studies. This is based on the observance that concordant DZ twin pairs do exist, therefore the DZ concordance rate is likely closer to the sibling recurrence rate of ~4-5%, considering DZ twin pairs share the same environment and proportion of genetic content. Although a shared environment cannot be eliminated for studies involving sibling recurrence, the much higher concordance rate between MZ and DZ twins suggests that genetic factors are the dominant factor in the development of the disease.

Finding the genetic elements that lead to human disease represent a powerful "bottom up" approach to understanding human disorders that have a genetic component. Insight into the genetic mutations responsible will provide

information on the molecular pathology of human disease and also be seamlessly integrated into neurological, physiological and clinical data. The discovery of these genetic elements that lead to the autistic phenotype is likely to be the most direct route towards early detection, as well as assembling a permanent cure that alleviates all associated autistic phenotypes.

The most important consideration when attempting to discover genetic factors that lead to disease, is to first understand the inheritance mechanism or genetic architecture of the disorder in order to determine the appropriate gene identification strategy. For example, in the case of a genetic condition resulting from a simple single-gene mendelian disorder, a family based linkage study can be performed and combined with other linkage studies to refine the segregating chromosomal region containing the mutation. However, it was clear from the beginning of the twin studies that AD is not a simple mendelian disorder, as one mutation in one gene would result in a MZ concordance rate closer to 100% and DZ concordance closer to 50%. The actual MZ and DZ concordance rates have led researchers to pursue gene identification strategies based on a polygenic model with as many as 2-15 genes interacting to produce the AD phenotype that dominates the literature today. However another model based on *de novo* mutations has been largely ignored by the scientific community and will be presented and contrasted within this thesis.

1.3 The polygenic model, common disease/common variant

The polygenic model for autism posits that multiple functional allelic variants are present within the global gene pool that each contribute to the autistic phenotype. The exact proportion, distribution, and genes involved within each patient will vary under this model, producing a large array of phenotypes under the autistic spectrum. This model was originally intuitive to the scientific community given the data from twin studies: if the MZ concordance is ~90% and the DZ concordance/sibling recurrence is ~4-5%, it is quite logical to assume a polygenic model for autism, given that the concordance will be essentially halved for each additional locus involved. For example, if five loci are involved then the probability of a twin pair being concordant for autism is MZ=100%, DZ= $(0.5)^5$ = 3.125%, which is very close to the observed value. It would also follow that in the twin that failed to meet diagnostic threshold there may be a broader spectrum. For example, if this twin inherited 2/5 of the AD conferring alleles rather than the entire suite, he or she may be partially affected. This is also the case as the twin studies report that discordant twins often show symptoms of the broader phenotype. The intuitive nature of this model for AD has led to countless sib-pair linkage and association studies, however these strategies have yet to identify a causal genetic variant or mutation that leads to the AD phenotype.

1.4 The multi-locus monogenic model dominated by low transmission of *de novo* mutations (DNM model)

The limitations of current sequencing technologies have made it difficult to accurately assess the rate and role of new mutations in complex genetic disorders. Although obtaining precise, multi-generational genome-wide sequence information is not currently plausible, it is still possible to examine the empirical data that are directly influenced by the underlying genetic architecture, to deduce the potential contributions *de novo* mutations may provide to the overall proportion of diagnosed AD cases.

The DNM model posits that the genetic basis of AD can be explained by monogenic *de novo* mutations in numerous genes across several molecular pathways in the absence of other predisposing variants. These *de novo* mutations would be highly penetrant, and rarely transmitted to offspring due to the strong negative selection against the phenotype. It has been well established that large scale *de novo* chromosomal anomalies occur frequently in patients with AD, explaining approximately 5-10% of cases[37, 38], including a recent report showing a strong association with *de novo* copy number variation (CNV)[39]. It is also believed that techniques providing higher resolution of CNV will likely identify smaller *de novo* copy number variants responsible for an even higher overall proportion of AD cases[40]. The remaining unidentified cases are currently believed to result from polygenic inheritance with multiple common and rare allelic variants contributing to the phenotype. However, little consideration has

been given to the possibility of even higher resolution *de novo* mutations (point mutations, expansions/contractions of repeated genomic segments of DNA, *de novo* epigenetic mutations).

To date, the mutations that have been discovered in individuals displaying the AD phenotype have arisen *de novo*, and are primarily involved in synaptic establishment, maintenance and plasticity[16, 17, 40, 41]. Mutations have also been described in transcription factors expressed in the fetal brain in the case of ARX[42], RNA binding translational repression in the case of *FMR1* (the gene responsible for fragile-X syndrome[43, 44]), transcriptional regulation in the case of *MECP2* and Rett syndrome, as well as the tumour suppressor genes *TSC1* and *TSC2*[44, 45]. These represent dysfunction in a vast array of molecular pathways and functional structures, all leading to the same general phenotype, implicating thousands of potential candidate genes. For example, using conservative numbers, there are approximately 5,000 genes expressed at the synapse that contribute to synaptic function, countless more are brain-expressed embryonic transcription factors, translational repressors and tumour suppressors. Certainly, mutations in all of these candidate genes and their downstream targets are not likely to produce the AD phenotype, however there is clearly a potential for the same phenotype to arise as a result of monogenic mutations in any one of several hundred genes.

Although not all amino acid substitutions will have a functional effect, a significant fraction will lead to disease. It has been estimated that in every zygote, there are approximately 1-3 new deleterious mutations that lead to an altered

amino acid per genome, this is on average 1 new mutation per 10,000 genes/zygote[46,47,48]. Therefore, for a disease such as AD that may result from dysfunction in any one of hundreds of different genes, new mutations may be responsible for a significant fraction of cases. Examining conditions for which genes have previously been identified, it is evident that new mutations are common. For example, 1/6,000 live births harbor a novel mutation causing neurofibromatosis type 1 (NF1)[49,50]. The frequency of new point mutations in Duchenne Muscular Dystrophy is similar: 1/10,500 live births[51]. These are large genes allowing for a high mutation rate, however their total genomic size is a small fraction compared to the genomic size of the hundreds of potential genes that may all produce the same general AD phenotype. Examining a smaller gene, Rett syndrome is a clinically characterized phenotype within the ASD spectrum, and results from mutations in the small *MECP2* gene involved in transcriptional regulation. The incidence of Rett syndrome is one in 10,000-15,000 females, and 99-99.5% of all cases are sporadic resulting from new mutations; this represents a new mutation rate of one in 5,000-7,500 live births for this small gene of 498 amino acids[52,53,54].

It has been estimated that the mutation rate of a repetitive stretch of genomic DNA is approximately 100,000 fold more frequent than common point mutations[55]. These stretches can code for poly-amino acid tracts, which are found in hundreds of genes throughout the genome, or for regulatory elements such as *FMR1*. For example, alanine tract expansions in the *ARX* gene, a transcription

factor expressed in the fetal brain, produces a broad spectrum of disorders including epilepsy, mental retardation and the AD phenotype[42, 56]. In the case of *FMR1*, an expansion of the repeat within the upstream regulatory region leads to hypermethylation and silenced gene expression[57]. Given the ubiquitous dispersal of repetitive stretches of DNA located throughout the genome within coding regions or regulatory genetic elements, *de novo* expansions or contractions of these repetitive elements could account for an appreciable fraction of AD cases as well.

The male:female ratio of mutation rate is estimated at about 4–6:1, presumably due to a higher number of germ-cell divisions with age in males. Therefore, the DNM model would predict that *de novo* mutations arising in the germ line would more frequently come from males, particularly with advancing paternal age[58]. The higher paternal origin of *de novo* mutations has been shown for many diseases, including Apert syndrome[59], Crouzon syndrome[60], NF1[61], Alexander disease[62] and Rett syndrome[63,64]. Some authors have even predicted a high incidence of male-derived novel mutations in human mental disorders[65,66]. A recent review has actually highlighted the association between advancing paternal age and autism [67], just as the DNM model predicts.

Considering the complete penetrance of the previously described mutations discovered in patients diagnosed with classic forms of AD[16, 17, 41], it is likely that these mutations produced phenotypic effects without the presence of other genomic variants. Taken together, these examples show that *de novo* mutations

can occur with a high enough frequency to explain the relatively high incidence of AD, as well as act dominantly to do so. Mutations in known genes for AD currently explain a small fraction of cases. Under the DNM model, over time, more monogenic mutations will be discovered in different genes, each contributing a small portion of the overall disease incidence. Eventually, a rare causal monogenic mutation could be identified for the majority of patients diagnosed with AD.

2.0 The strength of each model at predicting and explaining factors influenced by underlying genetic architecture

As previously mentioned, determining which model more accurately explains all of the data influenced by the underlying genetic architecture is extremely important as it will have a profound influence on the type of strategy employed to discover these genetic elements. Therefore, before beginning it is necessary to examine all of the recent literature relevant to the field of genetics and examine the power each model has at explaining the data.

2.1 Twin studies and sibling recurrence

As mentioned above, the MZ twin concordance rate for AD is 70-90% and the DZ concordance is similar to the sibling recurrence rate of 0-10%, with the extreme values likely representative of smaller sample sizes[7, 35,68]. Given these numbers, there are two possible explanations for the data; a polygenic model involving several genes interacting with environmental factors, or the *de novo*

mutations model mentioned above. An example of how the DNM model could hypothetically provide the observed sibling recurrence rate is shown in Table I.

Table I : Hypothetical example of an AD patient cohort that leads to the empirically observed values for sibling recurrence.

Inheritance	# Families	Recurrence	No Recurrence
Autosomal Dominant variable expressivity	**15**	**3**	**12**
Autosomal Recessive	**12**	**2**	**10**
X-Linked	**18**	**3**	**15**
Mutation in Imprinted Gene	**7**	**3**	**4**
***de novo* in affected**	**125**	**0**	**125**
Total	**178**	**11**	**166**

Sibling recurrence 11/178 = 6.2%

The majority of these *de novo* mutations will occur in the affected individuals and not be transmitted; however, they can be transmitted from unaffected parents to multiple affected siblings as a result of maternal transmission of mutations on the X chromosome to male offspring[16, 17] or of gonadal mosaicism[40, 41], which would all result in higher sibling recurrence rates. In fact, the transmission pattern of the four genes discovered to be mutated in patients with AD (*NLGN3*, *NLGN4*, *SHANK3*, and *NRXN1*) all fall within one of the above categories. Furthermore, it is also theoretically possible for transmission to occur from asymptomatic parents via autosomal recessive inheritance patterns, as well as through mutations in imprinted genes although specific examples of such are yet to be discovered.

Although perfect concordance is not observed between autistic MZ twins, this does not necessarily imply that a contribution of environmental factors must play a role in producing the AD phenotype in the affected twin. With the DNM model, discordance would occur as a result of X-linked inactivation[69], predicting that female MZ twin pairs will be more discordant than males, we are unaware of any data reporting separate concordance rates for male or female twin pairs.

Discordance between MZ twins would also occur from somatic mosaicism for a *de novo* mutation that occurred early in development in only one of the two developing zygotes, or autosomal dominance with variable expressivity potentially resulting from stochastic events during embryogenesis.

These novel considerations of the DNM model present implications for genetic counseling. Parents of autistic children are often informed that the sibling recurrence rate of autism is between 5-10% when planning for their next child. The DNM model predicts that the probability of future siblings being affected with autism is much higher when a second child is born with AD, as it shows that the underlying defect is being transmitted from one of the parents via one of the unknown transmission patterns mentioned in Table I (page 19).

2.2 Wide clinical spectrum associated with AD

Psychiatrists and psychologists are burdened with the task of collecting clinical information and developing diagnostic criteria for brain disorders. When examining a transient position on a scale of continuity, such as a poorly

developing child with global brain dysfunction and a fluctuating vocabulary, it becomes difficult to develop clinical categories and thresholds that will apply on the genetic level. This is especially true when attempting to account for the shifting landscapes of brain development into adolescence and adulthood. Therefore, a wide clinical spectrum is present in ASD, where some individuals are high functioning college students with above average IQs[70], some can perform astonishing memory feats and calculations[71], while others are non-verbal with severe mental retardation involving self injury[8]. Co-morbidity with epilepsy is estimated at approximately 30%[30, 31], while a strict definition of macrocephaly is observed in approximately 20% of children diagnosed with AD[29].

This wide variation in clinical presentation is currently explained with polygenic inheritance of many common alleles at multiple loci. These alleles are thought to combine in numerous permutations to produce a vast array of harmful phenotypes, the logistics of which will be confronted in a later section.

When accounting for this clinical variation with the DNM model, the wide variation in clinical presentation would be explained by new mutations in numerous genes that function in multiple molecular pathways producing phenotypes on a large spectrum. This spectrum makes it impossible to determine a boundary or threshold that will homogenize patients on the genetic level.

The clinical difficulty of homogenizing patients is even present in families where a single mutation is responsible for the phenotype. The large family presented by Laumonnier et. al, displayed several individuals with various clinical

alleles to produce the AD phenotype, or are they more simply a reflection of the clinical difficulties surrounding the ability to define AD on a level that is consistent with the underlying biological origin? Linkage analysis of this family identified a single mutation in *NLGN4* segregating in perfect X-linked transmission through all affected individuals, including obligate carriers with 100% penetrance[17]. This seems to indicate that this mutation is acting monogenically, and the diagnostic disparities are more likely confounded by varying stages of development at diagnosis and clinical heterogeneity rather than the presence of additional alleles that produce the AD phenotype.

Admittedly, the DNM model does not apply to high functioning adults leading relatively normal lives. Many individuals are suggested to fall within the broad ASD spectrum based on personality characteristics that are held in common with some patients diagnosed with AD, such as social difficulties and the inability to accept changes in their lifestyle. Although the wanton eye may be able to identify parallels on the phenotypic level between a high functioning asocial college student and an individual with mental retardation that requires constant supervision, this does not necessarily mean that there is a connection between their conditions on the genetic level. Many of these high functioning adults diagnosed with AD may carry a specific combination of common variants shrouded by environmental events that lead to social difficulties, just as any other individual may have a mix of common variants and environmental components that lead to their own behavioural quirks. It thus becomes a philosophical question

as to whether their behaviour lies within the *normal* range of human variation. This model is only meant to apply to individuals with a clear phenotypic developmental delay and mental retardation that progress into adulthood.

2.3 Failure of linkage and association studies to identify genes involved in AD

There have been a very large number of association studies performed with the goal of identifying the common alleles that contribute to the autistic phenotype under the pretense of a polygenic model[72]. None of these have been able to identify a single gene or variant with any considerable level of consistency. There have also been numerous linkage studies combining multiple small families that have produced inconsistent results, with linkage peaks spattered across the entire genome[72]. Although some regions have been replicated across studies, none of them have ever been replicated consistently in all studies involved. The explanation provided for the inadequacy of these studies to produce digestible results involves the addition of more and more allelic variants that contribute a smaller and smaller portion of the phenotype. The information gleaned from combining small families to perform linkage studies does in fact tell us that multiple loci are involved. Even under a monogenic model, a small number of genes will produce consistent linkage peaks if the majority of cases resulted from low transmission of *de novo* mutations occurring on different haplotypes at a small number of loci. However, since consistently replicated

linkage peaks have not been observed, multiple loci throughout the genome must be involved, whether one considers a polygenic model or the DNM model.

A problem arises when one considers the utility of using association studies to identify AD genetic variants under both models. When viewed from the perspective of a polygenic model, association studies should be the ideal method to identify the loci that contribute to the AD phenotype, especially considering the recent validation of the technique in other complex traits[73-75]. However, it should be noted that the genetic variants identified using this technique were represented in adult onset disorders, in which little selection occurs against the phenotype.

Assuming the polygenic model explains the majority of cases, the inability to globally replicate association studies must result from either choosing the wrong loci to study, or from a high number of loci that only contribute small percentages of the phenotype which can vary based on the cohort of patients involved in the study. The solution to this problem would be to increase power by adding more patients to these studies as well as the investigation of many more loci in an effort to identify these variants. However, under the DNM model, it is evident that association studies will fail to identify these variants, as new penetrant mutations that are strongly selected against will never become in linkage disequilibrium with any genomic marker.

The inability of association and small family linkage studies to identify genetic variants can always be explained by adding additional contributing loci under a polygenic model. The question that must be posed is how much further

should this approach be taken before an alternative genetic mechanism is considered?

2.4 Consistent global prevalence

Obtaining complete, global disease prevalence data for AD has not been accomplished, as diagnostic disparities coupled with little clinical investigation of AD in certain regions is difficult. However, to date epidemiological studies have failed to conclusively reveal an uneven distribution of AD from any ethnic or socioeconomic background[34, 76]. One of the many perplexing aspects of AD is its genetic basis, however an un-even global distribution is not observed. Under a polygenic model this could only be reconciled by the existence of allelic variants that confer a susceptibility to AD dating back to a common founder population for all of humanity, otherwise many pockets of the globe would be void of individuals with AD. These variants would have had to be spread globally through small human populations as they migrated to various locations around the world. Information gleaned from the International Hapmap project has shown that a considerable amount of human genetic variation is common between several populations around the globe, indicating that a substantial proportion of variants date back to a common founder population[77]. Therefore, on the surface the concept seems plausible.

In small isolated founder populations, such as those likely found in our ancestral environment, the genetic variation is reduced with a concomitant

increase in homozygosity at all loci. This would mean that under a polygenic model, the allelic variants that produce the AD phenotype must have combined more frequently to produce individuals that would be modernly diagnosed as autistic. In order for these variants to have remained globally consistent, there could not have been any selection acting against them, nor any fixation by genetic drift during small population migrations and settlements, otherwise an uneven disease distribution would be presently observed.

Considering the refusal of female chimpanzees to raise offspring that fail to develop a maternal bond, it is conceivable that there was a high degree of selection against the autism phenotype in our evolutionary ancestral environment. It has even been proposed that parental selection played a large role in human evolution, as raising children would have been difficult and expensive for hunter gatherer communities[78]. It is noteworthy that the decision to care for and raise an autistic child is still difficult for many parents in the recent western cultural environment, exemplified by the number of autistic children that are placed in psychiatric homes. In our modern environment, autistic individuals frequently survive to adulthood, yet strong negative selection still occurs against the phenotype evidenced by the low AD reproductive rates[79].

A gene pool is a reflection of successful reproductions between ancestral genomes that were composed of allelic variants whose collective actions conferred advantages to their historical possessors. The persistence of allelic variants under negative selection throughout evolutionary history in all human

populations disagrees with evolution at the genetic level and would require an explanation. Considering the low transmission of the AD conferring allelic variants, it is thus very unlikely that a polygenic model could explain an even disease distribution around the globe.

It is possible that particular variants present in the maternal and paternal lineage functioned properly/co-operated well separately in the ancestral gene pool, but can not execute the proper function when co-expressed in the same genome. For example, a poor fit between a common neurotransmitter variant from one population, and a variant present within its receptor in another population could result in altered neurotransmission. In fact, this would be predicted in a species with many isolated populations where there is a considerable amount of genetic variation within populations rather than between populations. However, this is not the case with humans, and there is no evidence suggesting that parents of mixed ethnic backgrounds are more likely to have autistic children than families with homogenous ethnic backgrounds.

Analyzing this observed feature using the DNM model, an even disease distribution is actually predicted, as mutation rates should not be considerably different between any specific global populations.

2.5 Skewed male to female AD ratio

It is already commonly understood that males are more prone to developing the autistic phenotype than females, by a ratio of ~ 4:1[32]. Under a

polygenic model, it has been suggested that intrinsic differences between male and female brains on the psychological level may account for the reasons why one sex is more vulnerable to the AD conferring allelic variants[80]. However, there is currently no strong evidence to support this. Furthermore, there has been little evidence from the neuroscience field to explain how inherent anatomical differences between males and females would predispose one sex to the AD phenotype, given the same genetic background and relative proportion of genetic variants.

The DNM model would actually predict an uneven sex ratio between males and females. When diploidy arose in our cellular ancestor, it likely conferred a selective advantage due to its protective mechanism against *de novo* mutations. The majority of *de novo* mutations that result in the production of a faulty protein will not display a phenotype, as its effects can be masked by the expression of the other allele. This is the case with all of the autosomes, however, in males the hemizygous X chromosome is exposed to the effects of every *de novo* mutation that occur. It would also be predicted that females with only one X-chromosome, such as a patient with Turner's syndrome, would be at an increased risk for AD compared to other women, although we are unaware if such a study has been performed.

Furthermore, *de novo* mutations occurring on the X chromosome in females can be briefly transmitted to produce affected males, unless unfavourable X-inactivation produces a phenotype in females. However, it would be rare for the

transmission to occur through multiple generations, as the penetrant mutation is unlikely to be transmitted from male offspring.

2.6 Presence of autistic features in first degree relatives

Many studies have observed that first degree relatives of individuals diagnosed with AD can often show some of the phenotypic symptoms of autism as well[81, 82]. These studies have been carefully conducted with proper controls and are able to consistently achieve statistical significance.

A polygenic model would explain this in terms of allelic AD variants for every autistic trait circulating in the population. The child diagnosed with AD would be considered unfortunate to have acquired by chance a higher number of these allelic variants, which caused him/her to surpass the threshold in the phenotypic triad (social interactions, language/communication and range of interests and activities). The existence and transmission of harmful allelic variants through multiple generations has been discussed in a previous section.

It has also been suggested that the allelic variants predisposing to the phenotypic triad are inherited separately[83]. However, this poses a perplexing question: how do the allelic variants predisposing to the phenotypic triad tend to co-occur? The existence of numerous tightly linked haplotypes located throughout the genome containing clusters of AD-conferring allelic variants that have managed to evade natural selection is unlikely. There could be an unknown mechanism that leads to this higher than expected co-segregation, or a synergistic

relationship between them. However, mutations in several genes have already been discovered whose global brain expression results in global deficits in all three domains without the contribution from other allelic variants[16, 17, 40, 41], which questions the necessity of investigating this avenue.

Under the DNM model, the examples of clearly identifiable, carefully controlled observation of milder forms of autistic symptoms in first degree relatives could result from X-linked inactivation in females, as well as autosomal dominant inheritance with variable expressivity. X-linked inactivation will produce a diverse array of phenotypes in female patients, dependent on the location of inactivation of the favorable chromosome. It is easy to imagine the implications of this if stochasticity resulted in the inactivation of the unaffected chromosome in only a portion of the brain of female carriers. This would conceivably result in various phenotypes depending on the function of the gene as well as the proportion and location of the expression of the mutant copy.

Autosomal dominant mutations in genes that are involved in embryonic development can result in an incredibly diverse array of phenotypes. Mutations in the Sonic Hedgehog protein (SHH), for instance, result in holoprosencephaly, a condition whose behavioural and cognitive phenotype can be similar to AD. Mutations in this gene can result in dramatically different phenotypes, from cyclopia in one family member to slight midline abnormalities in the next[84]. If the phenotype of mutant SHH-associated holoprosencephaly were defined on narrowed thresholds, there would undoubtedly be pure cases of

holoprosencephaly, with siblings showing the broader spectrum of holoprosencephalic features. The mechanism that leads to a phenotype with variable expressivity at most loci is unknown; it is therefore possible that other genetic variants as well as environmental factors may influence the AD phenotype when such a gene is mutated.

2.7 Lack of family history in majority of patients diagnosed with AD

The frequent occurrence of AD with no family history could be explained on multiple levels, depending on the particular family in question. In some families, the fluctuating diagnostic criteria as well as the difficulty of defining AD in adults that were not clinically assessed in their infancy, presents clear historical difficulties. However, in many families there is no history of mental retardation, or any phenotypes that could be confused with AD.

Under a polygenic model, multiple allelic variants would be assumed to be present within the maternal and paternal lineages that culminate in an offspring with no symptoms ever described in distant relatives, the logistics of which have been dealt with in a previous section.

The DNM model, on the other hand, would predict that the majority of individuals diagnosed as autistic would have no family history because their mutation would be an isolated event. If this is the case, offspring of individuals with AD would have a much higher frequency of AD. However, no data have

been generated or analyzed in this regard since, as previously mentioned, the reproductive rate of autistic individuals is very low[79].

2.9 Which model is more consistent with the data?

The DNM and polygenic models were presented separately as a means to show that the DNM model has the power to explain all of the data and observations surrounding the AD phenotype, whereas the polygenic model has several inconsistencies. However, this does not mean that polygenicity cannot play a role in AD. In fact, a polygenic model and the DNM model are not mutually exclusive. However, when coupled with the DNM model, the proportion of cases that will be explained by polygenic inheritance as well as the number of interacting variants must be far lower than is currently believed.

Although ASD represent a continuum on the phenotypic level, this does not necessarily imply the presence of a continuum at the genetic level. It is possible that numerous monogenic, discrete genetic mutations can produce a large spectrum of inextricable phenotypes. This would make it almost impossible to create phenotypic subsets of patients that correspond to the underlying genetic mutation without the conceit of hindsight provided after the genetic mutation is identified in a particular patient. Evidence has already shown that *de novo* mutations play a role in the onset of AD; the current remaining question is whether a DNM model underlies the majority or all of the remaining cases.

2.10 Environmental component analyzed with both models

Several environmental agents have been studied for a potential association with AD, including the measles, mumps, rubella (MMR) vaccine, obstetric complications[85], intrauterine bleeding, infection, caesarian delivery[86], as well as the exposure to harmful substances such as Thalidomide[87]. To date, there are no environmental agents or exposures that have been consistently replicated and definitively proven to lead to the AD phenotype, however examination of the DNM and polygenic models for AD can reveal the potential role for an environmental source or component.

Under a polygenic model for AD, it could be hypothesized that an environmental component may contribute to the phenotype by interacting with the disease causing alleles. In fact, under a polygenic model one must envision individual cases in which an unknown environmental component is solely responsible for the phenotype given that the concordance rate between MZ twins is less than 100%. It must be mentioned that individual cases where an environmental component is the sole causal factor is not mutually exclusive to the DNM model, however under this model the factors outlined in section 3.1 are more likely to create this imperfect concordance.

The DNM model largely precludes the role of an interaction between common genetic variants with an environmental component to produce the phenotype, however this does not eliminate an environmental component to AD. Environmental mutagens may raise a regional mutation rate, therefore predicting

that in regions such as Chernobyl, Nagasaki, and Hiroshima where an elevated level of environmental mutagens are present, a slightly increased incidence of AD would be discovered, although it is unlikely that such specific localized epidemiological information has been tabulated. This also has an implication with respect to the recent increased prevalence observed with AD[13]. Future studies may definitively determine whether an increased incidence of AD is at least slightly responsible for an increased global prevalence of AD, and not entirely due to fluctuating diagnostic criteria. The DNM model would imply that this increase would be due to an environmental agent that influences the mutation rate, although this contribution is likely to be slight. Examining other monogenic disorders that result largely from *de novo* mutations is unlikely to produce significant results as the genomic size upon which the specific mutation may arise is very narrow. However, with AD, where hundreds of genes may produce a vast spectrum of similar phenotypes, an increased mutation rate would likely influence the overall incidence of the disease. Also, in the case of an autosomal dominant mutation that produces variable expressivity, environmental components may influence the phenotype and expression of the mutant allele.

3.0 Strategies and tools for gene identification

As previously mentioned, the importance of determining the genetic architecture is extremely important as different strategies will be used in different scenarios. Under the polygenic model, association studies will be the method of

choice, with linkage studies ideally restricted to larger families. Under the genetic architecture of the DNM model, different strategies must be employed to identify mutations in genes that produce the AD phenotype, as linkage and association studies are based on linkage disequilibrium, which cannot be fruitful unless there is transmission of these genetic variants.

To date, most research attempts to discover the underlying genetic component of AD have been based on a polygenic model using linkage analysis and association studies.

3.1 Linkage analysis

A general description of a classic linkage study is performed by examining the segregation of genome wide markers, either polymorphic microsatellite loci or single nucleotide polymorphisms (SNPs), through a large multiplex pedigree of affected and non-affected individuals. If these markers are examined in tandem, they allow the researcher to construct haplotypes whose transmission can be easily observed through the pedigree, and whose size becomes smaller with each transmission resulting from recombination events during meiosis. The selective transmission of some haplotypes to affected individuals vs. non-affected individuals allows the researcher to employ a statistical analysis to determine the log of odds (LOD) that the disease allele, or mutation that causes the disease, is located within a given chromosomal loci on a specific haplotype. The LOD score will vary based on several conditions including the penetrance of the disease and

the type of analysis performed, and most importantly, the number of individuals in the pedigree.

Large multiplex pedigrees are exceedingly rare for AD which would either result from the low penetrance and requirement of multiple disease alleles in the case of the polygenic model, or the low reproductive rates of individuals with highly penetrant *de novo* mutations in the case of the DNM model. Only one pedigree large enough to refine a haplotype interval to the point of mutation discovery has been identified worldwide, and was used to reconfirm the role of *NLGN*4 as one of many causal genes for AD[17]. The lack of large pedigrees has led to a linkage strategy that involves combining sib-pairs from multiple families to examine the co-occurrence of haplotype sharing between affected siblings, regardless of whether the shared haplotype is common between different families (as mutations in the same gene on different haplotypes could lead to the AD phenotype). This strategy has been historically effective at identifying mutations in genes that lead to single gene disorders in which the phenotype is highly selected against, however in the case of AD this strategy has led to several linkage peaks scattered throughout the genome, none of which are common to every study. The distinction must be made at this point to clarify the difference between a single gene disorder of which the phenotype is *clinically homogenous*, and the DNM model for AD which states that a monogenic cause in multiple genes will account for the large spectrum observed in the *clinically heterogeneous* phenotype. The combination of sib pairs from multiple families has led to linkage

reports at several loci scattered throughout the genome. There are certain loci that have been replicated in more than one study located at 7q21.2-q36.2, 16p12.1-p13.3, 6q14.3-q23.2, 2q24.1-q33.1, 17q11.1-q21.2, 1q21-q44 and 3q21.3-q29[72].

3.2 Family based association studies

Association analysis is a powerful tool that can be used to identify common alleles in the population that may confer a risk for developing a particular disease. All association studies are based on the premise that a particular common marker in the gene pool (ideally a tandem segment of a particular arrangement of SNPs termed a haplotype) will be associated with the disease or in linkage disequilibrium with another variant that is associated with the disease. The most common type of association analysis used in AD research is the transmission disequilibrium test. This method examines the transmission of SNPs located within candidate regions from heterozygous unaffected parents to affected offspring. Transmission of a particular heterozygous SNP, perhaps a C or T would be random, each allele being transmitted with a probability of 50%. However if only affected individuals are examined, an excess transmission of a C or T may be observed if a particular variant either contributes to the disease or is in linkage disequilibrium with a different contributing variant. This experimental strategy is beneficial as the effects of population stratification and requirement of population matched controls are eliminated, which can create false positives with traditional association analyses. This strategy has been utilized in approximately

70 studies of AD cohorts as of August 2007 without leading to the identification of a single variant that is definitively associated with autism.

3.3 High resolution analysis of *de novo* copy number variation

During the process of meiotic recombination, genetic information is exchanged between the maternal and paternal chromosomes. This process is beneficial from the evolutionary perspective as it breaks down linkage disequilibrium allowing beneficial variants to escape negative selection of detrimental variants. However, it can often lead to cytogenetic anomalies such as duplications, deletions, and translocations. Depending on the genomic location and size that may affect several genes in tandem or in some cases may only involve a single gene. In many instances, the deletion or disruption of a gene may lead to the same phenotype as a point mutation that truncates the protein sequence. Therefore, techniques that provide high resolution of *de novo* copy number variation and chromosomal anomalies can be used to identify a region containing a gene of interest. The coding regions of this gene can then be sequenced in a cohort of autistic patients for the presence of point mutations as a means of validation. This strategy has been successfully employed to identify mutations in *SHANK3*[41].

3.4 Medical re-sequencing

If the DNM model accounts for the majority of patients with AD, large scale multi generational candidate gene sequencing projects would be the most efficient approach to identify mutations that produce the AD phenotype. Although financial restrictions may currently be prohibitive, it is now possible to identify a subset of candidate genes that may confer susceptibility to AD. For example, direct sequencing of candidate genes based on interactome maps, or genes belonging to a specific pathway or cellular structure, such as the synapse are currently plausible, whereas whole genome re-sequencing of 192 patients is currently out of the scope of all research labs. The success of this strategy relies heavily on the number of genes being screened as well as the number of subjects used in the study, which is unfortunately a direct reflection of the amount of financial assistance available. As sequencing costs decrease, the identification of *de novo* mutations using sequencing technologies coupled with efficient functional validation will ultimately be the future of gene identification strategies for idiopathic genetic disorders under strong negative selection.

3.5 Bioinformatics

The sequencing, mapping and subsequent online availability of the human genome project in 2001 has led to greater simplicity in identifying genetic variants and mutations that lead to disease[88, 89]. This project has led to a

considerable amount of information that is currently managed through several user-friendly web-based programs such as the UCSC Genome Browser (www.genome.ucsc.edu). These web based tools allow researchers to visually examine genomic regions for candidate regions that would likely harbor a mutation that leads to a given phenotype. For example, in the case of AD, if linkage analysis were to identify a 20 megabase segment on chromosome 4 that segregates with the disease in a given family, this program would allow detailed examination of genes located within this interval, intron/exon structure, expression information, promoter sequences, gene prediction tracks and evolutionary conservation among many other tools that would allow the researcher to make an informed decision regarding which regions are the best candidates for sequencing analysis. These tools also provide the user with a catalogue of known SNPs and microsatellites that aids the selection of suitable markers for association studies.

3.6 Validation of genetic variants

Although any two humans on the planet are 99.99% identical, this means that there are still approximately 12,000,000 variable genetic sites. This poses a difficulty in determining whether a SNP is responsible for producing a given phenotype. Therefore a validation process must be followed in order to determine the role of an identified SNP in causing the disease. The validation process is generally two fold, genetic and functional.

Genetic and statistical arguments should be satisfied before beginning functional work on a given genetic variant. For instance, many factors should be assessed depending on the type of study performed. As an example, in a large family linkage study, it should be determined first if the variant is segregating on the disease haplotype, and whether the variant is commonly found in controls (and if the proper control samples have been used). In the case of an association study, the sample size should be increased and replicated in different populations. For new mutations, care should be taken to determine whether the proper parents were obtained, and whether this site is highly conserved among other populations.

After genetic and statistical arguments have been satisfied, one would normally proceed with functional validation. This will vary widely depending on many factors including genomic location (promoter, enhancer, coding region, splice donor/acceptor site). For example, if a rare mutation has been discovered in a promoter element, a luciferase assay would be performed to determine the effect this variant has on expression. If a mutation were discovered in a splice donor site, RT-PCR would be performed on cell lines taken from patients with the variant to determine whether the variant affects splicing. Although one may show that a variant displays a functional effect on the cellular level, it does not necessarily mean that it will on a phenotypic level, therefore animal models are often developed for the variant such as transgenic or knockout mice.

3.7 The French Canadian population as a sample cohort for AD

The Rouleau lab has currently assembled approximately 600 DNA samples taken from patients with AD, of which ~ 70% are derived from the French Canadian (FC) population. These samples have been taken from patients that are have diagnosed according to DSM-IV criteria, administration of the ADI-R and/or ADOS-G and have participated willingly based on informed consent. The FC population is estimated to derive from a small founder population of 3000-4000 individuals which arrived from southern France between the years 1750-1850. Using a sample cohort derived from this population presents several considerations for gene discovery projects. This founder effect can provide benefits when one considers either the polygenic model or the DNM model. Under a polygenic model, the genetic variants that predispose to AD within the FC population would be homogenous and likely increase the probability that the genetic basis of AD in the FC population is based on a homogenous group of genetic variants. Under the DNM model, the majority of cases would not be influenced by the founder effect as *de novo* mutations are equally likely to strike a genetically homogenous population as they are a genetically diverse population, however transmission may occur via the mechanisms listed in Table I (page 19) and be more easily identified in a genetically homogenous population through haplotype sharing across multiple families.

Unfortunately, the prematurity of the FC population also likely means that fewer recombination events separate the individuals of the population from the

founders. Therefore, positional cloning techniques are likely to identify large haplotype blocks with many candidate genes delaying the mutation discovery.

4.0 Association study of the $GABA_A$ receptor subunit cluster at 15q12 in the French Canadian population

4.1 Abstract

Chromosomal region 15q11-13 has been consistently implicated in autism through numerous avenues such as maternal duplications, breakpoints in the region and presence of autistic features in patients with Prader-Willi and Angelman syndrome[90]. Furthermore, linkage to *GABRA3* within a cluster of $GABA_A$ receptor subunit genes has been previously described and replicated in different AD populations[91-93]. In this study, 185 trios were assembled with affected patients diagnosed according to DSM-IV criteria, administration of the ADI-R and/or ADOS-G. Genomic DNA was extracted from lymphocytes of patients, of which 70-75% are of French Canadian origin. With an average spacing of 30 Kb, 26 SNPs spanning 820Kb of the $GABA_A$ receptor subunit cluster at 15q12 were selected from the dbSNP and Celera databases based on minor allele frequency >0.05, spacing and haplotype structure. A transmission disequilibrium test (TDT) from the UNPHASED software suite was used to analyze the markers in question, 3 SNPs revealed a significant association including rs2873027.

4.2 Introduction and rationale

Gamma (λ) aminobutyric acid (GABA) is the primary inhibitory neurotransmitter in the adult brain, but likely acts an excitatory neurotransmitter

during embryonic development[94]. GABA plays an important role in signaling before, during and after synaptogenesis and synaptic maturation[95]. It is also a key signaling factor that controls the early and late phases of neurogenesis including cell proliferation, migration and cell fate commitment followed by the late stages of maturation and synaptic integration[95]. GABA interacts on the post synaptic membrane with a specific receptor that varies with cell type. GABA sub A ($GABA_A$) receptors are pentameric chloride channels composed of various subunits labeled from α1- α6, β1-β3, λ1- λ3, ρ1-ρ3, as well as δ,ε,θ,π [96, 97]. These subunits are found in several locations throughout the genome, however they combine in multiple permutations to produce very high functional heterogeneity[98].

Examining these receptor subunits for a potential role in AD is warranted due to the role they play in synaptogenesis and synaptic maturation. Furthermore, suppression of GABAergic inhibition has been hypothesized to lead to the AD phenotype[99] based on reduced binding of radiolabeled $GABA_A$ receptor ligands in autopsy brain specimens from individuals with autism[100], as well as elevated levels of circulating GABA in children with autism[101-103].

The observation of autistic features in individuals with Prader-Willi and Angelman syndrome, as well as the observation of chromosomal re-arrangements in autistic patients at the Prader-Willi/Angelman region at 15q11-13[104-106] initially led to the investigation of the $GABA_A$ receptor subunit cluster located at 15q12, a 1.1 Mb region containing three $GABA_A$ receptor subunit genes *GABRB3*, *GABRA5*, and *GABRG3* (Figure 1, page 48). Under the pretense of a

polygenic model, several studies have been performed on this region to examine association between common genetic variants and AD with mixed positive and negative results[91, 92, 107-115].

The gene that has shown more consistent reports of association is *GABRB3*. There have been reports of association with microsatellite marker 155CA-2 located near *GABRB3*[91, 114]. Recent studies have revealed five significant local p-values using SNPs across the gene[92], as well as an association with a haplotype located within an intron of *GABRB3*[113]. There are several imprinted genes in the 15q11-13 region, therefore a potential exists for imprinting at the $GABA_A$ receptor subunit cluster. Reports thus far have been conflicting with reports of bi-allelic expression in mice[116], as well as a paternal expression bias in humans[117, 118]. A more recent report has shown that these genes are bi-allelically expressed in humans, however maybe subject to epigenetic dysregulation in patients with AD and Rett Syndrome[119].

4.3 Materials and Methods

Clinical population

Three hundred and three trios (father, mother, affected) were tested in this study, of which 70-75% are of French-Canadian origin. Each affected patient was diagnosed with AD according to DSM-IV criteria using either the ADI-R and/or ADOS-G by at least one psychiatrist. Exclusion criteria from this study include a previous diagnosis or evidence of any psychiatric/neurological condition such as

Rett Syndrome, fragile-X syndrome, tuberous sclerosis, Tourette syndrome, or phenylketonuria.

Blood samples were collected from each participating member (mother, father, affected offspring) followed by DNA extraction from lymphocytes. The study received ethics approval from the local ethics committee, and the participants or parents provided informed consent.

SNP selection

Twenty-six SNPs with an average spacing of 30 Kb were selected spanning 820 Kb of the $GABA_A$ receptor subunit cluster at 15q12. These SNPs were selected based on minor allele frequency >0.05, potential functional candidacy, as well as distance from other selected SNPs to ensure adequate coverage and haplotype construction of the region. A distributional map of these SNPs is shown in Figure 1 (page 48).

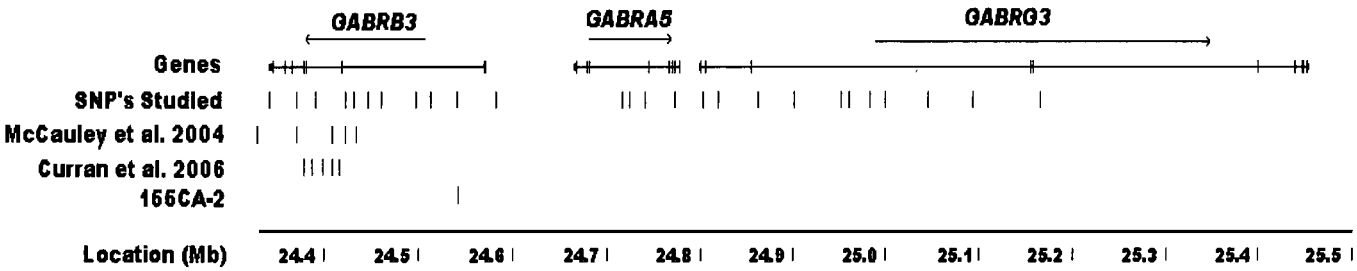

Figure 1: Map of the $GABA_A$ receptor subunit cluster at 15q12. The genes are displayed showing their approximate location represented as distance in Mb from the terminal segment of the small arm of chromosome 15. The structure of each gene is revealed with vertical strokes representing exons, and the 5' and 3' UTR represented as vertical strokes of intermediate length with arrows above revealing transcriptional orientation. The distribution of SNPs used in this study are displayed below the gene as vertical strokes that span 820 Kb of the total 1.1Mb represented in the diagram. The SNP rs2873027 whose association with AD has been replicated in this study from McCauley et al is shown in red. SNPs with positive local P-values within the GABRB3 gene are shown in red from McCauley et al, 2004. The 5 SNP haplotype whose excess transmission was observed in a patient cohort from Curran et al, 2006 is represented in green. The microsatellite 155CA-2 whose association with AD has been observed previously is also shown in a separate column.

Genotyping

One hundred and eighty-five trios were initially genotyped, however an additional 124 trios were genotyped at rs2873027. Twenty-six SNPs were selected for use in this study (see Table II, page 51). Genotyping was performed with a Taqman based allelic discrimination assay (Applied Biosystems, Foster City, CA). The polymerase chain reaction (PCR) was carried out in a total reaction volume of 5 µl using the following amplification protocol: denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 92°C for 15 s, and annealing and extension at 60°C for 60 s. Post-PCR, the genotype of each sample was assigned automatically by measuring the allele-specific fluorescence in the ABI 7900 HT Sequence Detection Systems, using the SDS 2.2.2 software for allelic discrimination (Applied Biosystems).

Statistical Analysis

Statistical analyses were performed using the program TDTPHASE from the UNPHASED software suite, a program written by Frank Dudbridge[120]. UNPHASED is a program that performs genetic association analysis. It implements maximum-likelihood inference on haplotype and genotype effects while dealing with missing data such as uncertain phase and missing genotypes.

The transmission/disequilibrium test (TDT) was developed in 1996 as a test for linkage disequilibrium[121]. The test is based on the principle that the probability of transmission of two different marker alleles from unaffected parents

to affected offspring will be significantly higher when the marker locus and the hypothetical disease locus are linked and are in linkage disequilibrium. When SNPs are used as the marker allele, it is therefore important to use SNPs with relatively similar population frequency to boost the number of heterozygous parents. This test is effective at identifying loci with less penetrant effects on the phenotype.

4.4 Results

Examination of local p-values from 26 SNPs across an 820 Kb region revealed three sites with significant transmission disequilibrium, $p<0.05$: rs2873027, rs6606905, and rs8043244 (Table II, page 51). Furthermore, three SNPs with relatively low p-values were found: rs75415 (0.08414), rs140682 (0.065482) and hCV42974 (0.08873) (Table II, page 51). Interestingly, an excess transmission of the C allele from SNP rs2873027 was observed in the initial screening of 179 families. Considering the previous results from McCauley et. al[92], as well as the excess transmission of a haplotype in linkage disequilibrium with this variant from Curran et. al[113], an additional 124 trios were screened for transmission disequilibrium at rs2873027 and added to the initial cohort of 179, for a total of 303 trios genotyped at rs2873027 (Table III, page 52). The p-value at this site increased from a significant 0.01246 to a non-significant 0.2894.

Table II: Results obtained from the TDT. The two columns on the left are the SNP references from the dbSNP and Celera databases. Empty fields represent SNPs found solely within one database. The distribution of these SNPs from top to bottom are shown from left to right on Figure 1. Transmissions (T) and non transmissions (NT) are listed along with the identity of the corresponding SNP for the significant as well as lower p-values. The furthest column on the right is the p-value obtained in previously published studies (PPS). * denotes significance at the 95% level.

Marker	dbSNP	Celera	Gene	SNP	p-value	T	NT	Allele	PPS
1	rs11637141	hCV2911915	GABRB3	c/t	0.851258				
2	rs1426217	hCV2901088	GABRB3	a/g	0.3887	124	114	g	0.21
3	rs1426208	hCV2901113	GABRB3	c/t	0.2529				
4	rs2873027	hCV2901140	GABRB3	c/t	0.01246*	179	149	c	0.04*
5	rs754185	hCV2901163	GABRB3	c/t	0.08414	193	173	c	0.82
6	rs12912421	hCV2901117	GABRB3	a/g	0.7423				0.94
7	rs890317	hCV8865209	GABRB3	a/c	0.6656				
8	rs11631421	hCV245488	GABRB3	c/t	0.4336				0.21
9	rs919075	hCV8864788	GABRB3	c/t	0.1732				
10	rs11161335	hCV9399204	GABRB3	a/t	0.1915				
11	rs8179184	hCV276328	inter	c/t	0.3186				
12	rs8032064	hCV416499	GABRA5	c/t	0.3715				
13		hCV42974	GABRA5	c/t	0.08873	114	94	t	0.08
14	rs140682	hCV1028938	GABRA5	c/t	0.06564	196	173	c	0.65
15	rs140685	hCV1028939	GABRA5	c/t	0.8066				0.84
16	rs6606905	hCV2078439	GABRG3	c/t	0.04957*	196	184	t	
17		hCV2078482	GABRG3	c/t	0.2675				0.21
18	rs2376483	hCV2078528	GABRG3	a/g	0.8048				
19	rs1432126	hCV8865801	GABRG3	c/t	0.634				
20	rs1029937	hCV2665757	GABRG3	a/g	0.4182				0.94
21	rs6606877	hCV2665737	GABRG3	c/t	0.2162				0.6
22	rs2286946	hCV2665715	GABRG3	a/g	0.905				0.66
23	rs208126	hCV2665687	GABRG3	g/t	0.2059				0.16
24	rs897173	hCV9408511	GABRG3	a/g	0.8403				0.06
25	rs8043244	hCV9408423	GABRG3	a/g	0.03936*	112	88	g	0.25
26		hCV435176	GABRG3	a/g	0.7394				1

Table III: TDT results of an additional 124 trios were genotyped at SNP rs2873027 bringing the total to 303. The p-value increased from 0.01246 to 0.2894

Marker	dbSNP rs#	Celera hCV #	Gene	SNP	p-value	T	NT	Allele
4	rs2873027	hCV2901140	GABRB3	c/t	0.2894	297	280	C

4.5 Discussion

In the initial screening, three SNPs showed significant transmission disequilibrium ($p<0.05$), rs2873027, rs6606905 and rs8043244, and three SNPs revealed lower p-values: rs754185 (0.08414) hCV42974 (0.08873) rs140682 (0.06564). All of these values must be examined with caution as they have not been corrected for multiple testing.

Our initial screen replicated previous work by McCauley et. al[92], showing a significant transmission disequilibrium for SNP rs2873027. It is also noteworthy that work by Curran et. al[113] showed significant transmission of a haplotype located 10 Kb upstream in linkage disequilibrium, however they did not include that particular SNP in their study. The inclusion of 124 additional trios totaling 303 trios at this site increased our p-value to 0.2894, questioning the replication and presents several scenarios which will need to be examined based on the polygenic and DNM model. This SNP is located within an unconserved region of intron 3 approximately 700 bp from the intron/exon boundary of exon 3.

Under the polygenic model, our results imply that, although this variant or a variant in linkage disequilibrium may contribute to the autism phenotype, it is clearly not required, and may only contribute to a small proportion of overall cases. Given that this gene encodes a GABA receptor subunit that interacts with several other subunits to create a functional receptor, a potential exists for rs2873027 to be in linkage disequilibrium with a common coding variant from exon 3 that interacts with another coding variant from a partnering subunit to

produce the phenotype. However, this potential is slight given that exon 3 is perfectly conserved among mammals, and no SNPs have been documented within it in humans. The potential for variants to co-exist through hundreds of ancestral generations that produce phenotypes under strong negative selection has been addressed in section 2.4. However, sequencing of this exon was not performed on individual cases in this study

Interpreting these results under the DNM model requires a slightly more abstract approach. Assuming that the association to rs2873027 as well as the haplotype in linkage disequilibrium with this SNP from Curran et. al[113], is functionally relevant and not a result of multiple testing, this association would stem from a recessive locus or from an association with a haplotype that has a tendency for non allelic homologous recombination (NAHR), which results in functional *de novo* deletions within this region. This could easily be imagined if the associated haplotype contained several extra repeats that would initiate homologous recombination with a downstream repeat resulting in a duplication or deletion. This is not an unprecedented mechanism, as several examples currently exist that lead to human genetic disorders[122-127]. Given the frequency of abnormal karyotypes in AD patients carrying duplications or deletions in this region, it is plausible that smaller scale duplications/deletions would produce the AD phenotype as well. Future techniques providing higher resolution of DNA copy number variants (CNVs) within this region in multiple patients containing the risk haplotype, may decipher whether this mechanism underlies the association to

rs2873027. However, the risk haplotype must first be properly characterized and identified through denser marker spacing in order to limit the cost of such a procedure, as simply screening multiple patients that contain a C at SNP rs2873027 would be quite inefficient.

The t allele from rs6606905 was transmitted in excess with slight significance (p=0.04957). This SNP lies within intron 1 of *GABRG3* approximately 1,000 bps from exon 1. We are unaware of the inclusion of this SNP in previous studies, however there have been no previously reported associations with any markers that would be in strong linkage disequilibrium. Replication from future studies will be required to pursue the nature of the association with this SNP any further.

Slightly significant excess transmission (p=0.03936) was also observed for the g allele from SNP rs8043244. This SNP is located within an unconserved region of intron 3 from the gene *GABRG3*. This SNP has been included in previous studies with a non significant p-value of 0.25[92]. Replication is required from future studies to determine if this association is functionally relevant or a chance result of multiple testing.

4.6 Conclusion

This study was performed to identify variants within the $GABA_A$ receptor subunit cluster, which may contribute or predispose to the AD phenotype. In addition to identifying an association with two SNPs that have not been reported

in the literature, some additional support has been given to the association of rs2873027 with AD. These results must be replicated in different populations for further validation.

5.0 Investigating the role of polyalanine expansions/contractions in AD

5.1 Abstract

Polyalanine repeats are under increasing scrutiny as they have been discovered to be involved in an increasing number of genetic disorders. In this study, 56 alanine repeats were screened for variants in 192 patients diagnosed with AD according to DSM-IV specifications and 192 controls, in an attempt to determine a link for alanine repeats with AD. This led to the identification of several interesting genetic variants from several genes including *ZIC2*, *TBLXR1, ARID1A, POUF3 and FOXD3*.

5.2 Introduction and rationale

Repetitive stretches of DNA are found throughout the human genome and are prone to intergenerational expansions and contractions due to their susceptibility to slipped strands during replication as well as unequal homologous allelic recombination. These expansions and contractions are generally neutral in the case of microsatellite DNA, however they can have phenotypic consequences if they produce a functional effect. These phenotypic consequences can lead to human genetic disorders, of which a growing class termed trinucleotide repeat disorders (TRD) is already well characterized. TRDs are the result of the expansion of nucleotide triplets and can be classified into three categories based

on the distinct disease mechanism[128]. In the first category, repeat expansions in non-coding sequences generally result in decreased transcription and mRNA formation. For example, the expansion of the (CGG)n repeat in the 5' UTR promoter region of *FMR1* results in hypermethylation and transcriptional repression which leads to Fragile-X Syndrome[129]. Secondly, expansions in coding but non-translated sequences can cause complex RNA pathology[130] as in the case of myotonic dystrophy[131]. The third category is the expansion of *translated* sequences coding for poly-amino acid tracts that lead to toxic mis-folded proteins, which is the most common disease mechanism leading to several human diseases. For example, the expansion of CAG repeats in the coding portion of the *Huntingtin* gene lead to the formation of a protein with a large toxic stretch of the amino acid glutamine. Expansion of polyalanine tracts have also been linked to human disease. To date, nine genes with polyalanine tracts have been identified for which an identifiable human disease has been found (Table IV, page 61). Although the function of these polyalanine stretches has not been determined, they are over represented in transcription factors, particularly those expressed during development. This fact presents itself when one looks at the list of genes involved in human disease from Table IV (page 61), as every gene aside from *PABPN1* is a transcription factor.

When one compares the location of the polyalanine tract within the gene, as well as the length of the repeat among mammals, considerable conservation is generally observed, which implies that there has been strong negative selection

against *de novo* mutations that produced any considerable functional or structural changes at the protein level. Alanine is encoded at the genetic level by the codon GCN, where N represents any of the four nucleotides. Therefore, the third base pair can also mutate freely and drift to fixation without producing a functional consequence on the protein level. As a consequence, the majority of alanine stretches are impure heteropolymeric repeats (e.g. GC(G)GC(C)GC(C)-GC(A)GC(T)), whereas the repeats found in other TRDs are pure homopolymeric repeats (e.g. CAGCAGCAGCAG) This explains why the expansion and contraction of polyalanine tracts are a result of unequal allelic homologous recombination during meiosis[132], whereas the expansion observed in other TRDs are likely a result of polymerase slippage during replication. This differing mutational mechanism leads to far less extreme shifts in repeat length for polyalanine tracts versus other trinucleotide repeats. For example, in the *Huntingtin* gene, the CAG repeat length can vary in the normal range from 6-35 and from 40-121 in the disease range[133], and in the case of *FMR1* the number of CGG repeats can vary enormously with the pathological range greater than 200 repeats[134]. In polyalanine disorders, the upper range in repeat length is always below 30 and the difference between wild type and disease length is far more subtle with ranges between 1-14[128]. As an example, in the case of ocular pharyngeal muscular dystrophy (OPMD), the wild type length of the polyalanine tract in the *PABPN1* gene is 10 with the mutant allele ranging from 11-17 repeats[135].

The pathological mechanism for the majority of polyalanine disorders is currently believed to be caused by protein misfolding[128]. The general concept is that mis-folded protein will form cytoplasmic aggregates that exceed their degradation and can interfere with the wild type protein, possibly explaining the dominant nature of many of these mutations. Mutant *ARX* (mutations in which have been discovered in patients with AD) also form aggregates, but these appear to occur predominantly in the nucleus of affected cells rather than the cytoplasm[136].

This type of pathology has generally led to adult onset disorders, however given the frequent appearance of these polyalanine repeats in embryonic transcription factors it had been theorized and recently shown that these alanine repeats are involved DNA binding as well. Deletions in the C-terminal alanine repeat of *ZIC2* have been shown to decrease its DNA binding efficiency[137].

Genes with polyalanine repeats are leading candidates for the investigation of a role in AD for several reasons. The majority of these tracts are found in embryonic transcriptions factors that are involved in neurological and physiological development, representing candidate genes for syndromic and non-syndromic forms of AD. Polyalanine repeats are also highly conserved in mammals in terms of length and site, as expansions of a single repeat have been shown to cause disease. Therefore slight variations in these tracts may produce profound phenotypic consequences. Considering that the majority of these genes are transcription factors, it is possible that slight variations in alanine length affect

protein folding and create functional differences that influence expression levels, which may trigger cascading events during embryological development.

The mutation rate of tandem DNA repeats has been estimated to be greater than 100,000 times that of a point mutation[55]. Although this represents the mutation rate of pure tandem repeats, the mutation rate of polyalanine stretches is undoubtedly greatly increased as well, and is a function of the repeat length, as longer repeats are generally less stable and more prone to expansion/contraction. A higher mutation rate coupled with the degree of conservation in these repeats implies that although *de novo* mutations are common, there is a considerable degree of selection against the resulting phenotype. This fits perfectly within the framework of the DNM model, however it is still possible that under a polygenic model, common variations of alanine repeat length can lead to, or predispose to the AD phenotype.

Table IV: List of genes whose alanine tracts have been implicated in human disease.

Gene	Disease
FOXL2	Blepharophimosis/Ptosis/ Epicanthus inversus[138]
PABPN1	OPMD[139]
ZIC2	Holoprosencephaly[140]
ARX	Mental Retardation, AD[42]
HOXA13	Hand-foot-genital syndrome[141]
HOXD13	Synpolydactyly[142]
RUNX2	Cleidocranial dysplasia[143]
SOX3	Mental Retardation[144]
PHOX2B	Congenital hypoventilation syndrome[145]

Given the clear potential for polyalanine expansion and contractions to be implicated in the phenotype of AD, this study was designed to perform an initial

broad screen of a fraction of alanine repeats in patients to compare with the normal population. If this study provides interesting results, it will lead to future screening of additional variants and functional studies. Furthermore, alanine expansions and contractions represent an important class of coding genetic variation within humans that has not yet been catalogued. Therefore, another outcome of this study will be to catalogue much of this variation.

5.3 Materials and Methods

Gene list

Online bioinformatics programs available at the NCBI database (http://www.ncbi.nlm.nih.gov/BLAST/) were used to develop a list of genes within the scope of this project using the Blast-P algorithm. The Blast-P search tool examines the coding portions of proteins for matches against the users search query. Tandem alanine repeats can vary in length from 2-20 with the higher repeat lengths decreasing in frequency, therefore criteria were established to determine a gene's candidacy for this project. The gene list was balanced on two main criteria, alanine repeat length and functional candidacy of the gene. Longer repeat lengths are less stable therefore an alanine repeats greater than ten were preferentially selected for this project, however several genes were selected with repeat lengths lower than this threshold based on functional candidacy with the shortest repeat length at five tandem alanine residues. Several of the genes contained more than

one alanine tract that were examined separately. In all 56 alanine tracts were successfully screened in this study.

Clinical population and controls

One hundred and ninety-two patients as well as 192 controls were selected for initial screening of these 56 selected alanine repeats. An additional 161 patients and approximately 460 additional controls were screened for variants that revealed rare polymorphisms. Each affected patient was diagnosed with AD according to DSM-IV criteria using either the ADI-R and/or ADOS-G by at least one psychiatrist. Exclusion criteria from this study include a previous diagnosis or evidence of any psychiatric/neurological condition such as Rett Syndrome, fragile-X syndrome, tuberous sclerosis, Tourette's syndrome, or phenylketonuria. Blood samples were collected from each participating patient followed by DNA extraction from lymphocytes. The study received ethics approval from the local ethics committee, and the participants or parents provided informed consent.

Determining alanine repeat lengths

Repeat lengths were determined by radiolabeled (alpha-35S)-dATP PCR amplification of genomic DNA using primers that surround the alanine repeats. The PCR reactions were performed with 75–150 ng of genomic DNA; the primer sequences and PCR reaction conditions are available upon request. These PCR products were run on 6% denaturing polyacrylamide gels and compared with the

product of a di-deoxy sequencing reaction to determine the length of the amplicon. This provides a rough but accurate assessment of the repeat length of the alanine residues, however all rare variants were directly sequenced to produce a more accurate representation of the repeat length, as well as to certify that the variations of amplicon length were a direct result of variations in the alanine repeat length. Furthermore a sample was sequenced from each of the other tracts as well to ensure that the correct fragment had been amplified.

5.4 Results

In all, 56 alanine repeats were successfully screened. The results of this screening created a natural categorization of these alanine tracts based on degree of variation and potential for disease allele based on discrimination between cases and controls: 1) alanine tracts which revealed no variation (Table V, page 65), 2) alanine tracts which were highly polymorphic with no identifiable wild type, (Table VI, page 65) 3) alanine tracts which had a clearly identifiable wild type repeat length with slight variations that were similar in cases and in controls (Table VII, page 66), 4) rare variants present in AD patients with little to no presence in controls (Table VIII, page 66). Several of the genes contained multiple alanine tracts therefore this is specified by noting if the repeat was located at the N-terminal, middle, or C-terminal end of the protein.

Table V: Gene list for which no variations in alanine length were determined. Several genes contained more than one repeat therefore this is specified in some cases by noting the location within the protein.

Gene	Locus	Repeat length
PBX3	9q33-q34	9
SOX3	Xq27.1 C-terminal	10
POU4F1	13q31.1	10
EN1	2q14.2	10
DACH	13q22.1	9
RBM9	22q13.1	8
SOX21	13q32.1 N-terminal	10
ZIC2	13q32 central tract	
ZIC2	13q32 N-terminal	9
HOXD13	2q31.1	6
FLJ23342	11q24.2	8
CRSP2	Xp11.4	8
MECP2	Xq28	5
GSH-2	4q11-q12	16
TBX1	22q11.21	15
LHX2	9q33-q34.1	10
OLIG2	21q22.11	13
SWAP2	19q13.32	10
HSRNAFEV	2q35	12
CHD3	17p13.1	10
ZBTB8	1p35.1	15
DMRTA1	9p21.3	11
NLK	17q11.2	13
ZIC5	13q32.3 N-terminal	9

Table VI: Gene list for which no discernible wild type repeat length was observed, therefore the repeat numbers given are approximations. Several genes contained more than one repeat therefore this is specified in some cases by noting the location within the protein

Gene	Locus	Repeat size
SOX21	13q32.1 C-terminal	~12
PRKCM	14q11	~10
ASCL1	12q22-q23	~13
NIPA1	15q11.2	~13

Given the potential for common alanine repeats within these genes to be involved in the AD phenotype as a polygenic contributor, the allele and genotype

frequencies were tabulated within the patient populations were tested for Hardy-Weinberg equilibrium with no significant deviations.

Table VII: Gene list for which several variants were observed whose frequencies matched those found in controls. Several genes contained more than one repeat therefore this is specified in some cases by noting the location within the protein. Several genes contained more than one repeat therefore this is specified in some cases by noting the location within the protein.

Gene	Locus		Repeat size
SOX3	Xq27.1	N-terminal	15
HOXD13	2q31.1		15
GUCY1A2	11q21-q22		8
HOXD8	2q31.1		9
IGSF4	11q23.2		8
ARID1B	6q25.1		11
HOXD13	2q31.1		15
NKX2-4	20p11		16
EGR2	10q21.1		10
EIF4G3	1p36.12		16
ZIC5	13q32.3	C-terminal	9
PTBP2	1p21.3		10
TLE4	9q21.3		11
KIAA0601	1p36.12		12
ASCL1	12q22-q23		13
RIS1	3p21.3		14
ASXL2	2p24.1		13
FBXO41	2p13.2		11
ARID1B	6q25.1		11
NKX2-3	10q24.2		16
ZIC3	Xq26.2		10
SOX21	13q32.1	Central Tract	13

Table VIII: List of alanine tracts whose rare variants were observed in patients and and unmatched in controls. Several genes contained more than one repeat therefore this is specified in some cases by noting the location within the protein.

Gene	Locus		Repeat size
ZIC2	13q32	C-terminal	15
TBL1XR1	3q23		11
POU3F3a	2q12.1	N-terminal	12
POU3F3b	2q12.1	C-terminal	16
FOXD3C	1p31.3	N-terminal	11
ARID1A	1p36.11		11

All variants from Tables VII and VIII (page 66) were verified by direct sequencing of PCR amplicons to ensure that fragment length variation on polyacrylamide gel were a result of alanine expansions/contraction and not an artifact or insertion/deletion of a different region. This is shown in Figure 2 (page 68) for the *ZIC2* C-terminal alanine deletion and Figure 3 (page 69) for the *TBLXR1* (alternative name *IRA1*) alanine expansion. Although all variants were screened in 192 patients and 192 controls, further screening of patients and controls was performed where possible of tracts listed in Table VIII (page 66) in an attempt to identify the rare variants found in larger sample cohorts.

ZIC2 was screened in 342 patients and a single large deletion was observed that was not seen in 384 controls (768 chromosomes). The N-terminal or 5' alanine repeat of *POU3F3* was screened in 342 patients as well as 192 patients (384 chromosomes) revealing a heterozygous deletion of seven alanines (12-5) in a single patient that was not observed in any controls. The C-terminal or 3'alanine tract of *POU3F3* (BRN-1 protein) is larger containing 16 alanines and was screened in 342 patients with AD and 192 controls (384 chromosomes). A heterozygous deletion of two alanines (16-14) was observed in one patient that was not observed in 384 control chromosomes. The C-terminal alanine repeat of *FOXD3* was screened in 384 patients and 192 controls (384 chromosomes) and an expansion of 4 alanines was observed in one patient not seen in controls. *ARID1A* was screened in 192 patients and 192 controls and single deletion of seven alanines was observed in one patient that was not observed in controls.

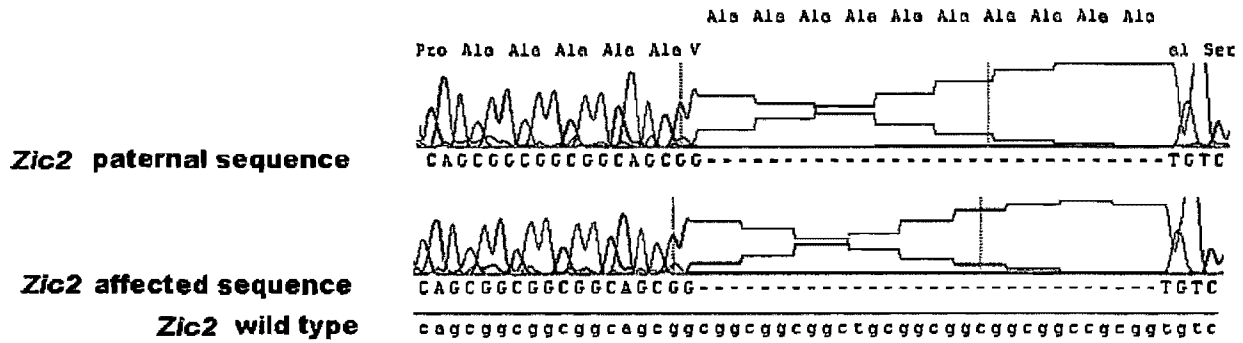

Figure 2: Sequence results showing an in frame 30 bp deletion in the affected offspring and the unaffected father resulting in a 15-5 deletion of 10 alanine residues. The reading frame consists of a six 5 alanines followed by a valine. The wild type protein contains 15 alanines followed by a valine residue (the program placed the first position valine codon (guanine) at the 5' end of the deletion, whereas it could have aligned it at the 3' end with equal merit given the wild type has a G at either end.

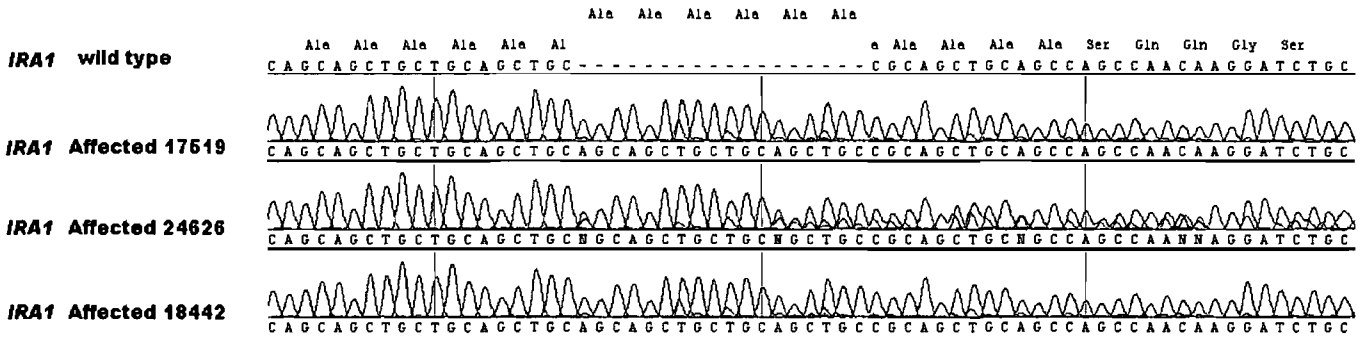

Figure 3: Sequence results showing an in frame insertion of 18 bp's resulting in an expansion from 11-17 alanine residues in three unrelated French Canadian patients. The patients were all heterozygous however the alleles were separated on an agarose gel allowing the two copies to be sequenced individually.

TBLXR1 was screened in 384 patients and over 500 population matched controls (>1,000 chromosomes). A heterozygous expansion of six alanines was observed in three patients of French Canadian origin that was not observed in any controls. Given the similar family background, microsatellite markers were genotyped at eight markers surrounding the expansion to determine whether the mutation was identical by descent (IBD). The resulting haplotype construction (Figure 4, page 70) confirmed that the expansion had a single point of origin that had been transmitted through several generations. None of the families reported a history of AD.

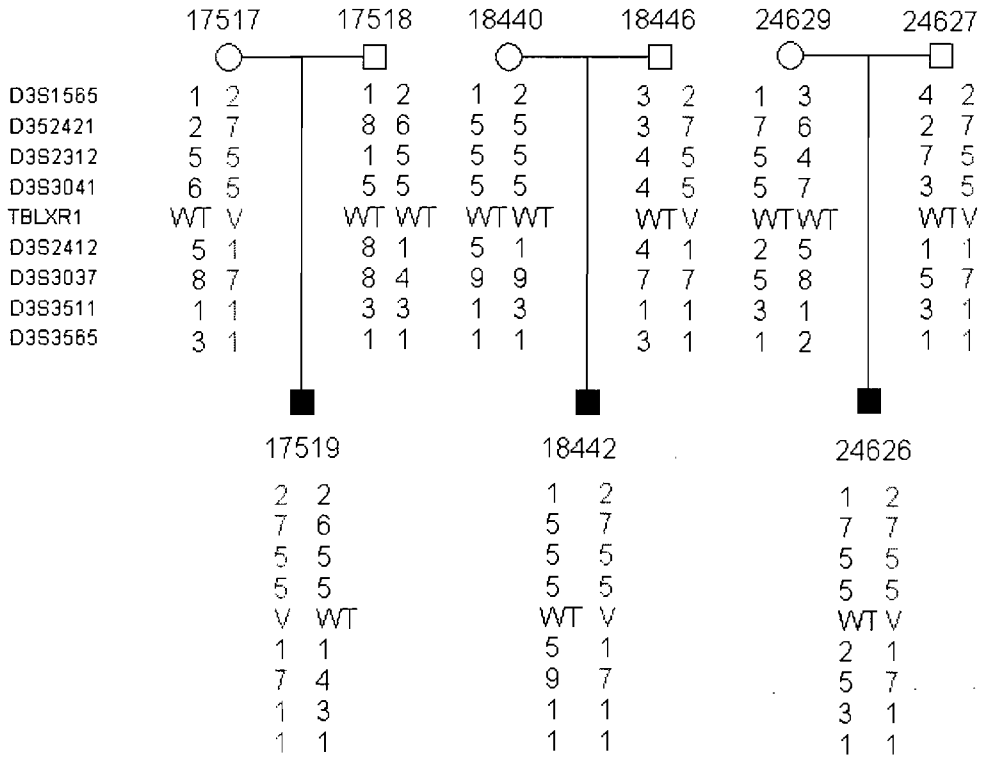

Figure 4: Haplotype construction of region surrounding *TBLXR1* alanine variant in three French Canadian trios carrying a six alanine expansion. The haplotype containing the alanine variant (V) was transmitted from the mother in one family (17518) and the father in the other two families (18446, 24627). The normal allele is labeled as wild type (WT).

5.5 Discussion

The tract list from Table V (page 65) shows that 24 of the 56 alanine tracts that had been screened revealed no variation in alanine repeat length among our 192 patients and 192 controls. There does not seem to be any correlation between alanine repeat length and degree of variability. Although local recombination rates may influence the mutation rate, the high degree of conservation is most likely a direct reflection of extreme structural and functional constraint on these sequences. Even though this group revealed negative results within the scope of this project, it is likely that future studies may discover that *de novo* polyalanine expansions or contractions are the causal role for a genetic disorder which is under strong negative selection or result in embryonic lethality.

Table VI (page 65) displays a list of four alanine tracts that were extremely polymorphic, containing no identifiable wild type length with several observed repeats. The frequencies of these alleles were compared in patients and controls, and the genotypes were examined for Hardy-Weinberg equilibrium within the autistic cohort to examine the possibility that certain combinations of alanine repeat lengths within these genes conferred a risk to developing the AD. This comparison did not produce any significant results. Although this reveals that these genes are provided with a degree of structural flexibility in terms of polyalanine length, it does not necessarily follow that the variations produce no functional effect. Previous work has suggested that alanine repeats may provide a mechanism for rapid morphological evolution, and therefore may represent an

important class of human genetic variation[146]. Future studies could focus on the allele and genotype frequencies of the alanine repeat variants between different populations to identify common variants that may be associated with common phenotypic variation between population groups.

Table VII (page 66) contains a list of alanine tracts with a clearly identifiable wild type length, which revealed several variants that were matched in the control populations. As mentioned above, although these variants could not be linked to the AD phenotype, these are rare coding variants that may have a phenotypic effect not identified within the scope of this study.

Table VIII (page 66) contains a list of genes whose rare variants were not observed in controls. It is worth noting that rare variants were observed in the control samples, which were not also observed in patients. Consequently, placement in this group is not a definitive link to the AD phenotype, it simply means that further screening of controls and patients is warranted. The work that has been performed to date on the genes within this group is summarized below.

A large 30 bp, deletion of 10 alanines (15-5) was discovered in the C-terminal alanine repeat of the *ZIC2* gene in one patient within the initial 192 affected sample cohort. This was not observed in our initial 192 controls. We subsequently screened an additional 150 AD subjects as well as an additional 192 population matched controls without observing this variant again. This resulted in a total of 1/342 patients and 0/384 controls. Given that this mutation would be producing the phenotype in a heterozygous manner, these numbers could be

looked at in terms of number of chromosomes: 1/684 patient chromosomes and 0/768 control chromosomes. Furthermore, this alanine stretch was perfectly conserved with the exception of this deletion. The N-terminal alanine stretch as well as the central alanine tract within *ZIC2* also revealed no variation in 192 AD patients and 192 controls. *ZIC2* belongs to a group of transcription factors termed "zinc finger of the cerebellum", which are involved in the development of neural tissues and the neural crest, left-right axis patterning, somite development, and formation of the cerebellum[147]. Interestingly, expansions of this C-terminal alanine tract are frequently observed in patients with holoprosencephaly (HPE)[140]. As this phenotype can be highly variable and lead to mental retardation, a patient with a lack of dysmorphic features could be misdiagnosed with AD.

The family of the patient with the *ZIC2* deletion reported no family history of AD or holoprosencephaly. Acrylamide gel electrophoresis as well as sequencing of DNA taken from lymphoblasts confirmed that the deletion had been transmitted from the unaffected father. This presents three possible scenarios: 1) the mutation is not related to the phenotype, or does not act dominantly to produce any phenotype, 2) the father is somatically mosaic for the mutation, which is a strikingly common scenario for expansions of this C-terminal alanine tract[148], or the mutation occurred in the father's mother and the gene is maternally imprinted, 3) this rare variant requires other genetic or non-genetic factors that were not present during the father's development to produce the phenotype. Due to the severity of the mutation, the function of the gene involved,

and the presence of expansions within this alanine tract that produce dominant phenotypes, it is likely that this mutation is responsible for the phenotype. However, the rare presence of this mutation indicates that mutations in *ZIC2* are not likely to explain a large proportion of AD cases. It is worth noting that the father's male sibling is dyslexic and all of his children have learning disabilities, however DNA is unavailable from them at this time. Further examination of this mutation to identify potential paternal somatic mosaicism, screening of the paternal family and siblings as well as a functional study of this given mutation is required to be certain of the phenotypic role this mutation has on the phenotype.

An 18 bp, six alanine expansion (11-17) was observed in an alanine tract from *TBL1XR1* (alternative name *IRA1, TBLR1*) in three heterozygous patients diagnosed with AD. This alanine tract was perfectly conserved among human patient samples. This variant or any other alanine length variants were not observed within this gene in >1,000 control chromosomes, whose primary origin is French-Canadian, French as well as many from mixed ethnic backgrounds. TBLXR1 is one of 10-12 proteins within the N-CoR complex[149]. In vitro, this protein binds histones H2B and H4, resulting in transcriptional repression. To date, this gene has not been implicated in any human disease.

All three heterozygous patients were of FC origin, with the alanine tract length variant having been inherited paternally in two cases and maternally in one. Genotyping of microsatellite markers surrounding this rare expansion revealed that it was identical by descent (IBD), however these families could not

be connected by an immediate common ancestor. This finding presents implications when interpreting our results, as it implies that the expansion was not only inherited from unaffected parents but also that it has been transmitted through several generations of unaffected individuals despite the non appearance of the variant in >500 controls of which a minimum of 1/5th are of French Canadian origin. Therefore, there are three possible scenarios for the contribution of this variant to the AD phenotype: 1) This relatively rare variant interacts with another variant not present in unaffected family members to produce the AD phenotype 2) this rare variant interacts with an environmental agent that the previous family members were not exposed to 3) this variant does not contribute to the AD phenotype. Determining which of these three scenarios is correct may potentially prove to be very difficult. Simply performing a functional study on the effect of this variant on the gene product may give unsatisfying results. For example, a study on this variant that revealed no functional impact in a cellular or animal model would not necessarily be conclusive that it does not have a functional impact in the presence of another genetic component or environmental agent. Therefore, the validation of *TBLXR1* as a contributing component to the AD phenotype may need to come through replication from different global populations.

POU3F3 located at 2q12.1 encodes Brn-1, a brain expressed transcription factor containing two separate alanine repeats, an N-terminal repeat of 12 alanine repeats and a C-terminal repeat of 16 alanines. This gene belongs to a family of

genes involved in development termed the POU specific homeobox DNA binding domain. This particular protein is also functional candidate for AD as it plays a critical role in neurogenesis acting as a transcriptional activator of Delta1[150]. To date, mutations in this gene have not been implicated directly in any human genetic disease, however the gene is highly conserved in mammals meaning that severe mutations would likely result in embryonic lethality.

The N-terminal, or 5' alanine repeat of 12 alanines was screened in 342 patients diagnosed with AD as well as 192 patients (384 chromosomes). A heterozygous deletion of seven alanines (12-5) was observed in a single patient that was not observed in 384 control chromosomes and was confirmed by sequencing and acrylamide gel electrophoresis. This deletion was paternally transmitted to an affected male child from a family with no history of AD. Interestingly, this affected male had a twin that died at 33 weeks *in utero.* Considering that this was not a *de novo* mutation, and was inherited from an unaffected father with no family history, the same three scenarios apply: 1) the mutation is not related to the phenotype, or does not act dominantly to produce any phenotype 2) the father is somatically mosaic for the mutation, or the mutation occurred in the father's mother and the gene is maternally imprinted 3) this rare variant requires other genetic or non-genetic factors that were not present during the father's development to produce the phenotype. This deletion is large and therefore would likely produce a functional effect on the protein, therefore the possibility of paternal somatic mosaicism should be investigated further, as well

as a screening of more controls. It should be mentioned that the control populations contained three variants in this tract that were not observed in any patients diagnosed with AD: in two controls a heterozygous expansion of two alanines was observed (12-14) as well as a heterozygous deletion of one alanine (12-11). However, these control variants were much smaller and would presumably have less functional impact on the protein.

The C-terminal or 3'alanine tract of *POU3F3* (BRN-1) is larger, containing 16 alanines, and was screened in 342 patients with AD and 192 controls (384 chromosomes). A heterozygous deletion of two alanines (16-14) was observed in one patient that was not observed in 384 control chromosomes. This deletion was once again paternally inherited in a French Canadian family, confirmed by sequencing as well as acrylamide gel electrophoresis. Deletions of a single alanine repeat have been shown to produce phenotypic results, therefore the smaller size of this deletion does not necessarily imply that it is not functional or causal. Nevertheless, a more extensive screening of controls is required before a decision on functional tests can be made.

A heterozygous 21 bp deletion of seven alanines was observed (11-4 alanines) in *ARID1A* (alternative name *SMARCF1*) located at 1p36.11. This gene encodes a member of the SWI/SNF family which is thought to regulate transcription of specific genes by altering chromatin structure in a broad range of tissues[151]. *ARID1A* has been distinctly shown to have tumour suppressor activity by inhibiting gene expression, which leads to cell cycle arrest[152-154]. However, this

gene has never been implicated in inherited forms of human genetic disease. The deletion was observed in 1/192 patients that was not observed in 192 controls and was paternally inherited from an unaffected father with no family history of AD. With the exception of this deletion, perfect conservation was observed in the remaining cases and controls. Although this is a large, rare deletion, both parents are of African descent for which a matching control population was not available at the time of the initial screening. Therefore, further screening of appropriate population matched controls is required before a functional examination of this mutation is warranted. Although outside the scope of this study, given the clearly defined role of this protein in tumour suppression, it may be interesting to investigate the paternal ancestry for a history of inherited forms of cancer, and potentially examine the effect of this mutation on cell cycle arrest.

The N-terminal alanine repeat of *FOXD3* located at 1p31 contains 11 alanine repeats. This gene belongs to the forkhead family of transcription factors involved in transcriptional repression as well as activation. *FOXD3* is thought to promote the specification of the neural crest by inducing the development of neural crest cells from neural tube progenitors[155]. This gene was screened in 384 patients revealing a 12bp, 4 alanine expansion (11-15) in one patient diagnosed with AD with a very diverse familial background that was not observed in 192 controls. Further screening of cases and controls is required to provide enough support to proceed with a functional study on this gene.

5.6 Conclusion

This study was designed to determine whether alanine repeat variants may be worthy of an in depth examination and screening for a contributing role to the AD phenotype. Several rare variants were identified, which may contribute to the AD phenotype and are candidates for functional studies of these variants. The genes selected for study in this group represent a subset of potential genes that could have been chosen. Further screening of additional genes with alanine repeats may be performed depending on whether some of the rare variants can be successfully validated.

Furthermore, this study identified an important class of functional coding variation in the human genome. Homogenizing the samples on different bases, besides those diagnosed with AD and controls, may identify certain variants which segregate with different phenotypes, even if they are neutral morphological features.

Afterword

The questions that most individuals are concerned with regarding autism did not find a suitable section within the body of this thesis. I will therefore address them here. These questions generally include the number of genes that may be involved in the phenotype, the environmental component, the increasing prevalence of autism, as well as future research and a potential cure for the disorder. After studying autism for a brief period during my M.Sc. research, I am left with some opinions on these questions. In my humble opinion, with an attempt at eliminating as much bias as possible, I believe that pure AD is a disorder with an almost exclusive genetic basis. I believe that highly penetrant *de novo* mutations, some of which are transmitted through a few generations in asymptomatic carriers, in as many as 200 and as few as 30 genes, as well as a multitude of *de novo* chromosomal translocations, duplications, deletions and inversions (some of which have yet to occur in nature), can all lead to a phenotype that psychologists currently diagnose as autistic.

I believe that research is proceeding rapidly enough at this point that whole genome sequencing and techniques involving high resolution analysis of copy number variation will soon be financially plausible. These techniques will initially be used as for diagnostic testing and will eventually be used during standard birth examinations. Research across many disciplines will unfold such a substantial amount of information regarding all proteomic interactions, regulation, and effects of interacting genetic variants that, coupled with the ongoing

advancement in computing technology, will lead to a highly accurate diagnosis. This diagnosis will not have to be based on precedence; it will be able to identify the developmental difficulties a child will face based on predictive algorithms regardless of whether a particular new mutation has been previously catalogued, and will account for other variants present within the individual's genome. This diagnosis will be more aptly classified based on the molecular pathology, rather than the psychological pathology which will lead to discontinued use of the term autism, pervasive developmental disorder not otherwise specified, and Asperger's syndrome. I believe these psychological terms will eventually be considered interesting solely from the perspective of medical history, as a reflection of how far things had come by the year 2107.

After research has led to such a profound understanding of systems at the molecular level, in some instances, it may become straightforward to use engineered proteins, or gene therapy as a means to assist the molecular developmental course of a child diagnosed with AD, based on an assessment of the particular molecular pathology in question. At one point in time, I believed that a molecular intervention (gene therapy, engineered proteins) would not be effective at curbing symptoms of AD if it were not administered before the age of onset, as the nervous system would have already failed to properly develop and missed several essential temporal developmental checkpoints that can not be replicated. However, I have reconsidered this position, given research showing that rats with mutations in *MECP2* have a phenotypic reversal when initiating the

expression of functional *MECP2* [156], as well as the difficulty in identifying physical abnormalities in the brains of autistic patients. Ultimately though, I believe the success of a molecular treatment will be dependent on the type of mutation (toxic gain of function, dominant negative) as well as the function of the gene or protein in question, as highly specific temporal and spatial concentrations of a protein may be required and difficult to replicate.

With regards to the increasing prevalence, I believe that the current level of awareness of autism in the community is based on consciousness raising and the shifting of arbitrary boundaries in clinical diagnoses. Although I do believe that higher levels of environmental mutagens in a local area may increase the local mutation rate, I believe that this increase could only go as far as to explain 0.1% or less of the total increase of the prevalence of AD. I am often struck by individuals that are skeptical of shifting diagnostic trends as an explanation for the increasing prevalence for AD, yet are quick to describe any variant psychological feature in children as "autistic" that would have clearly been given a different label ten years ago, from attention deficits, to shy behaviour. Another interesting note is the common belief among those that *are* willing to accept wavering diagnostic criteria for the recent surge in public interest in AD. They will attribute an *improvement* in diagnostic criteria for the increased prevalence. However, if child psychology has remained historically constant, while the behavioural categories that we arbitrarily assign shift within our cultural environment, it becomes a philosophical question as to whether the change was an *improvement*.

My belief is that the tendency to label the diagnostic criteria as an improvement is based on a myth of progress, assuming that all change must be better than the previous system. In the specific question of diagnosing AD, it puzzles me how a shifting diagnosis that increases heterogeneity, decreasing the ability to identify genes, delaying any cure, as well as creating an unfounded public hysteria could be viewed as an "improvement".

I have to re-iterate that I am far from an expert, and these thoughts are merely an interesting and amusing way of concluding a thesis written on a topic over which I have spent many hours deliberating. Should my beliefs turn out to be false, I will be equally excited to witness the actual unfolding of the true answers to these puzzling questions that surround this mysterious disorder.

References

1. Kanner, L., *Autistic disturbances of affective contact.* Acta Paedopsychiatr, 1943. **35**(4): p. 100-36.
2. Asperger, H., *Die autistischen psychopathen im kindesalter.* Archives fur Psychiatrie und Nervenkrankheiten, 1944. **117**: p. 76-136.
3. Folstein, S.E. and B. Rosen-Sheidley, *Genetics of autism: complex aetiology for a heterogeneous disorder.* Nat Rev Genet, 2001. **2**(12): p. 943-55.
4. Mead, M., *Coming of age in Samoa: A psychological study of primitive youth for Western Civilization*, ed. B.R. Books. 1928, New York.
5. Pinker, S., *The Blank Slate: The Modern Denial of Human Nature*, ed. P. Books. 2002, New York.
6. Eisenberg, L., *The fathers of autistic children.* Am J Orthopsychiatry, 1957. **27**(4): p. 715-24.
7. Folstein, S. and M. Rutter, *Infantile autism: a genetic study of 21 twin pairs.* J Child Psychol Psychiatry, 1977. **18**(4): p. 297-321.
8. Muhle, R., S.V. Trentacoste, and I. Rapin, *The genetics of autism.* Pediatrics, 2004. **113**(5): p. e472-86.
9. Fombonne, E. and S. Chakrabarti, *No evidence for a new variant of measles-mumps-rubella-induced autism.* Pediatrics, 2001. **108**(4): p. E58.
10. Bernard, S., et al., *Autism: a novel form of mercury poisoning.* Med Hypotheses, 2001. **56**(4): p. 462-71.
11. Stehr-Green, P., et al., *Autism and thimerosal-containing vaccines: lack of consistent evidence for an association.* Am J Prev Med, 2003. **25**(2): p. 101-6.
12. Thompson, W.W., et al., *Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years.* N Engl J Med, 2007. **357**(13): p. 1281-92.
13. Fombonne, E., *Is there an epidemic of autism?* Pediatrics, 2001. **107**(2): p. 411-2.
14. DSM-IV, A.P.A.T.F.o., *Diagnostic and Statistical Manual of Mental Disorders: DSM IV-TR*, ed. A.P. Association. Vol. 4th Edition. 2000, Washington, DC.
15. Organization, W.H., *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*, ed. W.H. Organization. 1992, Geneva.
16. Jamain, S., et al., *Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism.* Nat Genet, 2003. **34**(1): p. 27-9.
17. Laumonnier, F., et al., *X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family.* Am J Hum Genet, 2004. **74**(3): p. 552-7.

18. Huttenlocher, P.R. and A.S. Dabholkar, *Regional differences in synaptogenesis in human cerebral cortex.* J Comp Neurol, 1997. **387**(2): p. 167-78.
19. Association, A.P., *Diagnostic and Statistical Manual of Mental Disorders.* Fourth Edition ed. 1994, Washington, DC: American Psychiatric Association.
20. Schumann, C.M., et al., *The amygdala is enlarged in children but not adolescents with autism; the hippocampus is enlarged at all ages.* J Neurosci, 2004. **24**(28): p. 6392-401.
21. Wolterink, G., et al., *Early amygdala damage in the rat as a model for neurodevelopmental psychopathological disorders.* Eur Neuropsychopharmacol, 2001. **11**(1): p. 51-9.
22. Baron-Cohen, S., et al., *The amygdala theory of autism.* Neurosci Biobehav Rev, 2000. **24**(3): p. 355-64.
23. Piven, J., et al., *Platelet serotonin, a possible marker for familial autism.* J Autism Dev Disord, 1991. **21**(1): p. 51-9.
24. Leboyer, M., et al., *Whole blood serotonin and plasma beta-endorphin in autistic probands and their first-degree relatives.* Biol Psychiatry, 1999. **45**(2): p. 158-63.
25. Persico, A.M., et al., *Lack of association between serotonin transporter gene promoter variants and autistic disorder in two ethnically distinct samples.* Am J Med Genet, 2000. **96**(1): p. 123-7.
26. Tordjman, S., et al., *Role of the serotonin transporter gene in the behavioral expression of autism.* Mol Psychiatry, 2001. **6**(4): p. 434-9.
27. Yirmiya, N., et al., *Evidence for an association with the serotonin transporter promoter region polymorphism and autism.* Am J Med Genet, 2001. **105**(4): p. 381-6.
28. Betancur, C., et al., *Serotonin transporter gene polymorphisms and hyperserotonemia in autistic disorder.* Mol Psychiatry, 2002. **7**(1): p. 67-71.
29. Fombonne, E., et al., *Microcephaly and macrocephaly in autism.* J Autism Dev Disord, 1999. **29**(2): p. 113-9.
30. Fombonne, E., et al., *Autism and associated medical disorders in a French epidemiological survey.* J Am Acad Child Adolesc Psychiatry, 1997. **36**(11): p. 1561-9.
31. Tuchman, R. and I. Rapin, *Epilepsy in autism.* Lancet Neurol, 2002. **1**(6): p. 352-8.
32. Chakrabarti, S. and E. Fombonne, *Pervasive developmental disorders in preschool children.* Jama, 2001. **285**(24): p. 3093-9.
33. Folstein, S.E. and M.L. Rutter, *Autism: familial aggregation and genetic implications.* J Autism Dev Disord, 1988. **18**(1): p. 3-30.
34. Fombonne, E., *Epidemiological trends in rates of autism.* Mol Psychiatry, 2002. **7 Suppl 2**: p. S4-6.

35. Bailey, A., et al., *Autism as a strongly genetic disorder: evidence from a British twin study.* Psychol Med, 1995. **25**(1): p. 63-77.
36. Steffenburg, S., et al., *A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden.* J Child Psychol Psychiatry, 1989. **30**(3): p. 405-16.
37. Veenstra-Vanderweele, J., S.L. Christian, and E.H. Cook, Jr., *Autism as a paradigmatic complex genetic disorder.* Annu Rev Genomics Hum Genet, 2004. **5**: p. 379-405.
38. Freitag, C.M., *The genetics of autistic disorders and its clinical relevance: a review of the literature.* Mol Psychiatry, 2007. **12**(1): p. 2-22.
39. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.
40. Szatmari, P., et al., *Mapping autism risk loci using genetic linkage and chromosomal rearrangements.* Nat Genet, 2007.
41. Durand, C.M., et al., *Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders.* Nat Genet, 2007. **39**(1): p. 25-7.
42. Turner, G., et al., *Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation.* Am J Med Genet, 2002. **112**(4): p. 405-11.
43. Reddy, K.S., *Cytogenetic abnormalities and fragile-X syndrome in Autism Spectrum Disorder.* BMC Med Genet, 2005. **6**: p. 3.
44. Fombonne, E., *Epidemiological surveys of autism and other pervasive developmental disorders: an update.* J Autism Dev Disord, 2003. **33**(4): p. 365-82.
45. Harrison, J.E. and P.F. Bolton, *Annotation: tuberous sclerosis.* J Child Psychol Psychiatry, 1997. **38**(6): p. 603-14.
46. Eyre-Walker, A. and P.D. Keightley, *High genomic deleterious mutation rates in hominids.* Nature, 1999. **397**(6717): p. 344-7.
47. Giannelli, F., T. Anagnostopoulos, and P.M. Green, *Mutation rates in humans. II. Sporadic mutation-specific rates and rate of detrimental human mutations inferred from hemophilia B.* Am J Hum Genet, 1999. **65**(6): p. 1580-7.
48. Crow, J.F., *The origins, patterns and implications of human spontaneous mutation.* Nat Rev Genet, 2000. **1**(1): p. 40-7.
49. Stephens, K., et al., *Preferential mutation of the neurofibromatosis type 1 gene in paternally derived chromosomes.* Hum Genet, 1992. **88**(3): p. 279-82.
50. Hudson, J., et al., *Novel and recurrent mutations in the neurofibromatosis type 1 (NF1) gene.* Hum Mutat, 1997. **9**(4): p. 366-7.
51. Grimm, T., et al., *On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis.* J Med Genet, 1994. **31**(3): p. 183-6.

52. Hagberg, B. and G. Hagberg, *Rett syndrome: epidemiology and geographical variability.* Eur Child Adolesc Psychiatry, 1997. **6 Suppl 1**: p. 5-7.
53. Van den Veyver, I.B. and H.Y. Zoghbi, *Mutations in the gene encoding methyl-CpG-binding protein 2 cause Rett syndrome.* Brain Dev, 2001. **23 Suppl 1**: p. S147-51.
54. Hagberg, B., *Rett syndrome: Swedish approach to analysis of prevalence and cause.* Brain Dev, 1985. **7**(3): p. 276-80.
55. Ellegren, H., *Microsatellite Mutations in the Germline: implications for evolutionary interference.* Trends in Genetics, 2000. **16**(12): p. 551-558.
56. Stromme, P., et al., *Mutations in the human ortholog of Aristaless cause X-linked mental retardation and epilepsy.* Nat Genet, 2002. **30**(4): p. 441-5.
57. Verkerk, A.J., et al., *Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome.* Cell, 1991. **65**(5): p. 905-14.
58. Li, S., et al., *Matrix assembly, regulation, and survival functions of laminin and its receptors in embryonic stem cell differentiation.* J Cell Biol, 2002. **157**(7): p. 1279-90.
59. Moloney, D.M., et al., *Exclusive paternal origin of new mutations in Apert syndrome.* Nat Genet, 1996. **13**(1): p. 48-53.
60. Glaser, R.L., et al., *Paternal origin of FGFR2 mutations in sporadic cases of Crouzon syndrome and Pfeiffer syndrome.* Am J Hum Genet, 2000. **66**(3): p. 768-77.
61. Jadayel, D., et al., *Paternal origin of new mutations in von Recklinghausen neurofibromatosis.* Nature, 1990. **343**(6258): p. 558-9.
62. Li, R., et al., *Propensity for paternal inheritance of de novo mutations in Alexander disease.* Hum Genet, 2006. **119**(1-2): p. 137-44.
63. Trappe, R., et al., *MECP2 mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin.* Am J Hum Genet, 2001. **68**(5): p. 1093-101.
64. Girard, M., et al., *Parental origin of de novo MECP2 mutations in Rett syndrome.* Eur J Hum Genet, 2001. **9**(3): p. 231-6.
65. Chumakov, I., et al., *Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13675-80.
66. Preuss, T.M., et al., *Human brain evolution: insights from microarrays.* Nat Rev Genet, 2004. **5**(11): p. 850-60.
67. Cantor, R.M., et al., *Paternal age and autism are associated in a family-based sample.* Mol Psychiatry, 2007. **12**(5): p. 419-421.
68. Le Couteur, A., et al., *A broader phenotype of autism: the clinical spectrum in twins.* J Child Psychol Psychiatry, 1996. **37**(7): p. 785-801.
69. Abbadi, N., et al., *Additional case of female monozygotic twins discordant for the clinical manifestations of Duchenne muscular dystrophy due to*

opposite X-chromosome inactivation. Am J Med Genet, 1994. **52**(2): p. 198-206.

70. Baron-Cohen, S., et al., *The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians.* J Autism Dev Disord, 2001. **31**(1): p. 5-17.
71. Treffert, D.A. and D.D. Christensen, *Inside the mind of a savant.* Sci Am, 2005. **293**(6): p. 108-13.
72. Yang, M.S. and M. Gill, *A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence.* Int J Dev Neurosci, 2006.
73. Sladek, R., et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. **445**(7130): p. 881-5.
74. Grant, S.F., et al., *Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes.* Nat Genet, 2006. **38**(3): p. 320-3.
75. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration.* Science, 2005. **308**(5720): p. 385-9.
76. Fombonne, E., *Epidemiology of autistic disorder and other pervasive developmental disorders.* J Clin Psychiatry, 2005. **66 Suppl 10**: p. 3-8.
77. *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
78. Harris, J.R., *Parental selection: a third selection process in the evolution of human hairlessness and skin color.* Med Hypotheses, 2006. **66**(6): p. 1053-9.
79. Nicolson, R. and P. Szatmari, *Genetic and neurodevelopmental influences in autistic disorder.* Can J Psychiatry, 2003. **48**(8): p. 526-37.
80. Baron-Cohen, S., R.C. Knickmeyer, and M.K. Belmonte, *Sex differences in the brain: implications for explaining autism.* Science, 2005. **310**(5749): p. 819-23.
81. Murphy, M., et al., *Personality traits of the relatives of autistic probands.* Psychol Med, 2000. **30**(6): p. 1411-24.
82. Piven, J. and P. Palmer, *Psychiatric disorder and the broad autism phenotype: evidence from a family study of multiple-incidence autism families.* Am J Psychiatry, 1999. **156**(4): p. 557-63.
83. Happe, F., A. Ronald, and R. Plomin, *Time to give up on a single explanation for autism.* Nat Neurosci, 2006. **9**(10): p. 1218-20.
84. Cordero, D., et al., *Temporal perturbations in sonic hedgehog signaling elicit the spectrum of holoprosencephaly phenotypes.* J Clin Invest, 2004. **114**(4): p. 485-94.
85. Deykin, E.Y. and B. MacMahon, *Pregnancy, delivery, and neonatal complications among autistic children.* Am J Dis Child, 1980. **134**(9): p. 860-4.
86. Finegan, J.A. and B. Quarrington, *Pre-, peri-, and neonatal factors and infantile autism.* J Child Psychol Psychiatry, 1979. **20**(2): p. 119-28.

87. Stromland, K., et al., *Autism in thalidomide embryopathy: a population study.* Dev Med Child Neurol, 1994. **36**(4): p. 351-6.
88. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
89. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
90. Veltman, M.W., E.E. Craig, and P.F. Bolton, *Autism spectrum disorders in Prader-Willi and Angelman syndromes: a systematic review.* Psychiatr Genet, 2005. **15**(4): p. 243-54.
91. Buxbaum, J.D., et al., *Association between a GABRB3 polymorphism and autism.* Mol Psychiatry, 2002. **7**(3): p. 311-6.
92. McCauley, J.L., et al., *A linkage disequilibrium map of the 1-Mb 15q12 GABA(A) receptor subunit cluster and association to autism.* Am J Med Genet B Neuropsychiatr Genet, 2004. **131**(1): p. 51-9.
93. Shao, Y., et al., *Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes.* Am J Hum Genet, 2003. **72**(3): p. 539-48.
94. Jentsch, T.J., et al., *Molecular structure and physiological function of chloride channels.* Physiol Rev, 2002. **82**(2): p. 503-68.
95. Bordey, A., *Enigmatic GABAergic networks in adult neurogenic zones.* Brain Res Rev, 2007. **53**(1): p. 124-34.
96. Whiting, P.J., *GABA-A receptor subtypes in the brain: a paradigm for CNS drug discovery?* Drug Discov Today, 2003. **8**(10): p. 445-50.
97. Wafford, K.A., *GABAA receptor subtypes: any clues to the mechanism of benzodiazepine dependence?* Curr Opin Pharmacol, 2005. **5**(1): p. 47-52.
98. Benarroch, E.E., *GABAA receptor heterogeneity, function, and implications for epilepsy.* Neurology, 2007. **68**(8): p. 612-4.
99. Hussman, J.P., *Suppressed GABAergic inhibition as a common factor in suspected etiologies of autism.* J Autism Dev Disord, 2001. **31**(2): p. 247-8.
100. Blatt, G.J., et al., *Density and distribution of hippocampal neurotransmitter receptors in autism: an autoradiographic study.* J Autism Dev Disord, 2001. **31**(6): p. 537-43.
101. Dhossche, D., et al., *Elevated plasma gamma-aminobutyric acid (GABA) levels in autistic youngsters: stimulus for a GABA hypothesis of autism.* Med Sci Monit, 2002. **8**(8): p. PR1-6.
102. Aldred, S., et al., *Plasma amino acid levels in children with autism and their families.* J Autism Dev Disord, 2003. **33**(1): p. 93-7.
103. Moreno-Fuenmayor, H., et al., *Plasma excitatory amino acids in autism.* Invest Clin, 1996. **37**(2): p. 113-28.
104. Mao, R. and S.M. Jalal, *Characteristics of two cases with dup(15)(q11.2-q12): one of maternal and one of paternal origin.* Genet Med, 2000. **2**(2): p. 131-5.

105. Repetto, G.M., et al., *Interstitial duplications of chromosome region 15q11q13: clinical and molecular characterization.* Am J Med Genet, 1998. **79**(2): p. 82-9.
106. Bundey, S., et al., *Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia.* Dev Med Child Neurol, 1994. **36**(8): p. 736-42.
107. Curran, S., et al., *An association analysis of microsatellite markers across the Prader-Willi/Angelman critical region on chromosome 15 (q11-13) and autism spectrum disorder.* Am J Med Genet B Neuropsychiatr Genet, 2005. **137**(1): p. 25-8.
108. Salmon, B., et al., *Absence of linkage and linkage disequilibrium to chromosome 15q11-q13 markers in 139 multiplex families with autism.* Am J Med Genet, 1999. **88**(5): p. 551-6.
109. Maestrini, E., et al., *Serotonin transporter (5-HTT) and gamma-aminobutyric acid receptor subunit beta3 (GABRB3) gene polymorphisms are not associated with autism in the IMGSA families. The International Molecular Genetic Study of Autism Consortium.* Am J Med Genet, 1999. **88**(5): p. 492-6.
110. Solis-Anez, E., et al., *[Molecular analysis of the GABRB3 gene in autistic patients: an exploratory study].* Invest Clin, 2007. **48**(2): p. 225-42.
111. Kim, S.A., et al., *Association of GABRB3 polymorphisms with autism spectrum disorders in Korean trios.* Neuropsychobiology, 2006. **54**(3): p. 160-5.
112. Ashley-Koch, A.E., et al., *An analysis paradigm for investigating multi-locus effects in complex disease: examination of three GABA receptor subunit genes on 15q11-q13 as risk factors for autistic disorder.* Ann Hum Genet, 2006. **70**(Pt 3): p. 281-92.
113. Curran, S., et al., *An association analysis of candidate genes on chromosome 15 q11-13 and autism spectrum disorder.* Mol Psychiatry, 2006. **11**(8): p. 709-13.
114. Cook, E.H., Jr., et al., *Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers.* Am J Hum Genet, 1998. **62**(5): p. 1077-83.
115. Menold, M.M., et al., *Association analysis of chromosome 15 gabaa receptor subunit genes in autistic disorder.* J Neurogenet, 2001. **15**(3-4): p. 245-59.
116. Buettner, V.L., et al., *Analysis of imprinting in mice with uniparental duplication of proximal chromosomes 7 and 15 by use of a custom oligonucleotide microarray.* Mamm Genome, 2004. **15**(3): p. 199-209.
117. Meguro, M., et al., *Evidence for uniparental, paternal expression of the human GABAA receptor subunit genes, using microcell-mediated chromosome transfer.* Hum Mol Genet, 1997. **6**(12): p. 2127-33.
118. Kubota, T., et al., *GABAA receptor beta 3 subunit gene is possibly paternally imprinted in humans.* Am J Med Genet, 1994. **49**(4): p. 452-3.

119. Hogart, A., et al., *15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders.* Hum Mol Genet, 2007. **16**(6): p. 691-703.
120. Dudbridge, F., *Pedigree disequilibrium tests for multilocus haplotypes.* Genet Epidemiol, 2003. **25**(2): p. 115-21.
121. Spielman, R.S. and W.J. Ewens, *The TDT and other family-based tests for linkage disequilibrium and association.* Am J Hum Genet, 1996. **59**(5): p. 983-9.
122. Shaw-Smith, C., et al., *Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability.* Nat Genet, 2006. **38**(9): p. 1032-7.
123. Koolen, D.A., et al., *A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism.* Nat Genet, 2006. **38**(9): p. 999-1001.
124. Raedt, T.D., et al., *Conservation of hotspots for recombination in low-copy repeats associated with the NF1 microdeletion.* Nat Genet, 2006. **38**(12): p. 1419-23.
125. Gimelli, G., et al., *Genomic inversions of human chromosome 15q11-q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions.* Hum Mol Genet, 2003. **12**(8): p. 849-58.
126. Kurotaki, N., et al., *Fifty microdeletions among 112 cases of Sotos syndrome: low copy repeats possibly mediate the common deletion.* Hum Mutat, 2003. **22**(5): p. 378-87.
127. Visser, R., et al., *Identification of a 3.0-kb major recombination hotspot in patients with Sotos syndrome who carry a common 1.9-Mb microdeletion.* Am J Hum Genet, 2005. **76**(1): p. 52-67.
128. Albrecht, A. and S. Mundlos, *The other trinucleotide repeat: polyalanine expansion disorders.* Curr Opin Genet Dev, 2005. **15**(3): p. 285-93.
129. Oberle, I., et al., *Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome.* Science, 1991. **252**(5010): p. 1097-102.
130. Ranum, L.P. and J.W. Day, *Dominantly inherited, non-coding microsatellite expansion disorders.* Curr Opin Genet Dev, 2002. **12**(3): p. 266-71.
131. Ranum, L.P. and J.W. Day, *Myotonic dystrophy: clinical and molecular parallels between myotonic dystrophy type 1 and type 2.* Curr Neurol Neurosci Rep, 2002. **2**(5): p. 465-70.
132. Warren, S.T., *Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13.* Science, 1997. **275**(5298): p. 408-9.
133. Reddy, P.H., M. Williams, and D.A. Tagle, *Recent advances in understanding the pathogenesis of Huntington's disease.* Trends Neurosci, 1999. **22**(6): p. 248-55.
134. Oostra, B.A. and P. Chiurazzi, *The fragile X gene and its function.* Clin Genet, 2001. **60**(6): p. 399-408.

135. Brais, B., *Oculopharyngeal muscular dystrophy: a late-onset polyalanine disease.* Cytogenet Genome Res, 2003. **100**(1-4): p. 252-60.
136. Nasrallah, I.M., J.C. Minarcik, and J.A. Golden, *A polyalanine tract expansion in Arx forms intranuclear inclusions and results in increased cell death.* J Cell Biol, 2004. **167**(3): p. 411-6.
137. Brown, L., et al., *In vitro analysis of partial loss-of-function ZIC2 mutations in holoprosencephaly: alanine tract expansion modulates DNA binding and transactivation.* Hum Mol Genet, 2005. **14**(3): p. 411-20.
138. Crisponi, L., et al., *The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome.* Nat Genet, 2001. **27**(2): p. 159-66.
139. Brais, B., et al., *Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy.* Nat Genet, 1998. **18**(2): p. 164-7.
140. Brown, S.A., et al., *Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired.* Nat Genet, 1998. **20**(2): p. 180-3.
141. Goodman, F.R., et al., *Novel HOXA13 mutations and the phenotypic spectrum of hand-foot-genital syndrome.* Am J Hum Genet, 2000. **67**(1): p. 197-202.
142. Muragaki, Y., et al., *Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13.* Science, 1996. **272**(5261): p. 548-51.
143. Mundlos, S., et al., *Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia.* Cell, 1997. **89**(5): p. 773-9.
144. Laumonnier, F., et al., *Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency.* Am J Hum Genet, 2002. **71**(6): p. 1450-5.
145. Amiel, J., et al., *Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome.* Nat Genet, 2003. **33**(4): p. 459-61.
146. Fondon, J.W., 3rd and H.R. Garner, *Molecular origins of rapid and continuous morphological evolution.* Proc Natl Acad Sci U S A, 2004. **101**(52): p. 18058-63.
147. Merzdorf, C.S., *Emerging roles for zic genes in early development.* Dev Dyn, 2007. **236**(4): p. 922-40.
148. Brown, L.Y., et al., *Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination.* Hum Mol Genet, 2001. **10**(8): p. 791-6.
149. Yoon, H.G., et al., *Purification and functional characterization of the human N-CoR complex: the roles of HDAC3, TBL1 and TBLR1.* Embo J, 2003. **22**(6): p. 1336-46.
150. Castro, D.S., et al., *Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif.* Dev Cell, 2006. **11**(6): p. 831-44.

151. Wang, X., et al., *Two related ARID family proteins are alternative subunits of human SWI/SNF complexes.* Biochem J, 2004. **383**(Pt 2): p. 319-25.
152. Nagl, N.G., Jr., et al., *The p270 (ARID1A/SMARCF1) subunit of mammalian SWI/SNF-related complexes is essential for normal cell cycle arrest.* Cancer Res, 2005. **65**(20): p. 9236-44.
153. Wang, X., et al., *Expression of p270 (ARID1A), a component of human SWI/SNF complexes, in human tumors.* Int J Cancer, 2004. **112**(4): p. 636.
154. Huang, J., et al., *Genomic and functional evidence for an ARID1A tumor suppressor role.* Genes Chromosomes Cancer, 2007. **46**(8): p. 745-50.
155. Barembaum, M. and M. Bronner-Fraser, *Early steps in neural crest specification.* Semin Cell Dev Biol, 2005. **16**(6): p. 642-6.
156. Guy, J., et al., *Reversal of neurological defects in a mouse model of Rett syndrome.* Science, 2007. **315**(5815): p. 1143-7.