# Stony Brook University

# The Effect of chr16p11.2 Microdeletions and Microduplications on Gene Expression in Autism Spectrum Disorders and Schizophrenia.

A dissertation presented by
Mary Kusenda

to
The Graduate School
In partial fulfillment of the requirements
For the degree of

Doctor of Philosophy
in
Genetics

Stony Brook University
August 2010

Stony Brook University
The Graduate School

Mary Kusenda

We the dissertation committee for the above candidate for the Doctor in Philosophy degree, hereby recommend acceptance of this dissertation

Dr. Jonathan Sebat - Thesis Advisor
Assistant Professor, Psychiatry, Cellular and Molecular Medicine
University of California, San Diego

Dr. Turhan Canli – Chairperson of Defense
Associate Professor, Biopsychology
Stony Brook University

Dr. Marian Evinger
Associate Professor, Pediatrics Neurobiology and behavior
Stony Brook University

Dr. Joshua Dubnau
Associate Professor, Genetics
Cold Spring Harbor Laboratory

Dr. Lilia Iakoucheva
Assistant Professor, Psychiatry
University of California, San Diego

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

**The Effect of chr16p11.2 Microdeletions and Microduplications on Gene Expression in Autism Spectrum Disorders and Schizophrenia.**

By

Mary Kusenda

Doctor of Philosophy

in

Genetics

Stony Brook University

2010

The number of rare variants found to be associated with multiple psychiatric disorders is growing. One such locus is a recurrent ~600kb copy number variant (CNV) at 16p11.2, occurring in approximately 1% of autism and 0.3% of schizophrenia cases, as compared to 0.01% of the general population. (Sebat 2007, Kumar 2007, Weiss 2008, McCarthy 2009). This mutation has been found at a higher frequency in autistics and schizophrenics, but is also found in patients with developmental delay without an autistic diagnosis, and is also rarely seen in healthy individuals. Head circumference is observed to be smaller in deletion cases versus duplication cases. (McCarthy 2009) We hypothesize that one or more of the 25 genes at this locus contribute to the neurodevelopmental phenotype observed in patients with psychiatric disorders. To determine how gene function is altered by this CNV, we analyzed genome wide expression data from Epstein Barr Virus (EBV) transformed Lymphoblast cell lines (LCL), of patients with autism or schizophrenia who have deletions of reciprocal duplications of 16p11.2. Using RNA expression profiling by Affymetrix Human Genome U133 Plus 2.0 chip, we examined differential *dosage* and *trans* gene expression in individuals with 1, 2, or 3 copies of the genomic region. (6, 19, 16 respectively) To avoid skewing of data due to limited sample size we customized the Significance Analysis of Microarrays (SAM) method to utilize all samples while accounting for sources of bias. Our data highlighted 7 genes located both within and outside of the mutation which expression correlates with genotype. Some of these genes play a role in development while others have been associated with psychiatric disorders. We have analyzed our list of 7 dysregulated genes to identify pathways and functions relevant to neurodevelopment, and psychiatric disorders. Data generated by this study will give insight into dosage sensitive genes within the risk variant, and may help pinpoint genes which are relevant to pathology of the psychiatric disorders associated with this region.

## TABLE OF CONTENTS

## LIST OF FIGURES

**Chapter 1**



**Chapter 2**



**Chapter 3**



**Chapter 4**



**Chapter 5**



**Supplementary Chapter A**

## LIST OF TABLES

### Chapter 1



### Chapter 2



### Chapter 3



### Chapter 4



### Chapter 5

**Supplementary Chapter A**

## LIST OF ABBREVIATIONS

ABI – Applied Biosystems
aCGH array Comparative Genome Hybridization
ADI-R Autism Diagnostic Interview-Revised
ADOS – Autism Diagnositc Observation Schedule
AGRE - Autism Genetics Resource Exchange
ASD - Autism Spectrum Disorders
BDNF – Brain-Derived Neurotropic Factor
BDV- Borna Disease Virus
BMP – Bone Morphogenic Factor
CATIE - Clinical Antipsychotic Trials of Intervention Effectiveness
CDC – Centers for Disease Control
CHOP - Children's Hospital Pennsylvania
CMP1A Charcot-Marie Tooth neuropathy type A
CNS- Central Nervous System
CNV, Copy Number Variations
DGS/VCFS - DiGeorge velocardiofacial syndrome
DNA Deoxyribonucleic Acid
dpc – days post conception
DSM- V - The Diagnostic Manual of Mental Health
EBV, Epstein Barr Virus
EFT – Embedded Figure Test
FDR – False Discovery Rate
FISH Fluorescence in-situ Hybridization
GABA – Gamma-aminobutiric Acid
GATHER- Gene annotation Tool to Help Explain Relationships
GO- Gene Ontology
HD2 – High density 2 million
HIV Human Immunodeficiency Virus
ID - Identifier
IGA – Individual Gene Analysis
IPA, Ingenuity Pathway Analysis
IQ – Intelligence Quotient
LCLs Lymphoblast Cell Lines
LCRs Low Copy Repeats
MAPK- Mitogen Activated Protein Kinase
MCHG- Marshfield Clinical Human Genetics
MR- Mental Retardation
NAHR - Non-Homologous Recombination
NF1 - Neurofibromatosis
NIMH- National Institute of Mental Health
OFC – Occipital- Frontal-Circumference
OFD - oro-facio-digital
OMIM – Online Inheritance in Man

PDD-NOS - Pervasive Developmental Disorder Not Otherwise Specified
PFO - Patent Foramen Ovale
Q-Q – Quantile-Quantile Plot
RMA – Robust Medioid Averaging
RNA - Ribonucleic Acid
ROH - Region of Homozygosity
RT-PCR, Reverse Transcriptase Polymerase Chain Reaction
SAM - Significance Analysis of Microarrays
SNPs - Single Nucleotide Polymorphisms
TGF- beta – Transforming Growth Factor
UCLA – University of California Los Angeles
UCSC University of California Santa Cruz
UGET – UCLA Gene Expression Tool

# CHAPTER 1
# Introduction

**Introduction:** A rare recurrent deletion on 16p11.2 is associated with Autism Spectrum Disorders (ASD) (0.5-1%), and developmental delay (0.1%). The reciprocal duplication is associated with schizophrenia (0.3%), bipolar disorder and developmental delay (0.4%). The deletion and duplication are also rarely found in the general population at the rate of 0.01%, and 0.03% respectively [1-5]. One or more of the 27 genes associated with the variation may be contributing to the psychiatric phenotype associated with the recurrent mutations. Some of these 27 genes may be dosage sensitive. Altering the copy number of such genes may result in altered expression, which may affect neurodevelopmental pathways, resulting in a neuropsychiatric phenotype.

## Copy Number Variations and Neurological disorders

### Structural Variations and Disease

Structural variation or Copy Number Variations (CNVs) make up to 12-15% of genetic variability and are the greatest source of genetic variation between healthy individuals [6-8]. Unlike Single Nucleotide Polymorphisms (SNPs), which are single nucleotide base changes, CNVs can affect between 1kb and 1Mb of genetic material in the form of deletions, duplications, translocations, inversions, isochromosomes or other large alteration of genetic structure [7, 9]. CNVs which are rare or sporadic, are more likely to be *de novo*, or resulting from either a mutation in a parent's germ line during meiosis, or from a mutation in a somatic cell during development. It is estimated that 1 in 8 infants are born with a *de novo* deletion, and 1 in 50 with a *de novo* duplication [10]. Additionally, *de novo* deletions are more pathologically hazardous than duplications [11]. Deletions larger than 50kb are found randomly through out the chromosomes, and are rare which suggests purifying selection. Duplications may remain longer in the population [12]. The majority of CNVs are passed on from parent to child and do not cause maladaptive consequences [7, 11]. However in some cases rare or rare-recurrent CNVs have been associated with genetic syndromes, autoimmune disease, neuropsychiatric disorders, and even an increased risk of HIV infection [6].

*De novo* CNVs have been associated with neurological disorders. In a landmark study *de novo* copy number variations were found in 10% of families with sporadic autism, or families with only one autistic member, as compared to 1% of the general population [11]. This shows a 10 fold increase in *de novo* mutations in the sporadic autism population [13]. Another study determined that *de novo* CNVs were 8 times more likely to be found in schizophrenia cases as compared to controls (p value =0.00078) [14].

### Genomic Disorders

Our knowledge on the role of structural variations in human disease comes from literature on genomic disorders. A genomic disorder is a specific disease phenotype which is associated with recurrent structural mutations at a single locus. Among the best described is Charcot-Marie-Tooth neuropathy type A (CMP1A) a neurodegenerative disorder associated with a microduplication of a

locus on chromosome 17. Hereditary neuropathy with liability to pressure palsies was associated with the reciprocal deletion [15-17]. Recently a rare recurrent deletion on 16p11.2 is associated with Autism Spectrum Disorders (ASD), and developmental delay, while the reciprocal duplication is associated with schizophrenia, bipolar disorder and developmental delay. [1-5]. In both of these examples different copies of the same genomic region can cause such different phenotypical manifestations.

Structural variations can result in phenotypic effects [18]. One hypothesis explaining how certain CNVs contribute to a disease phenotype is that a CNV contains dosage sensitive genes in which the alteration of copy number alters the level of gene expression resulting in the disruption of molecular pathways and a disease phenotype [19]. Cases where the expression of those genes is critical for normal function and development, may result in disease and disorders [19]. A well known disorder resulting from chromosomal dosage effect is trisomy 21, or down syndrome. Patients have an extra copy of chromosome 21, and usually exhibit common characteristics resulting from a gene dosage effect [20, 21]. Alternatively deletions and insertions can disrupt DNA coding regions, resulting in truncated non functional genes, or fusion genes [6, 22]. An example of a fusion gene is Disrupted in Schizophrenia 1, or DISC1 a gene implicated in schizophrenia. This fusion gene was found in a family in which psychiatric disorders segregated with the mutation [23-25]. Sometimes the deletion of a genomic region may uncover deleterious SNPs on the remaining heterozygous allele which may result in a disease phenotype [9]. Deletions and duplications which take place in the intergenic regions may still influence gene expression, by changing the position of enhancers or other regulatory elements. Disorders associated with these rare recurrent CNVs may be influenced by many genetic, epigenetic, and environmental factors [1].

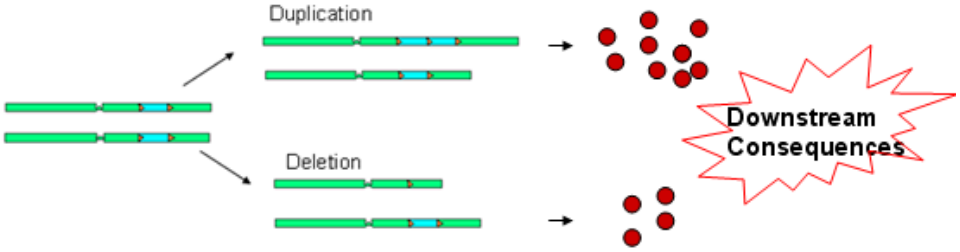



**Figure 1.1 Expression changes resulting from deletions and duplications**
There is an association between a 16p11.2 CNV and a neurodevelopmental phenotype. We hypothesize that changes in expression of one or more of the 27 genes within the 600 Kb region leads to a dysregulation of developmental pathways. This cartoon depicts the rational for our study: Figure 1.1 is a cartoon of chromosome 16, depicted here in green. The region in blue represents the 16p11.2 microduplication, and microdeletion. In an individual with a microdeletion, we expect gene expression of genes within the CNV to be decreased as compared to individuals with three copies of the microduplication.

The dots in red represent the expression level of genes as a result of a microdeletion, or microduplication. We also hypothesize that the dysregulation of these genes will affect other genes downstream of their respective molecular pathway.

**Mechanisms by which structural variations form.**

Certain repetitive sequences which make up the genomic architecture are more fragile and susceptible to chromosome breakage. These repetitive sequences increase the chance for mediating structural rearrangements resulting in deletions, and duplications [19]. These Low Copy Repeats (LCRs) LCRs are also called duplicons, or segmental duplications, and commonly flank CNVs associated with genomic disorders. These are stretches of DNA which have 90% sequence similarity between them, are greater than 1kb and account for 5% of the human DNA sequence [22, 26]. Regions of LCRs are 4-5 times more likely to be associated with CNVs, and are rearrangement hotspots, as compared to control regions [9, 22, 26, 27].

There are certain mechanisms which cause these rearrangements to occur. Homologous recombination is a process during meiosis where the sister chromatics exchange DNA material [19]. This process is important in increasing genetic variation [15]. Non-homologous recombination (NAHR) occurs if the regions of recombination are in regions of repeat sequences. NAHR is one of several mechanisms for structural rearrangement, and is the predominant mechanism driving reciprocal genomic disorders. During NAHR sequences which recombine may be on different chromosomes, or on different sequences of the homologous chromosome [15]. Due to unequal crossing over, this process results in a structural rearrangement, such as a chromosomal deletion or chromosomal duplication [2, 9, 15, 28, 29]. If the LCRs are directly oriented the results are duplicated or deleted sequences, while if they are inversely oriented inversions are generated. The results of this mechanism can sometimes lead to genomic disorders [15].

With the advent of array technology, an increasing number of CNVs flanked by segmental duplications that are associated with clinical disorders, have been discovered. This list includes 1q21.1, 3q29, 17q12, 22q11.2 [1]. The 16p11.2 microdeletion and reciprocal microduplication is known to occur as a result of NAHR [2]. This 600kb CNV is flanked by 147kb segmental duplications with ~99% sequence similarity [2]. LCRs which flank the variation contain 27 annotated genes inside of the variation, of which three are found in the segmental duplication [2, 5, 30].

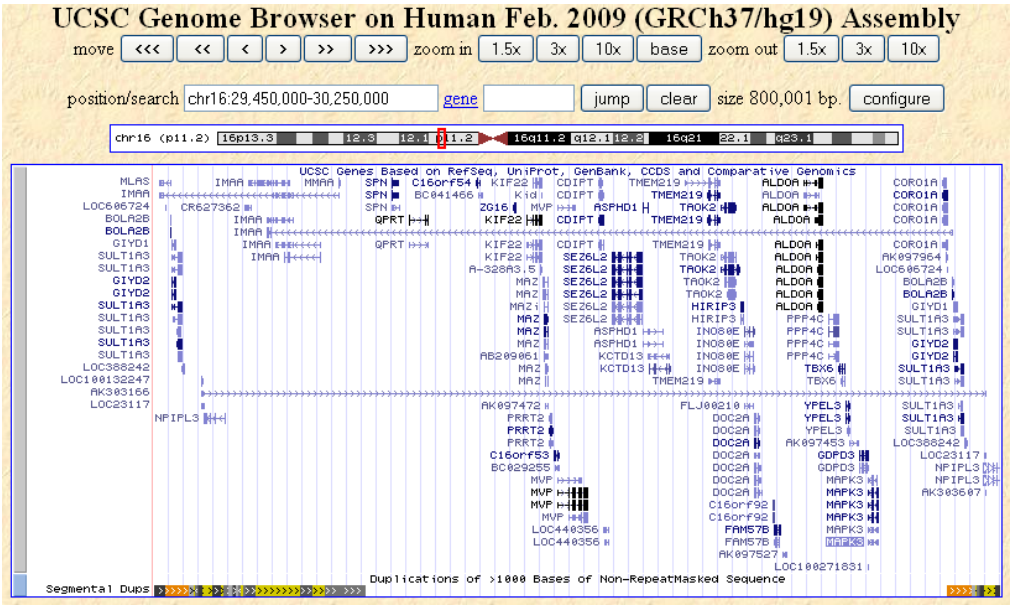



**Figure 1.2 The 16p11.2 interval from the UCSC genome Browser**
Figure 1.2 shows the 600kb interval on chromosome 16, which is associated with autism, schizophrenia, and developmental delay. Towards the bottom of the figure, are rectangles which represent repeat sequences and the segmental duplications, which are believed to mediate the variation. Approximate coordinates of the 1611.2 variation are chr16:29557498-30107355.

From: UCSC Genome Browser: Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res.* 2002 Jun;12(6):996-1006.

**Methods for detection of CNVs**

Tools such as array Comparative Geneome Hybridization microarray (aCGH) technology, and cytogenetic techniques have been used to determine genomic regions which contribute to disease etiology. In the past cytogenetic methods such as Fluorescence in situ hybridization (FISH) were routinely used in clinics to diagnose children with large chromosomal variants [31]. Chromosomal variants which are smaller than 3Mb are difficult to visualize by cytogenetic methods, but aCGH technology has rapidly advanced the ability to visualize CNVs accurately and at a high resolution [8]. This technology makes it possible to correlate the smallest structural variation with increased disease risk. In one example, DNA microarrays have been used to detect genome rearangments which exist in tumor samples as compared to normal [8, 31]. Currently gene chips are being widely used to discover genomic variants in neuropsychiatric disease. aCGH consists of using thousands or millions of short DNA fragments called probes which are fixed onto a glass plate [31].Chemically or mechanically digested DNA fragments from a test sample, and control samples are placed on the array to competitively to hybridize to their complementary sequence. Comparing the test genome to the reference genome can help visualize copy

number gains and losses [31]. Once a recurrent CNV or other region of interest is found, tiling and targeted array platforms can further discern the genomic breakpoints of the CNV [8]. The improvement in targeted array platforms can enhance the discovery of smaller *de novo* and inherited CNVs. After finding a candidate region studies can be done to determine the gene or genes within the variation which may be contributing to pathogenesis. Sequencing genes found in candidate regions or using expression microarrays to determine dosage sensitive genes are methods used to determine the association between genomic region and a disease or syndrome [8, 32].

**History of discovery:**

The association of the 16p11.2 mutation with an increased risk for neuropsychiatric disease was recently discovered through several independent microarray studies. The variation was first found in single cases and then as a rare recurrent variation associated with autism, schizophrenia and developmental delay. Currently more work is being done to analyze the phenotype associated with the variation, in the hopes of determining better genotype-phenotype correlations. Initially the 16p11.2 deletion was believed to be solely associated with autism due to ascertainment bias. The samples included in the initial studies were those which had a diagnosis of autism. Since this time more cases have been analyzed and the 16p11.2 variation has been associated with other psychiatric disorders and cognitive deficits.

Initially, the16p11.2 variation was reported as single cases. A study which discovered an increased *de novo* CNV rate in simplex autism families included a female Asperger patient, with a *de novo* 16p11.2 deletion [11]. Around the same time Kumar *et al* and Weiss *et al* almost simultaneously published papers which used different aCGH platforms to independently discover the association of the CNV with autism (~1%), and the enrichment of the 16p11.2 microdeletion in developmental delay [2, 5]. Bijlisma *et al*, also showed that cases with the 16p11.2 microdeletion also exhibited mental retardation without an autism diagnosis.

McCarthy *et al*, reported a significant association of the microduplication with schizophrenia, and bipolar disorder. Interestingly, the reciprocal microdeletion was not associated with schizophrenia, or bipolar disorder. The 16p11.2 duplication was also reported in a separate study of early onset schizophrenia patients at an even higher rate (2%) than found in adults [33].

In light of these studies, one can theorize that the deletion is specific to the neurodevelopmental changes involved in the development of autism, while the duplication of the region is involved in the pathogenesis of schizophrenia, autism, and bipolar disorder.

Several papers which tried to determine a common clinical phenotype associated with the microdeletion and microduplication were published. Fernandez *et al*, and Shinawi *et al* both reported detailed cases reports of 16p11.2 patients with the goal of defining a common set of facial dysmorphisms, and behavioral characteristics common to 16p11.2 carriers. [29, 34]. Common facial dysmorphisms reported in the deletions include broad forehead, flat midface,

hypertelorism, and micrognathia [29]. Deletion cases in particular had structural brain abnormalities which did not show any commonality between patients. Duplications did not exhibit common facial dysmorphisms, nor had as many reports of brain lesions. Congenital abnormalities found in deletion cases include numerous organ systems including congenital heart defects, cleft palate, multicystic dysplastic kidney, pyloric stenosis, fusion of the lower ribs. Duplication cases had less severe abnormalities including cleft lip, pectus excavatum, torticollis, pes planus, and mild scoliosis [29]. Interestingly a 16p11.2 mosaic deletion patient reported in Bijlisma *et al.* had the most severe phenotype observed in their study [35]. Several interesting individual case reports can be found in table 1.1.

The most recent data on the 16p11.2 region is an association with severe obesity. The 16p11.2 deletion was significantly associated with obesity and morbid obesity while the reciprocal duplication was not associated with either.

With every additional case more light is shed on the phenotype associated with these patients. Although there is a broad psychiatric phenotype associated with 16p11.2 patients one common thread associated with cases is speech and language delay [29, 34, 36]. Less common but frequent is motor delay. In addition to the cognitive phenotypic variability, patients with the 16p11.2 variation have other physical defects, or congenital abnormalities [36].

| Phenotype | Reference | CNV |
|---|---|---|
| Mental retardation, developmental delay, bone malformations or ribs and vertabrae, inguinal hernai, astigmatism | Shimojima (2009) | Deletion |
| Immunodeficiency, low/average cognitive function, attention difficulties | Shiow (2009) | Deletion and 2bp deletion on COROLA1 |
| seizures, mild developmental delay, multiple congenital abnormalities, bicuspid aortic valve malformation | Ghebranious (2007) | Deletion |
| Wilms Tumor (2 years 9 months) | Bijlisma (2009) | Deletion |

**Table 1.1 Phenotype of select 16p11.2 carriers**

### Neuropsychiatric disorders associated with the 16p11.2 region

Both autism and schizophrenia are classified as neuropsychiatric disorders, a class of disorders involving higher cognitive functions associated with the cerebral cortex, and the limbic system [37, 38]. It is believed that psychiatric disorders are influenced by genetic, epigenetic and environmental

factors, and affect thought processes and behavior [9, 39]. Most neuropsychiatric diseases are highly heritable suggesting a strong genetic component [12].

**Autism Spectrum Disorders (ASD)**

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder categorized as a pervasive developmental disorder. Inheritance rates are between 70-90% for monozygotic twins. Results from linkage and association studies strongly support a strong genetic component influencing pathogenesis [4, 40]. The Diagnostic Manual of Mental Health (DSMV) describes autistics as having three core symptoms which include problems with (1) social skills, (2) deficits in verbal communication or language, and (3) restricted or repetitive behaviors.

Autism Spectrum Disorders is an umbrella term for anyone on the autistic spectrum. These individuals can be diagnosed with autistic disorder, Asperger syndrome, Rett syndrome, and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). Asperger syndrome is a milder version called autsim. Autistic disorder is generally considered more severe. Rett syndrome is a monogenic disorder with autistic features. This syndrome has been linked to a gene on the X chromosome MECP2 and affects only girls [41-43]. Autistic individuals present with a range of heterogeneous symptoms which include seizures (found in 6-60%) macrocephaly (20% of 2-3 year olds) gastrointestinal problems (4-45%), sensory abnormalities (94%), and co-morbid psychiatric diagnosis (25-70%) [41] [44].

Heterogeneous physical symptoms and a range of severity make autism difficult to diagnose. Tools such as the Autism Diagnostic Observation Schedule (ADOS), and Autism Diagnostic Interview-Revised (ADI-R) are seen as the gold standard, and have been crucial to containing uniformity in diagnostics. Symptoms commonly occur before the age of 18 months with diagnosis around 2 years of age [44]. Race, ethnicity, and socioeconomic status do not significantly affect severity, or onset of the disorder [45]. In both Europe and the United States there is believed to be a rise in diagnosed cases. An estimated prevalence rate in the US is 60 in 10,000 children. (US CDC 2007) This estimate is higher for males, as they are four times more likely to develop autism than females.

Pinpointing a concrete cause for ASD remains elusive. Candidate genes and CNVs associated with ASD are found among the 23 autosomes and the X, and Y chromosomes [41]. The most frequent rearrangements, (15q11, 22q13, 2q31) are observed in less than 1% of patients [41]. Other large cytogenetic abnormalities are found in only 6-7% of ASD cases [41]. This genetic heterogeneity makes it difficult to pinpoint a biological mechanism or pathway which contributes to pathogenesis [42]. With the knowledge from rare recurrent CNVs, and other candidate regions and genes a common biological similarities have emerged which include synaptic dysfunction, and defective brain connectivity, serotonin transport, dopaminergic activity, and oxytocin [41, 42].

**Schizophrenia**

Schizophrenia is a deteriorating, debilitating, psychiatric disorder characterized by positive symptoms, (hallucinations, delusions, racing thoughts), negative symptoms (apathy, lack of emotion, low energy), cognitive disorganized thinking, difficulty concentrating, completing tasks, expressing thoughts, and memory problems [14, 33, 46, 47]. Schizophrenia is a complex genetic disorder with a heritability of 50% between monozygotic twins [47]. Patients exhibit risky behavior such as smoking, and suicide, thus it is responsible for more deaths than most cancers [48]. It is estimated that the worldwide prevalence is 1%, placing a great burden on society government systems [39]. Onset is usually in late teens or twenties, with later onset in females [48]. Although rare there is a type of early onset schizophrenia in which symptoms appear between ages 6-12. Cytogenetic abnormalities have been associated with early onset cases at the rate of 10% with the 16p11.2 duplication associated with 2% [33]. This rate is significantly higher than seen in adult schizophrenia. The early onset, coupled with varied response to medication creates a strain on individual and family members [33, 47].

Both environmental and genetic factors are thought to contribute to increased risk factors for developing schizophrenia. Environmental factors which influence schizophrenia susceptibility include vitamin D intake during pregnancy, winter birth, birth during times of famine, maternal infection, prenatal exposure to viral infections, living in an urban area, exposure cannabis use and others [47, 48]. The major gene which contributes to schizophrenia is *disrupted in schizophrenia 1*(DISC-1). DISC-1 is an important scaffold protein which was discovered as a translocation of chromosome 1 in a large Scottish family with psychiatric illness [49-51]. There are many pathophysiological models regarding the biological manifestation of schizophrenia. Several plausible mechanisms include abnormal neurotransmitter function, disturbances in the synaptic cleft between axons and dendrites, abnormal connections between neurons, glutamate signaling (NMDA), and Gamma-aminobutyric acid GABA signaling **[46].**

**Autism and Schizophrenia – linked?**

Autism, schizophrenia, bipolar disorder, and developmental delay have all been associated with the 16p11.2 mutation. Today autism and schizophrenia are considered to be separate disorders, but almost a hundred years ago they were considered the same disorder. What we now call “autism” used to be characterized as a subset of schizophrenic patients who were withdrawn from reality [52]. With the current increase in aCGH studies more genes and genetic variants which overlap between autism and schizophrenia have been discovered [53]. Recently the field of psychiatric genetics has shown an increased interest in rare recurrent CNVs which are not disease specific, but are associated with a heterogeneous psychiatric phenotype [53, 54].

Genetic regions which are associated with a heterogeneous psychiatric phenotype have led to the hypothesis that autism and schizophrenia are linked, or even on opposite sides of the spectrum. While schizophrenics tend to read into social relationships, autistics are challenged in social situations [53, 55]. Autism is associated with macrocephaly, or a larger head circumference, while

schizophrenia is associated with reduced brain growth [53]. Autistics have a preference for local processing, and pay close attention to details [54]. They do very well in Embedded Figure Tests (EFTs) which required finding a shape inside of a large picture [56]. Schizophrenics have a preference for global processing, and are better at seeing the bigger picture. This evidence suggests that autism and schizophrenia may be two sides of the same coin [54].

CNVs and candidate genes which are associated with a heterogeneous phenotype include 1q21.1, 15q13.3, 16p11.2, 16p13.1, 17p12, 22q11.21, 22q13.3, Neurexin1, DAO, DISC1, and can be found in patients diagnosed with autism, schizophrenia, bipolar disorder, mental retardation, ADHD or patients which are co-morbid for more than one disorder [39, 53, 57]. Conversely, families in which one of these rare recurrent CNVs segregates are found to have a spectrum of these psychiatric disorders, which can include autism and schizophrenia [3]. 16p11.2 and 22q13.3 microdeletion carriers are associated with increased risk of developing autism, as microduplication carriers have an increased risk of developing schizophrenia [53]. This puts the 16p11.2 variation in the growing list of CNVs and candidate genes which are challenging the separation of autism and schizophrenia into two separate disorders.

**Table 2 Meta-analysis of 16p11.2 rearrangements in schizophrenia, autism and developmental delay, and bipolar disorder**

| | Subjects | Deletions | | | | Duplications | | | |
|---|---|---|---|---|---|---|---|---|---|
| Diagnosis | *n* | *n* | % | OR (95% CI) | *P* value | *n* | % | OR (95% CI) | *P* value |
| Schizophrenia | 8,590 | 3 | 0.03 | NC[a] | | 26 | 0.30 | 8.4 (2.8, 25.4) | $4.8 \times 10^{-7}$ |
| Controls | 28,406 | 9 | 0.03 | | | 8 | 0.03 | | |
| Autism or developmental delay | 2,172 | 17 | 0.78 | 38.7 (13.4,111.8) | $2.3 \times 10^{-13}$ | 10 | 0.46 | 20.7 (6.9,61.7) | $1.9 \times 10^{-7}$ |
| Controls | 24,891 | 5 | 0.02 | | | 6 | 0.02 | | |
| Bipolar disorder | 4,822 | 4 | 0.08 | NC[a] | | 6 | 0.12 | 4.3 (1.3, 14.5) | 0.017 |
| Controls | 25,225 | 6 | 0.02 | | | 7 | 0.03 | | |

**Table 1.2 Association of 16p11.2 region with neuropsychiatric disease**
Data from four studies reporting an association of 16p11.2 with neuropsychiatric disorders were combined with a primary sample in McCarthy *et al* set to determine the strength of the association with a particular disorder. Both the microdeletion and microduplication is significantly associated with autism and developmental delay. Only the microduplication is significantly associated with schizophrenia and bipolar disorder.
Table taken from: McCarthy *et al*. Microduplications of 16p11.2 are associated with schizophrenia Nature Genetics 41, 1223 - 1227 (2009) Published online: 25 October 2009 | doi:10.1038/ng.474

**Penetrance and phenotypic variability of the 16p11.2 mutation.**

Penetrance is the percentage of individuals with the same genetic variation that present with the associated phenotype [58]. These gene variations are influenced by genetic, epigenetic, and environmental factors [1]. Inheritance patterns suggest that the 16p11.2 microdeletion has greater penetrance than the microduplication. Deletion patients seem to be more severely affected than duplication patients, while both are reported to have facial dysmorphisms, and

cognitive dysfunction [29]. Duplications associated with disease have a tendency to be under diagnosed because of reduced penetrance [12]. This is evident as 16p11.2 duplication cases are three times more likely to be found in the general population than deletions [1, 12]. Although it is rare there have been a few cases of deletion patients which have not presented with neuropsychiatric phenotype, as the microdeletion is present in 0.01% of the general population [1]. A few mosaic 16p11.2 cases have been discovered. The one case which had phenotypic data showed a severe phenotype, which may suggest another mutation elsewhere in the genome is influencing phenotype. More needs to be done to determine if mosaicism contributes to penetrance and variability [35].Together this shows the extensive phenotypic variability associated with this mutation. This may suggest that there is another locus outside of the 16p11.2 variation influencing the neuropsychiatric phenotype, although this is not likely as 90-100% of carriers have cognitive deficits, regardless of proper psychiatric diagnosis. The genes within the region are not believed to be imprinted, which suggests that imprinting is not a cause for the variability in phenotype. (http://geneimprint.com as of July 2010)The phenotypic variability may be caused by deleterious SNPs or CNVs on the remaining 16p11.2 allele, but more likely it is caused by variable gene expression [1, 59]. Certain dosage sensitive genes in the variation may be influencing the variable phenotype.

**Candidate genes within the 16p11.2 genomic interval. (See Table 1.3)**

The 16p11.2 mutation associated an increased risk for neuropsychiatric disease encompasses twenty five genes, with three found in the segmental duplications. It is unclear which one or more of the 16p11.2 genes play a role in the pathogenesis of neuropsychiatric disease. Candidate genes found in the variation include seventeen genes which are expressed in the brain (QPRT, c16orf54. PRRT2, CDITP, C16ORF53, SEZ6L2, ASPHID1, KCTD13, LOC124446, HIRIP3, FLJ90652, DOC2A, FAM57B, ALDOA, TBX6, MAPK3) [2, 30], and several transcription factors, TBX6, TAOK2 and MAZ. Mitogen activating protein kinase 3 (MAPK3) also influences the transcription of other genes through the MAPK/ERK pathway. [2]. (See supplementary table 1.1 for list of 16p11.2 genes and gene names)

Based on function, some genes in the 16p11.2 variation may be good candidate genes. T-box 6 (TBX6), one of the transcription factors, is a key developmental gene responsible for mesoderm formation [60]. Homozygous Tbx6-/- mice have rib fusions, rib anomalies, asymmetrical pairing of ribs, with varying penetrance [61, 62]. This is similar to what is seen in a 16p11.2 deletion carrier who has rib and vertebral anomalies [59]. Mutations in genes involved in metabolic pathways have been linked to mental retardation. Two genes in the genomic interval involved in metabolism include Fructose-bisphosphate aldolase A (ALDOA), and Aspartate beta-hydroxylase domain containing 1 (ASPHID1). Mutations in ALDOA have been associated with human disorders involving weekness. [63, 64]. Genes which function in the brain include Double C2-like domains alpha (DOC2A), which is involved in calcium dependent neurotransmitter release [65]. Doc2a-/- mice have altered synaptic transmission

and behavioral defects [66]. HIRA interacting domain (HIRIP3) plays a role in chromatin remodeling by interacting with HIRA a gene found in the 22q region associated with DiGeorge/velocardiofacial syndrome (DGS/VCFS)

Some studies have attempted to sequence candidate genes including ALDOA, TBX6, Sialophorin (SPN), DOC2A, Seizure related 6 homolog (mouse)-like isoform SEZ6L2, and Quinolate phosphoribosyltransferase (QPRT). These studies did not succeeded in finding sequence alterations within the coding regions [2, 35, 67]. Functional, molecular, bioinformatic, and expression studies can be conducted on genes in the 16p11.2 region to elucidate the genes role in the pathogenesis of the neurological disorders associated with the 16p11.2 mutation.

The list of 27 candidate genes can be narrowed down through cases with smaller mutations of the variation. In unpublished research presented at the World Congress of Psychiatric Genetics (Nov 2009) a group from UZ Leuven (Belgium) reported a case with a smaller mutation within the 16p11.2 region. The smaller variation encompasses only 5 genes within the common 600kb 16p11.2 variation (MVP, SEZ6L2, KCTD13, CDIPT, and ASPHID1). Two of these genes play a role in the brain: SEZ6L2, and Major Vault Protein (MVP). SEZ6L2 is expressed in various regions of the brain, including the cortex, hippocampus, and amygdala, and thalamus [68]. The child's maternal uncle and maternal grandmother who also have the smaller deletion have autistic tendencies but no formal autism diagnosis. The mother, from whom the proband inherited the deletion does not have either traits and is normal. It remains unclear as to whether the smaller deletion contributes to autism [69].

| Genes found in the deletion | Gene Name | In same network * | Expressed in brain ** | Function - Gene Atlas |
|---|---|---|---|---|
| ALDOA | Fructose-bisphosphate aldolase A | YES | YES | reversibly cleaving FI-P and FBP to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate |
| ASPHD1 | Aspartate beta-hydroxylase domain containing 1 | | YES | aspartate beta-hydroxylase domain containing 1 A transcript abundant in the brain, catalyses oxidative reactions in a range of metabolic processes |
| BOLA2 | | | | BOLA-like protein |
| C16orf53 | open reading frame | | YES | Histone methyl transferase complex |
| C16orf54 | open reading frame | | Yes | hypothetical protein |
| CDIPT | CDP-diacylglycerol-inositol | | YES | Enzyme, phosphatidylinositol synthase and phosphatidylinositol inositol exchange activity, implicated in a process cellular process physiological cellular trafficking transport, phosphatidylinositol transporter, lipid/lipoprotein, signaling, phospholipid biosynthesis |
| CORO1A | Coronin actin binding 1A | | | |
| DOC2A | Double C2-like domains alpha | | YES | may have regulatory role in membrane interactions during trafficking of synaptic vesicles at the active zone of the synapse C2-like domains that interact with Ca(2+) and phospholipid |
| FAM57B | hypothetical protain LOC83 | | YES | family with sequence similarity 57, member B, transmembrane protein of the cerebellum |
| FLJ90652 | | | YES | coiled-coil domain containing 95 |
| GDPD3 | Glycerophosphodiester phosphodiesterase domain 3 | | | glycerophosphodiester phosphodiesterase activity |
| GIYD2 | GIY-YIG domain-containing protein | | | Role in DNA repair and recombination |
| HIRIP3 | HIRA interacting domain | | YES | role in chromatin assembly, controls gene transcription. |
| KCTD13 | potassium channel tetramerisation domain | YES | YES | voltage-gated potassium channel, protein binding activities, |
| KIF22 | Kinesin family member 22 | YES | | involved in spindle formation and the movements of chromosome during mitosis and meiosis. The kinesin superfamily proteins transport membranous organelles and protein complexes in a microtubule- and ATP-dependent manner |
| LOC124446 | | | YES | |
| MAPK3 | Mitogen activated protein kinase 3 isoform 1 | | YES | MAPK/ERK pathway |
| MAZ | MYC associated zinc finger protein isoform 2 | | | transcription factor, antioncogene, growth suppressor in non transformed cells by affecting the levels of key cell cycle regulatory proteins may function as a transcription factor with dual roles in transcription initiation and termination |
| MVP | Major Vault Protain | YES | | structural protein, associates with RNA and cytoplasmic ribonucleoprotein complexes, Functions in secretory and excretory functions |

| | | | | |
|---|---|---|---|---|
| PPP4C | Serine/threonine-protein phosphatase 4 catalytic subunit | | | serine/threonine protein phosphatase 4C,catalytic subunit, involved in microtubule organization, activator of cRel/NF kappa B, stimulating the DNA-binding activity of cRel and activated NF kappa B-mediated transcription (overexpression of PPX) |
| PRRT2 | proline-rich transmembrane protain 2 | | YES | |
| QPRT | Quinolate phosphoribosyltransferase | | YES | quinolinate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (carboxylating)) |
| SEZ6L2 | Seizure realted 6 homologue (mouse) - like isoform | YES | YES | |
| SPN | Sialophorin | YES | | important for immune function, plays a role in the physicochemical properties of the T-cell surface, lymphocyte adhesion and activation through lectin |
| SULT1A3 | Sulfotransferease (SULt) 1A | | | catalyzing the sulfate conjugation of phenolic monoamines (neurotransmitters such as dopamin, serotonin) phenolic and catechol drugs |
| TAOK2 | TAO kinase 2 isoform 1 | | YES | TAO kinase 2 (Serine/threonine-protein kinase TAO2) belongs to the MAP kinase family. It activates the JNK MAP kinase pathway through the specific activation of the MAP2Ks MEK3 and MEK6. |
| TBX6 | T box 6 | | YES | transcription factor, involved in paraxial mesoderm formation and somitogenesis |
| YPEL3 | Yippie-like 3 | | | playing an important roles in the cell division and in the maintenance of life, highly homologous genes are found thought life including plants and fungi. It is associated with the mitotic spindle. |

**Table 1.3 Genes found in the 16p11.2 region.**

**Which 16p11.2 gene is contributing to disease pathogenesis?**

The 16p11.2 mutation is associated with an increased risk for autism spectrum disorders, schizophrenia and developmental delay. There are many genes within the region which can be theorized to contribute to the pathogenesis of the neuropsychiatric phenotype associated with the region. Since MAPK3 (ERK1) has such an important role in cellular regulation, and mutations in other genes in the ERK/MAPK pathway result in developmental syndromes, MAPK3 seems like a likely candidate in contributing to the pathogenesis of the psychiatric diseases associated with the variation. MAPK3 is involved in the RAS/MAPK pathway which has previously been associated with developmental syndromes, known as the RASopathies [70]. The MAPK pathway is involved in cell differentiation, survival, proliferation, learning and memory [71, 72]. MAPK3 forms a complex with MAPK1 (ERK2) which acts as a kinase in the ERK/MAPK pathway, and plays a role in controlling cell regulation including those events involved in early neural development [71, 72]. Blocking the ERK1/2 complex in mouse, and chick ES cells, has been show to inhibit the expression of key neural development genes [73]. Mapk3 null mice are hyperactive and have reduced long term potentiation [74].

The RASopathies are developmental syndromes due to germline mutations, of genes in the ERK/MAPK pathway [70]. Many of the symptoms in this class of syndromes include craniofacial dysmorphology, cardiac abnormalities, skin lesions, muscular skeletal abnormalities, and cognitive and learning defects and cancer due to up regulation of RAS [70]. Common

syndromes in the RASopathies include Neuro-cardio-facial-cutanious syndromes (NCFC) which include Noonan, LEOPARD, and Costello. NCFC is associated with mutations affecting the activation of ERK. These patients exhibit a high rate of cognitive deficits and psychiatric disease. Patients with a 1MB deletion in 22q which includes MAPK1 have neurodevelopmental deficits, and macrocephaly.

Determining which gene or genes in the 16p11.2 mutation are contributing to the neuropsychiatric phenotype would enable further understanding of the pathogenesis of the psychiatric diseases associated with the region. Expression studies in particular can reveal dysregulated genes resulting in disrupted molecular pathways. This may allow us to determine a gene on the 16p11.2 region which is playing a role in disease pathogenesis. In our study we will use genome-wide expression analysis from lymphoblast cell lines to determine dysregulated genes in patients as compared to controls.

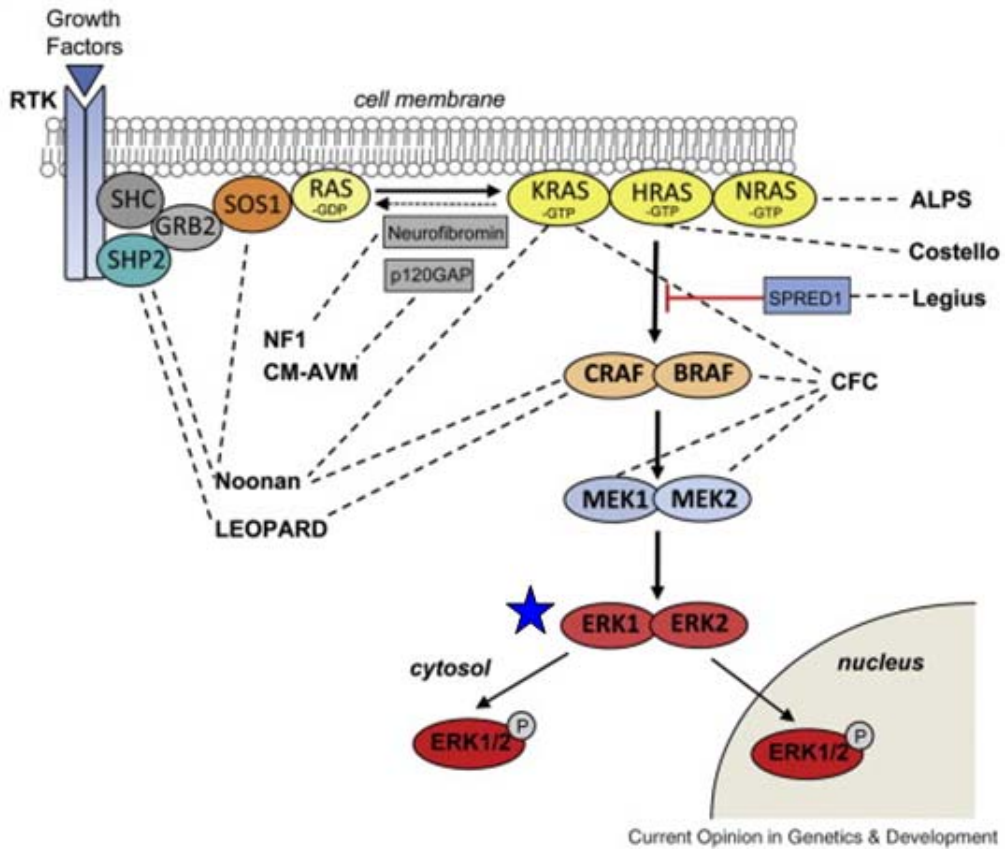



**Figure 1.3. Genes in the MAPK pathway are associated with developmental disorders.**
Various components of the RAS/Mitogen activated protein kinase (MAPK) pathway have been associated with developmental disorders. Disorders such as ALPS, Costello, Legius, CFC, CM-AVM, NF1, Noonan, and Leopard all have been associated with one of more molecules involved in the pathway. A small percentage of NF1 patients have been diagnosed with autism. ER1 (MAPK3) is a gene found in the 16p11.2 interval is labeled with a blue star in the diagram.

**From:** Tidyman *et al*. *The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation* Current Opinion in Genetics & DevelopmentVolume 19, Issue 3, June 2009, Pages 230-236
Genetics of disease

## Chapter 2.

### Using a high resolution microarray to discern the breakpoints of the 16p11.2 microdeletion and microduplication in 10 cases.

Mary Kusenda[1,2], Patricia Roccanova[1], Ray Kim[1], Shane McCarthy[1].

The breakpoint analysis in this chapter was performed by Mary Kusenda. The HD2 microarray (Roche Nimblegen Inc) hybridization, and data processing was performed by Patricia Roccanova at Cold Spring Harbor Laboratory. Ray Kim helped in data processing. Shane McCarthy provided the Splus script used to plot chromosome 16.

[1]Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[2]Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, NY 11724

**Summary:** The breakpoints for the 16p11.2 microduplication and microdeletion may vary between individuals. Using a high resolution microarray technology NimbleGen HD2 (High Density 2 million probe array) we can determine whether the recombination breakpoints differ among 10 screened patients. Fusion genes, truncated proteins, and identification of smaller deletion regions within the variation can all contribute to variation in phenotype. This microarray technology enables us to determine patients which have different breakpoints, and whether this corresponds to differences in phenotype. We have used HD2 arrays to determine the breakpoints of 5 autistic, and 5 schizophrenic 16p11.2 carriers. Breakpoints between individuals did not vary, but we were able to determine a sample which did not have the 16p11.2 mutation as previously thought. This sample will no longer be used in our analysis.

**Introduction:**

Studies using array comparative hybridization (aCGH) microarray technology and cytogenetic techniques have attempted to determine candidate genomic regions which contribute to disease etiology. aCGH uses a test and a reference sample which is co-hybridized onto a microarray platform [31]. The sample DNA hybridizes to its complementary sequence on the array [31, 75]. Data is then processed to determine deletions and duplications found in the test sample [31]. Finding the correct breakpoints flanking a genomic variation is crucial in determining an accurate set of candidate genes associated with the phenotype.

The early publications regarding the 16p11.2 variant produced different breakpoint coordinates based on study, and microarray method. Initial reports ranged from chr16: 29,578,715, 29,557,173 and 29.5 for the left boundary, and 30,081,289, 30,253,852 and 30.1 for the right boundary. (Sebat *et al*, Weiss *et al*, ad Kumar) [2, 5, 11]. One reason for the great discrepancy between breakpoints of the 16p11.2 mutation is each study used different microarray platforms. Additionally microarray platforms have poor probe coverage in regions of sequence similarity or Low Copy Repeats (LCRs) by which the region is flanked [2, 76]. New high resolution microarray technology can be used to determine the breakpoints of a CNV. Roche NimbleGen HD2 array CGH platform has 2.1 million probes at a ~5kb resolution, at a median probe spacing of 1.1kb genome wide. In addition to this superior resolution, the technology has enhanced probe selection in regions of LCRs. These improvements in array technology have enabled us to determine breakpoints within the CNV region, and will allows us to determine shared genes among patients, or fusion genes which may have occurred as a result of recombination.

Finding the recombination breakpoint among 16p11.2 microdeletion, and microduplication patients is an important task. These three genes in the LCRs include Bola like protein member 2 (BOLA2), GIY-YIG domain-containing protein (GIYD1/2), and sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3 (SULT1A3/4). Exons overlap between GIYD1/2, and SULT1A3/4. By finding the recombination breakpoints we are able to better understand whether all patients share the same number of genes within the variation. Additionally it

would be interesting to see whether genes within the segmental duplications were known to be dosage sensitive, and whether these changes in dosage resulted in any functional consequences. Finding truncated and fusion genes in the segmental duplications flanking the 16p11.2 variation would be interesting to our results.

Finding the breakpoints of the 16p11.2 mutation can result in narrowing down the list of candidate genes. A patient with neurological deficits associated with the 16p11.2 mutation and a smaller variation containing only a subset of genes would greatly help in narrowing down the list of candidate genes within the variation. Determining the precise genomic breakpoints in 16p11.2 carriers would enable us to discover patients who have truncated or fusion genes, or cases which have varying number of genes in the variation. By comparing the patients phenotype of those which have a particular breakpoint variation would enable us to make correlations between certain genomic breakpoints and the severity of the phenotype.

**Materials and Methods:**

**Collection of DNA samples.** We have collected five DNA samples from individuals diagnosed with autism and the 16p11.2 mutation, provided from the Autism Genetic Resource Exchange (AGRE) and five DNA samples from individuals diagnosed with schizophrenia and the 16p11.2 mutation from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE)

**Hybridization of sample DNA on NimbleGen HD2 array.** DNA samples were labeled with fluorescent dyes. Test samples were labeled with Cy3, and reference samples with Cy5. These were pipetted into PCR strip tubes (Ambion). Samples were heat denatured in a thermocycler at 98C for 10 min, followed by chill on ice for 2 min. dNTP/Kenlow is added to samples. Samples are incubated for 2 hours at 37C. The reaction is stopped by adding 0.5M EDTA. 5M NaCl is then added. Samples are incubated for 10min at room temperature. The samples are centrifuged, and rinsed with 80% ice cold ethanol, followed by centrifugation. DNA is dried using a SpeedVac, and then hybridized onto NimbleGen HD2 array using the protocol given. Sample mixers are provided by NimbleGen. Test and reference samples hybridize on to the array for 60-72 hours. Arrays were scanned on the Axon 4000A scanner (Molecular Devices Inc) Data was gathered using GenePix software.

**Data processing of Nimblegen HD2 data.** Color intensity data was normalized using spatial normalization to correct for intensity variations on the surface of the array, using an R-module (Roche-Nimblegen Inc). This step is an image processing step. Data from Cy5, and Cy3 intensities, were corrected to a fitting curve, using intensity data from autosomal probes with the least variability between test and reference samples. Since the hybridization method involves a dye swap, this normalization is repeated, by exchanging the test and reference sample. Using geometric mean of the normalized and raw intensity data the log2 ratio is then found using the test and reference.

**Determining breakpoint of the 16p11.2 deletion:** An S+ script provided by Shane McCarthy (Cold Spring Harbor Laboratory) was used to plot the intensities on chromosome 16 capturing probes which correspond to the genomic

coordinates within and surrounding the variation. Zoom features and scatter plots in S-Plus were then employed to determine the breakpoints. All breakpoints were recorded for each sample.

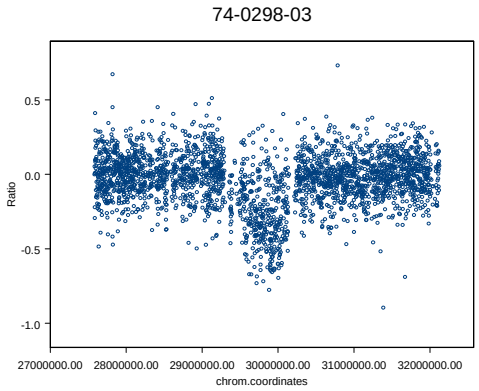
**Results:**

We were not able to discover any significant variations in the breakpoints between samples. We confirmed the two duplications, and three deletions which were found in our autism samples and the one deletion and three duplications in the schizophrenia samples. One sample did not confirm to have the deletion or duplication. We were also interested in determining whether there was a common breakpoint unique to microdeletions, as opposed to the microduplication. This again was not the case as all breakpoints were similar through out the samples.

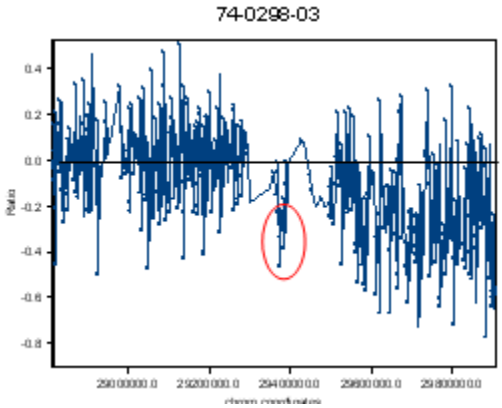
**Figure Description for figure 2.1**

The first scatter plot in figure 2.1 has the wide view of the 16p11.2 variant in each patient as visualized by Splus from the Nimblegen HD2 array. The following image for the sample shows the left followed by the right breakpoint. For all graphs the x axis is the genomic coordinate on chromosome 16. The y axis is the intensity ratio. The very first probe which is believe to be part of the variant, meaning that the intensity ratio is falls suddenly lower in the deletion, and suddenly higher in the duplication is circled. The breakpoint estimate is determined by reading the genomic coordinate where the probe falls in the x axis of the graph. Each breakpoint estimate, for the left and right breakpoints is circled in red. Data for each sample includes identifiers such as Cell Identification number (Cell ID), individuals identification number (ID#), and experiment identification number (Experiment name). It is also reveled whether the variation in each sample is a deletion or duplication, and whether or not if it is *de novo* or inherited (if known). This was plotted for all 10 samples analyzed using the Nimblegen HD2. Figure 2.1 is an example of these plots.

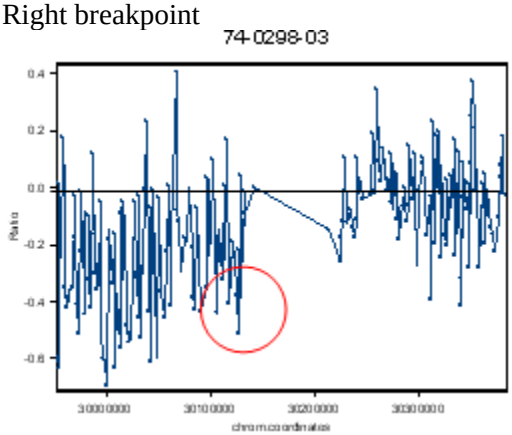
**Figure 2.1:**
ID# 74-0298-03
Cell ID 03C16872
Subject has de novo deletion.
Breakpoint coordinate estimate: Chr16: 29373319-30126984
Experiment name is PR2003
Chromosome 16p11.2:




Left breakpoint




Right breakpoint



| Sample | Left Breakpoint | Right Breakpoint | CNV |
|---|---|---|---|
| 1 | Chr16: 29502742 | Chr16: 3010634 | Duplication |
| 2 | Chr16: 29373319 | Chr16: 30126984 | Deletion |
| 3 | Chr16: 29372219 | Chr16:30245792 | Deletion |
| 4 | Chr16: 29496614 | Chr16: 30241392 | Duplication |
| 5 | Chr16: 29368919 | Chr16: 30241392 | Deletion |
| 6 | Chr16:23975619 | Chr16: 30213836 | Deletion |
| 7 | No deletion or duplication found | No deletion or duplication found | |
| 8 | Chr16:29425000 | Chr16: 30143384 | Duplication |
| 9 | Chr16:29293203 | Chr16: 30322000 | Duplication |
| 10 | Chr16: 29356806 | Chr16: 30143384 | Duplication |

**Table 2.1. List of left and right breakpoints for the 16p11.2 mutation in 10 cases.**
This table shows the breakpoint estimate for each sample, and whether the sample had a microdeletion or microduplication at the 16p11.2 locus. The first 5 samples are Autism samples, the next 5 are from Schizophrenia samples.

**Discussion:**

We used the NimbleGen HD2 microarray to determine the breakpoints of the 16p11.2 variation. The ten samples which we analyzed did not vary much in their breakpoints at the 16p11.2 locus. One sample which had initially been determined to have the 16p11.2 variation was now determined it does not have the variation, and will no longer be included in our study.

The twenty five genes which were within the variation were common to all samples. We were not able to find samples which only contained a subset of genes from the twenty five in the variation. It would have been of great advantage to find such a sample through our study. One drawback from our study was the limited number of samples which were scanned with the HD2 arrays.

Although we were not able to determine whether the breakpoints differed greatly between individuals, the fact that breakpoints between individuals were within the 147kb segmental duplication gives us confidence that most patients with the 16p11.2 microdeletion and microduplication have a mutation which encompasses the same 27 genes associated with the CNV. This is a great advantage when looking at genome-wide expression data of known 16p11.2 carriers. We would expect that the 16p11.2 deletion or duplication would produce the same dysregulated genes downstream in their respective molecular pathway.

## Chapter 3.

## The effect of chr16p11.2 microdeletions and microduplications on gene expression in Autism Spectrum Disorders and Schizophrenia

Mary Kusenda[1,2], Vladmir Vacic[1], Seungtai Yoon[1], Joseph Sarrro[1], Chris Johns[1], Sohail Khan[1].

Culture of lymphoblast cell lines, RNA extraction, genome wide expression data processing, fold change analysis, and analysis of gene function was performed by Mary Kusenda. RNA hybridization, array data processing was performed by Joseph Sarro, Chris Johns, and Sohail Khan at the Cold Spring Harbor Laboratory Microarray Core Facility. Vladimir Vacic coded all scripts for Significant Analysis of Microarrays (SAM), and modified SAM using the R computer language. Vladimir also helped with data analysis, and experimental design. Seungtai Yoon, gave advice on use of statistical methods. Most cell lines were acquired from Rutgers University Cell and DNA Repository AGRE sample repository. All other cell lines were shipped live from Ravi Kumar, (University of Chicago) Judy Rappaport (NIMH), and Hilde Peters (UZ Leuven).

[1]Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[2] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, NY 11724

**Summary:** The 16p11.2 microdeletion is associated with an increased risk for Autism Spectrum Disorders, and developmental delay, while the reciprocal duplication is associated with schizophrenia, bipolar disorder, autism and developmental delay. [1-5]. We hypothesized that there are dosage sensitive genes within the variation that are contributing to the pathogenesis of psychiatric disease. We generated genome wide expression data using lymphoblast cell lines from 16p11.2 carriers and controls to determine a list of dysregulated genes which may be contributing to the psychiatric phenotype seen in 16p11.2 patients. We determined seven genes outside of the 16p11.2 region which expression correlates with copy number at the variation. (URB1, PTGS1, BMP7, C6orf35, ARL17, c19orf36, SIAE). One of these genes, BMP7 has been shown to be induced through the BDNF/MAPK/ERK pathway by direct activation through MAPK3/1 (ERK1/2) during brain development. MAPK3 is a gene on the 16p11.2 region. Inhibition of ERK1/2's induction of BMP7 results in abnormal neuronal migration, reported through in-vivo, and in-vitro studies [77]. The dysregulation of BMP7 may be caused by MAPK3, which may be contributing to the increased risk of neurological diseases associated with the 16p11.2 mutation.

**Introduction:**

One of most recent rearrangement hotspots associated with autism and schizophrenia is on 16p11.2, which contains 27 genes [2, 3, 5, 30, 34, 35, 59, 78-80]. The overwhelming rate that microdeletion and the microduplication carriers will exhibit some psychiatric or cognitive disorder makes the 16p11.2 region a good candidate CNV to study gene expression [2, 3, 68]. The genomic region is flanked by segmental duplications, increasing the probability that individuals share similar breakpoints therefore gene expression analysis across individuals would study the same genes [2, 5, 30]. Although there is no candidate gene within the variation that has been implicated in psychiatric disorders, many of the genes are expressed in the brain, and also including some transcription factors (TBX6, MAZ) [2, 68]. We hypothesized that the microdeletion and microduplication would affect the expression of dosage sensitive genes within the variation. Changing the dosage of certain regulatory genes could alter subsequent genes downstream in the respective molecular pathway, or dosage effects. We hypothesize that together this common set of dysregulated genes will pinpoint a possible developmental pathway that is affected, which may guide us towards understanding the pathological manifestation of the phenotype associated with the 16p11.2 CNV.

Gene expression profiles reveal complex biological processes found in vivo, and contribute to the discovery of indicators of disease [81-86]. Possibly the best example is clinical cancer research, where the gene expression patterns found in tumor subtypes have identified molecular markers, and subsequently paved the way for new cancer targeting therapies [85, 86].

The brain undergoes many gene expression changes due to external and internal stimuli, and neuronal activity, which influences several transcription factors [87]. Previous gene expression studies have compared the expression

profiles between cases and controls in the hopes of elucidating out a list of dysregulated candidate genes [84, 86, 88, 89]. Expression studies have to be done with some caution. Differential expression may be due to a functional variant, cell line/tissue type artifact, changes upstream or downstream to the gene, or by statistical chance [90]. The experimental design is important for getting accurate results, and with the advances in psychiatric genetic research gene expression studies on autism and schizophrenia have become more developed. Previous studies did not consider autism and schizophrenia as heterogeneous disorders. Studies in the past have used monozygotic twins discordant for autism, and determined out numerous discordant genes between the two twin pairs [84]. Others have used age matched gender matched sibling pairs in their studies, without looking first into the genetic makeup of the patients and even unknowlingly used a family with the 16p11.2 mutation [89]. Recently researchers have made a more focused attempt in searching for discordant genes, by focusing on dysregulated genes as a result of the same genotype [91]. In Nishimura *et al*, two candidate genes (JAKMIP, GPR15) were discovered by investigating dysregulated genes in patients with fragile X syndrome, and the maternally inherited duplication of 15q11-13, both regions previously associated with autism [91]. Our approach follows this method as all our cases have either the 16p11.2 microdeletion or microduplication. Using the genome-wide expression data from these carriers will help discover dysregulated *dosage* and *trans* effects.

In our experimental design we utilized lymphoblast cell lines as our tissue of choice, as brain samples from 16p11.2 patients are unavailable. Obtaining blood from patients is a simple, relatively non-invasive, and useful alternative when brain banks do not exist [89]. Studies done on post mortem brain tissues on schizophrenic individuals have revealed dysregulated genes. Although it is ideal to use brain tissue when studying a neurological disorder, some caution should be used. RNA from post-mortem brain may not be high quality, due to delayed time to RNA extraction, or traumatic death. Alternativly mRNA from blood has been used in the study of gene expression changes resulting from neurological injury (stroke, seizure, hypoglycemia), hereditary neurological disorders (Huntington's disease), and neurological disease (Autism, Schizophrenia) [87, 89]. A study on the neurological disorder Huntington's disease showed that genes expressed in peripheral blood, were similarly expressed in post mortem tissue. Lymphoblast cell lines, the immortalized derivative of peripheral blood have been used to generate candidate genes in studies in a number of neurological diseases and disorders including of Rett Syndrome, Autism, Autism due to Fragile X, 15q11-q13 dup, schizophrenia, and bipolar disorder [91]. There are certain drawbacks when using lymphoblast cell lines. Immortalization by Epstein Bar virus may affect gene expression, although all samples used were immortalized which would negate any effect of gene expression. However differences in gene expression are caused by differences in the immortalization process. All samples in our study were not from the same source and may not have been immortalized using the same method. Therefore, although lymphoblast cell lines are not the ideal tissue to used in the study of a neurological disorders we hope it can be a useful alternative in our study, as brain tissue is unavailable [89].

Genome wide expression profiles have been collected on 21 carriers of chr16p11.2 microdeletion or microduplication carriers and 14 controls with the intention of defining a common set of dosage sensitive genes associated with the phenotype. Gene expression profiles will be able to reveal dosage effects but more importantly *trans* effects or dysregulated genes outside of the CNV. In our experimental design we have cultured Epstein Barr Virus (EBV) immortalized lymphoblast cell lines (LCLs) of patients which have the 16p11.2 deletion or duplication along with healthy controls. We have extracted RNA from these lines in biological triplicates and hybridized it onto the Affymetrix U133 Plus 2.0 Array. This particular array was chosen for its 54,000 probe sets, which capture expression of up to 38,000 characterized genes and 47,000 transcript variants. After normalization and data processing using Gene-Spring GX software (Agilent Technologies) we used a Significant Analysis of Microarrays (SAM) a microarray analysis tool, to analyze the expression data in order to determine genes which expression correlates with dosage.[92]

We employed SAM to discover a set of statistically significant dysregulated genes. In order to analyze our genome wide expression data we used 'samr' package for R (version 1.26) [92]. The SAM algorithm is non-parametric, and empirically estimates the null distribution of gene association scores by randomly permuting genotype labels. We performed SAM quantitative analysis using copy number (1, 2, and 3) as a response variable. SAM prioritizes significant genes using a SAM score which corresponds to the slope of the regression line by a variance penalty terms. Higher SAM scores correspond to genes which expression values deviate from expected, in this case expression which responds to dosage at the 16p11.2 variation.

Patients with the 16p11.2 CNV are limited and those which have available lymphoblast cell lines are even fewer. In order to utilize each sample we modified SAM to take into account the unequal number of samples in each copy number category in addition biases. Lymphoblast cell lines from cases which had the microdeletion were the limiting factor in our analysis. The liner regression and permutations of the SAM method is sensitive to an unbalanced number of samples in copy number category which results in an inflation of SAM scores. We had several other biases in our samples due to age ethnicity and gender. Our experimental design had many experimental samples from younger children with autism, and control samples from adults; the mother and fathers of the children. Gene expression profiles from whole blood show a negative correlation between Iq expression and age, as a result of decreases IgM, and IgG serum levels with age [85, 87]. The majority of our samples were caucasian, with one family which was Pacific Islander. In order to mollify these biases and limitations we modified SAM so the program would randomly choose 18 samples (6 deletions, 6 controls, and 6 duplications) in each run. (See Table 3.1) Additionally each copy number category was gender balanced (4 males, and 2 females). We ran 6000 iterations using SAM. Our final SAM scores is the average, of the 6000 runs. Through random sampling of our samples our approach placates the biases which occur in our data set and captures the microarray probes which expression correlates positively or negatively with copy number. Probes corresponding to genes within

the 16p11.2 region should be expected to have the highest SAM score and show the highest departure from background distribution.

**Materials and Methods:**

**Lymphoblast cell line sample collection.** We collected the Epstein Barr Virus (EBV) immortalized Lymphocytes cell lines from cases containing a CNV at the 16p11.2 locus and an Autism or Schizophrenia diagnosis, and control lymphocyte cell lines which do not have the mutation or a known psychiatric disorder. Cryopreserved lymphocyte cell lines were acquired from Rutgers University Cell and DNA Repository AGRE sample repository. All other samples were shipped live from Ravi Kumar, (University of Chicago) Judy Rappaport (NIMH), and Hilde Peters (UZ Leuven). (See Table 3.1 for Sample Summary) We were limited by the number of samples we could find which had the microdeletion.

**Maintenance of Lymphoblast cell lines.** All Lymphocyte cell lines are cultured in RPMI1640 Medium (Gibco) supplemented with 15% Fetal Bovine Serum (Atlas Biological), 1% Penicillin/Streptomycin (Gibco), 1% MEM Non-Essential Amino Acid solution (Gibco), and 1% L-Glutamax (Gibco). Cells are then placed into 37C incubator at 5% C02. Complete Media is doubled every 2-3 days until cells are approximately 2.5x10^6 per ml, and in log phase. Cells are counted by hemocytometer, and are seeded as biological replicates in triplicate at a 5x10^6 cells per flask, and 10mls Media.

**RNA Extraction and Hybridization to Expression Array.** Cells are pelleted 24 hours post-seeding and homogenized using Qiagen's Qiashredder. RNA extraction is performed using the RNeasy Mini kit (Qiagen, Valencia, CA) following the protocol for RNA extraction of animal cells exactly. Concentration and preliminary quality control of RNA is performed by Nanodrop ND-1000 spectrophotometer (Nanodrop technologies, Willmington DE). Each RNA sample was submitted to the Microarray Shared Resource at Cold Spring Harbor Laboratories. Quality control was checked by Agilent 2100 bioanalyzer (Agilent technologies, Palo Alto, CA). ~2.0ug of RNA from each sample was amplified and biotin labeled and converted into cDNA and hybridized onto the Affymetrix Human Genome U133 Plus 2.0 Array (Affymetrix Gene Chip technologies, Santa Clara, CA). This particular array was chosen for its 54,000 probe sets, allowing analysis of up to 38,000 characterized genes and 47,000 transcript variants.

**Analysis of the Expression Microarray.** After arrays are scanned by Axon 4000A scanner, (Molecular Devices Inc) data is converted into Cel Files. Gene Spring GX software (Agilent Technologies) was used for processing raw data. Robust Medioid Averaging (RMA) was performed in order to take the median intensity value of probe set. Data was then transformed to log base 2, and the average between the 3 replicates is computed for each sample. Differential analysis was then completed to give the fold change of expression as compared to the experimental. All data was imported into SPlus (S plus Programming Language software; Insightful Corp). The fold change difference between cases and controls was determined. Fold change expression values were determined by taking the mean expression values across all samples for each copy number category (copy number 1, 2, 3) and dividing copy number 1 category expression

values by values for copy number 2 category for a given probe, and copy number 3 category expression values by values for copy number 2 category for a given probe.

**Statistical Analysis of Genome wide Expression Data.** Gene expression values which correlated with genotype were discovered using (Significant analysis of Microarrays; Tusher 2001). Copy number (1, 2, and 3) was used as an response variable using 'samr' package for R (version 1.26) We were limited by the number of samples in the microdeletion (1 copy number) category. In order to maximize statistical power and use our small sample set we modified SAM in order to account for sources of bias contributed by relatedness between samples, gender, ethnicity, age, and an unequal number of individuals in a copy number category. We ran 6000 iterations of SAM analysis each on random combinations of 18 out of the total number of 35 samples. Each combination had a balanced number of samples with corresponding copy number (6 deletions, 6 duplication, 6 normal). For copy number of 3 we further restrained the program to always pick 3 samples with autism and 3 samples with schizophrenia. The out put of the SAM score is the slope of the regression line, with a variance penalty term. Both linear regression and permutations (which SAM uses to estimate statistical significance of observed SAM score) are sensitive to the sizes of groups of samples with a specific genotype. SAM scores for each probe are the means of observed SAM scores from each iteration. Probes which mean expression intensity were lower than 6.0 were removed as a precaution that the significant positive correlation of expression with genotype may have been due to noise, or random data. A 9.88% False Discovery Rate (FDR) cut off value was used in the final analysis.

**Real-time PCR of lymphoblast cell line RNA.** Significant genes from the expression analysis which show positive correlation with copy number were confirmed using real-time PCR. RNA extracted from lymphoblast cell lines was converted to cDNA using the Taqman RT PCR reagents (Applied Biosystems). Each sample which was used in the genome wide expression analysis was used in the RT-PCR analysis. Every lymphoblast cell line was cultured in biological triplicates, and so three RNA samples exist for each sample. Each of these biological triplicates has three experimental triplicate RT-PCR reactions for a total of 9 RT-PCR reactions for each lymphoblast cell line. The average of these nine expression values was used in the final analysis. The real-time PCR primers were designed for BMP7, SIAE, PTGS1, and Actin-beta as a positive control. Primers for BMP7 (5' CAGAGCATCAACCCCAAGTT 3') and reverse (5' CAGGTCTCGGAAGCTGACAT 3'),
SIAE (5' ATCGAAGACTGGCGTGAAAC 3') and reverse (5' TCTTCTCAGGCAGTGGTCCT 3'), and for
PTGS1 ( 5' GGAGTTTGTCAATGCCACCT 3') and reverse (5' CTGGTGGGTGAAGTGTTGTG),
ACTIN-Beta, ( 5' TGTGATGGTGGGAATGGGTCAG 3') and reverse (5' TTTGATGTCACGCACGATTTCC 3'). 96-well optical plates were used for the Real-time PCR in triplicate for each gene tested. Real-time PCR reaction was carried out using 2X SYBR Green PCR master mix, 200 reactions (Applied Biosystems) in a 10ul volume. Real-time PCR was run using Real-time PCR ABI

Prism 7000 Sequence Detection System (Applied Biosystems). The Real-time PCR cycling parameters were as follows: 48°C for 30 min, 95°C for 10 min, 40 cycles of (95°C for 15 sec, and 60°C for 30 sec, and 72°C for 30 sec) followed by a disassociation step of (95°C for 15 sec, and 60°C for 15 sec, and 95°C for 15 sec.) Results were processed using ABI Prism 7000 SDS software (Applied Biosystems). At the end of each run, the baseline and Cr was adjusted to ideal linearity for each standard curve. As samples were across many RT-PCR plates final expression value was determine by using the equation (2^(expression of gene x – expression of endogenous control)) (Comparative Method) Quantitative values were determined using the standard curve.

**Gene Enrichment Profiler**: (Center for Computational and Integrative Biology.) This tool does not identify expression levels in tissues but determines tissues in which the expression of an Affymetrix probe is enriched. (http://xavierlab2.mgh.harvard.edu/EnrichmentProfiler/)

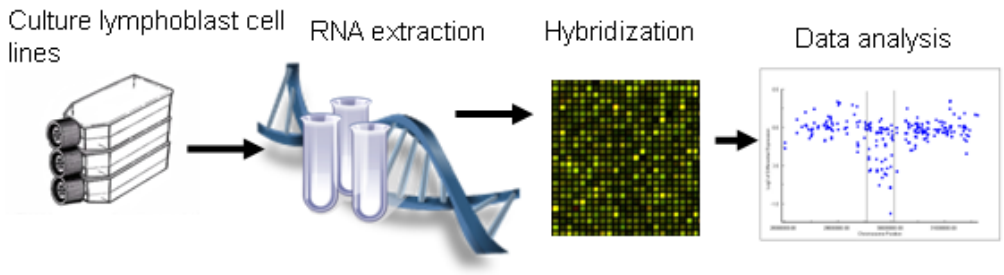



**Figure 3.1 Diagram of method used for genome-wide expression experiment.** EBV immortalized lymphoblast cell lines from cases and controls were cultured in biological triplicates. RNA was extracted from the lymphoblast cell lines, followed by hybridization onto Affymetrix U133 2.0 Plus, an expression array. Genome-wide expression data is processed and analyzed.

Original Data Set

| | Autism | Schizophrenia | Unaffected | Total |
|---|---|---|---|---|
| 1 copy | 5 | 1 | 0 | 6 |
| 2 copies | | | 14 | 14 |
| 3 copies | 6 | 9 | | 15 |
| | | | | 35 |

**Table 3.1 Number of samples in expression study**
Table 3.1 includes the number of individuals for each copy number category, (1 copy as the microdeletion, 2 copy as the controls, 3 copy as the microduplication) and their respective diagnosis. 35 samples were used in the SAM analysis. We were limited by the number of samples with a microdeletion.

**RESULTS:**

**SAM analysis of genome wide expression data.**

The SAM statistical software was used to determine genes whose expression correlates with genotype at the 16p11.2 locus, and shows the highest departure from background distribution. When the observed mean SAM score and the expected SAM score was plotted in a Quantile distribution, all genes which deviated from the linear regression line were positively correlated with genotype, meaning that gene expression increased with increasing gene dosage at the 16p11.2 locus. (See Q-Q plot, Figure 3.4) The first twenty four significant probes comprised of dosage effects, or genes found in the 16p11.2 locus. Expression values plotted versus copy number also showed that expression level correlates with dosage. At a 10% false discovery rate (FDR) (SAM g<0.1) we called thirteen significant probes which were captured outside of the 16p11.2 interval. Stringent criteria were applied to these thirteen genes, to remove probes which sequence had sequence similarity to genes in the 16p11.2 region, and probes with expression level lower than 5.0. This removed probes which significant correlation with genotype may be explained by noise from the data.

Our final list included seven significantly affected *trans* effects, which are URB1, ribosome biogenesis homologue, a gene which functions in ribosome biogenesis, and is ubiquitously expressed in tissues; PTGS1 prostaglandin-endoperoxide synthase 1, a key enzyme in prostaglandin biosynthesis, and a major target of non-steroidal anti-inflammatory drugs; C6orf35, a chromosome 6 open reading frame, BMP7, bone morphogenic protein 7, a gene in the TGF-beta super family, and a transcription factor in early development; ARL17, ADP-ribosylation factor like 17; SIAE, sialic acid acetylesterase; cDNA FLJ30565 fis, c19orf36, chromosome 19 open reading frame.

Real-time polymerase chain reaction (RT-PCR) was used to validate three of the seven significant *trans* effects including BMP7, SIAE, and PTGS1. The same RNA which was used for the genome wide expression analysis using lymphoblast cell lines was used for the RT-PCR analysis. For all three genes analized by RT-PCR expression value was lower in deletion cases as compared to duplication cases, which replicated with that of the genome-wide expression microarray.

We input all seven significant probes into a gene enrichment analyzer [93]. This tool does not identify expression levels in tissues but determines tissues in which the expression of an Affymetrix probe is enriched. All seven probes were used in the analysis but data only existed for BMP7, and PTGS1. Our significant *trans* effect BMP7 (probe number 211260_at) was enriched in the trigeminal ganglion, the cerebellum peduncles, and the superior cervical ganglion. Enrichment in regions of the brain was not seen in the other two probes for BMP7 (probe numbers 209590_at, 209591_s_at) which were not significant in our genome wide analysis. PTGS1 (215813_s_at) was enriched in smooth muscle, mast cell Ige, and mast cell.

A group in Cold Spring Harbor Laboratory led by Alea Mills created a 16p11.2 mouse model. They analyzed the genome wide expression profile from four different brain tissues including the cortex, cerebellum, olfactory bulb, and brain stem. BMP7, and SIAE two significant *trans* genes from our analysis, overlapped with the significant genes from their analysis. It is encouraging that there is some concordance between human lymphocyte and mouse brain expression data.

In this study we have used Lymphoblast cell lines from 16p11.2 carriers and healthy controls to determine seven significant *trans* effects. Three of three *trans* effects analyzed replicated using RT-PCR, which validates the microarray based data. Two of these genes were dysregulated in mouse brain tissue. Together the RT-PCR results and genome wide expression results from mouse tissue, gives us confidence that our genome wide expression analysis is capturing genes which are dysregulated in 16p11.2 carriers as compared to controls.

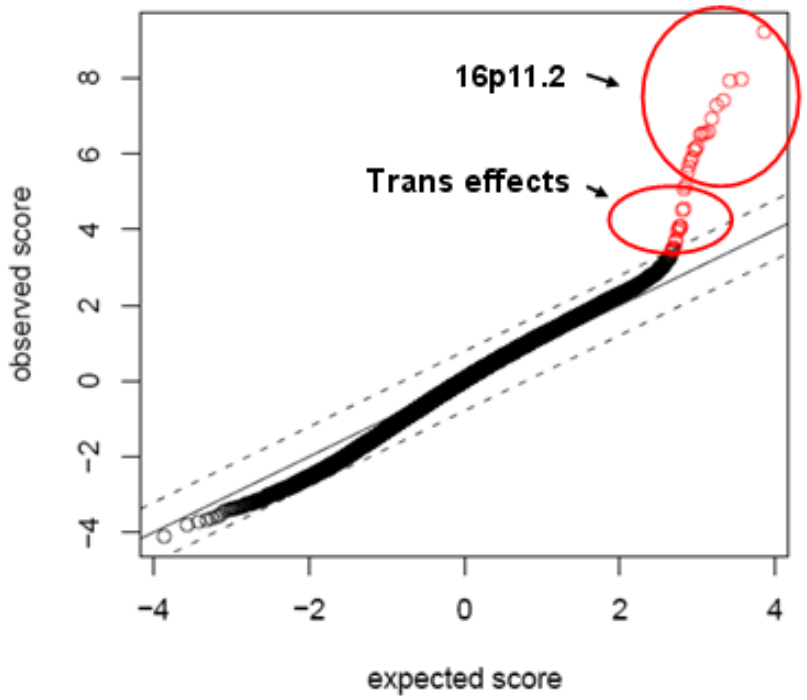
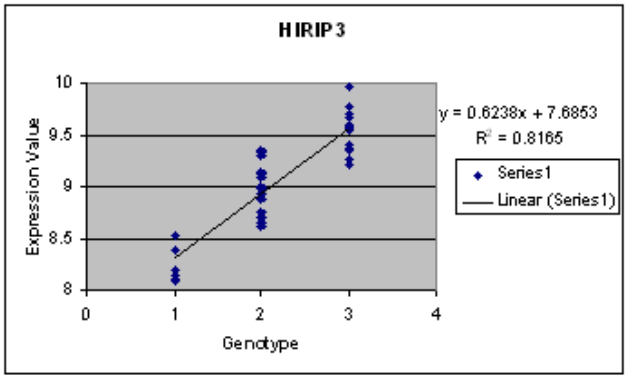


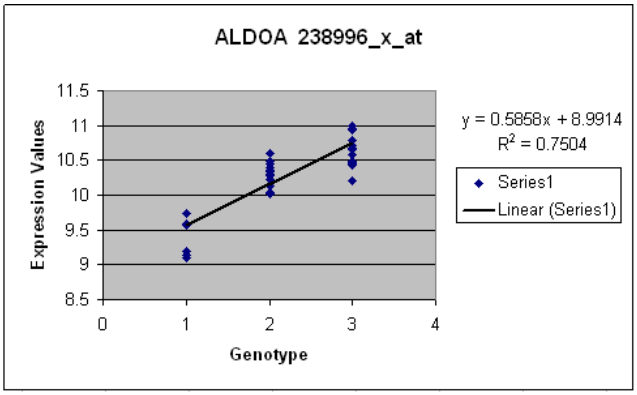

**Figure 3.2. Quantile-Quantile Plot depicting probes which expression correlates with genotype at 16p11.2**

Here is the Quantile-Quantile (Q-Q) Plot from our SAM analysis. This plots the observed SAM score/expected score. Each probe is represented by a small circle. There is an inflation of probes which are positively correlated with copy number, in the upper right corner of the plot. The probes in red are probes which had the highest SAM score and highest departure from background distribution (x=y line), and reached significance in our analysis. The dashed line corresponds to an empirical 9.88% False Discovery Rate (FDR) significance cut off. The first 24 genes are genes within the 16p11.2 region. The next 13 red probes are the 13 which made the FDR cut-off. Only 7 of the 13 significant *trans* effects made our stringent criteria to be part of our seven significant *trans* effects.

| | Affymetrix Probe | HUGO | Cytoband | FDR |
|---|---|---|---|---|
| 1 | 204504_s_at | HIRIP3 | chr16p11.2 | 0 |
| 2 | 202183_s_at | KIF22 | chr16p11.2 | 0 |
| 3 | 208932_at | PPP4C | chr16p12-p11 | 0 |
| 4 | 238996_x_at | ALDOA | chr16p11.2 | 0 |
| 5 | 201253_s_at | CDIPT | chr16p11.2 | 0 |
| 6 | 200966_x_at | ALDOA | chr16p11.2 | 0 |
| 7 | 214687_x_at | ALDOA | chr16p11.2 | 0 |
| 8 | 221889_at | KCTD13 | chr16p11.2 | 0 |
| 9 | 1568964_x_at | SPN | chr16p11.2 | 0 |
| 10 | 224981_at | TMEM219 | chr16p11.2 | 0 |
| 11 | 209836_x_at | BOLA2/B2 | chr16p11.2 | 0 |
| 12 | 231878_at | C16orf53 | chr16p11.2 | 0 |
| 13 | 227286_at | INO80E | chr16p11.2 | 0 |
| 14 | 218300_at | C16orf53 | chr16p11.2 | 0 |
| 15 | 209083_at | CORO1A | chr16p11.2 | 0 |
| 16 | 242414_at | QPRT | chr16p11.2 | 0 |
| 17 | 238142_at | --- | --- | 0 |
| 18 | 1559584_a_at | C16orf54 | chr16p11.2 | 0 |
| 19 | 204044_at | QPRT | chr16p11.2 | 0 |
| 20 | 229697_at | HIRIP3 | chr16p11.2 | 0 |
| 21 | 45653_at | KCTD13 | chr16p11.2 | 0 |
| 22 | 210580_x_at | SULT1A3/4 | chr16p11.2 | 0 |
| 23 | 202180_s_at | MVP | chr16p13.1-p11.2 | 0 |
| 24 | 212064_x_at | MAZ | chr16p11.2 | 0 |
| 25 | 228513_at | TMEM219 | chr16p11.2 | 0 |
| 26 | 218317_x_at | GIYD1/2 | chr16p11.2 | 0 |
| 27 | 206057_x_at | SPN | chr16p11.2 | 0.028959436 |
| 28 | 209607_x_at | SULT1A3/4 | chr16p11.2 | 0.028959436 |
| 29 | 1560841_at | LOC389247 | chr4q35.1 | 0.050679012 |
| 30 | 215813_s_at | PTGS1 | chr9q32-q33.3 | 0.050679012 |
| 31 | 238921_at | LOC644794 | chr7q11.21 | 0.050679012 |
| 32 | 244766_at | SMG1 | /// chr16p12.2 /// | 0.050679012 |
| 33 | 218453_s_at | C6orf35 | chr6q25.3 | 0.071546841 |
| 34 | 211260_at | BMP7 | chr20q13 | 0.071546841 |
| 35 | 1554245_x_at | ARL17 /// ARL17P1 | chr17q21.31 | 0.092670194 |
| 36 | 223744_s_at | SIAE | chr11q24 | 0.098885878 |
| 37 | 207925_at | CST5 | chr20p11.21 | 0.098885878 |
| 38 | 208118_x_at | IMAA | 3p11.2 /// chr16p | 0.098885878 |
| 39 | 1556103_at | --- | --- | 0.098885878 |
| 40 | 217633_at | URB1 | chr21q22.11 | 0.098885878 |
| 41 | 215734_at | C19orf36 | chr19p13.3 | 0.098885878 |
| 42 | 233334_x_at | SULT1A3/4 | chr16p11.2 | 0.130784548 |
| 43 | 232620_at | WDR93 | chr15q26.1 | 0.130784548 |
| 44 | 228083_at | CACNA2D4 | chr12p13.33 | 0.158753533 |
| 45 | 216981_x_at | SPN | chr16p11.2 | 0.158753533 |
| 46 | 215261_at | --- | --- | |
| 47 | 223179_at | YPEL3 | chr16p11.2 | |

**Table 3.2 List of top 45 significant genes from genome wide expression analysis**
Here is the list of significant genes as ranked by SAM analysis. We wanted to capture genes in which expression correlated with genotype. Genes in blue are within the 16p11.2 mutation. The top 28 significant probes are within the 16p11.2 variation. This gives us confidence that we are capturing genes which expression correlates 16p11.2 dosage. Genes in white are *trans* effects or genes outside of the variation which are positively correlated with genotype. *Trans* effects not included in the final analysis which are shown here in gray either had low expression or sequence similarity to a gene on the 16p11.2 variation.



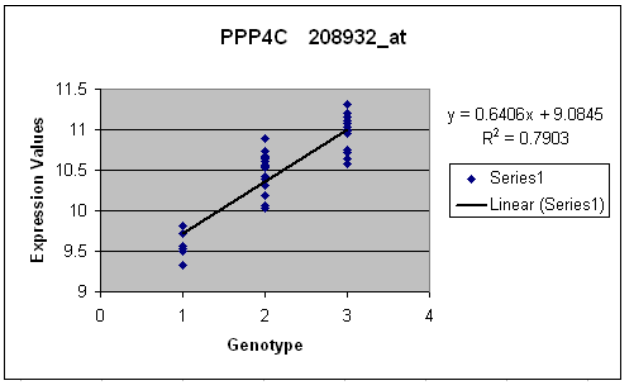




**Figure 3.3 Three probes on 16p11.2 showing that expression correlates with genotype**

Here are the top 3 probes from the SAM sampling analysis. The x axis shows the genotype at the 16p11.2 locus (1,2,3 copies). All points in the graph include all 35 samples in the analysis. The y axis shows the expression value (log base 2). All three probes show that expression levels increase with increased dosage of the 16p11.2 locus.

| Affymetrix Probe Set ID | Gene Symbol | Location | FDR | Selected GO Biological Process |
|---|---|---|---|---|
| 215813_s_at | PTGS1 | chr9q32-q33.3 | 0.050679012 | Response to oxidative stress |
| 218453_s_at | C6orf35 | chr6q25.3 | 0.071546841 | Integral to membrane |
| 211260_at | BMP7 | chr20q13 | 0.071546841 | BMP signalling pathway |
| 1554245_x_at | ARL17 | chr17q21.31 | 0.092670194 | Vessicle mediated transport |
| 223744_s_at | SIAE | chr11q24 | 0.098885878 | Carboxylesterase activity |
| 1556103_at | --- | --- | 0.098885878 | |
| 217633_at | URB1 | chr21q22.11 | 0.098885878 | |
| 215734_at | C19orf36 | chr19p13.3 | 0.098885878 | |

**Table 3.3 List of top seven probes which expression correlates with genotype**
Here is a list of the 7 genes outside of the 16p11.2 region whose expression is significantly correlated with copy number. This list was generated by removing all of the probes in the 16p11.2 CNV, probes with low expression, and probes which had sequence similarity to genes in the 16p11.2 variation.

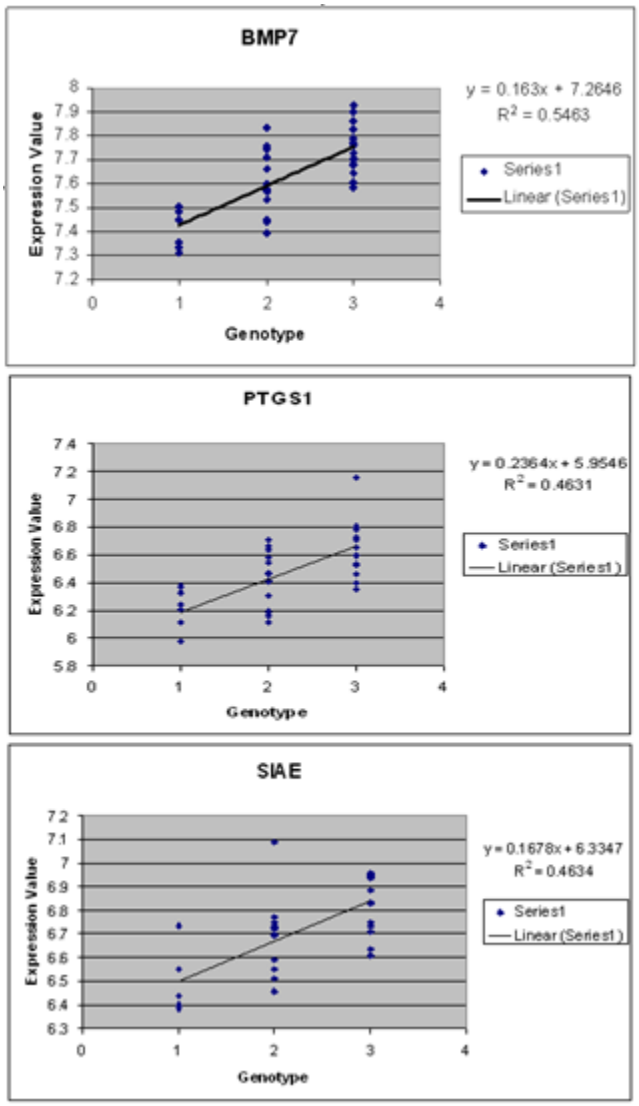



**Figure 3.4 Three probes outside of the 16p11.2 locus which show that expression correlates with genotype**

Here are three genes outside of the 16p11.2 region which expression is significantly correlated with copy number. The x axis shows the genotype at the 16p11.2 locus (1,2,3 copies). All points in the graph include all 35 samples in the analysis. The y axis shows the expression value (log base 2). Here is BMP7 Bone morphogeneic protein 7 which plays a role in development, URB1 which is ribosome biogenesis 1 homolog, and SIAE sialic acid acetylesterase.

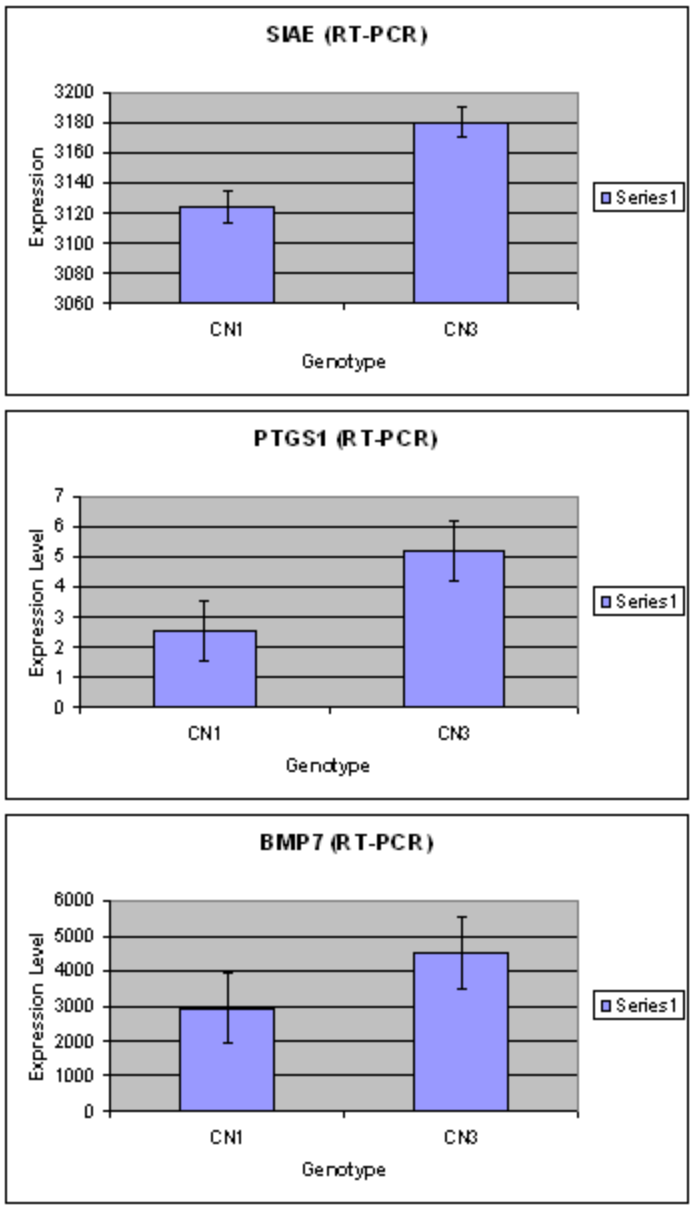



**Figure 3.5 RT-PCR confirmations of significant *trans* effects SIAE, PTGS1, and BMP7.**

mRNA expression level of three significant *trans* effects for all samples from lymphoblast cell lines used in genome wide expression analysis with a genotype of 1, and 3. This provides validation to our microarray data.

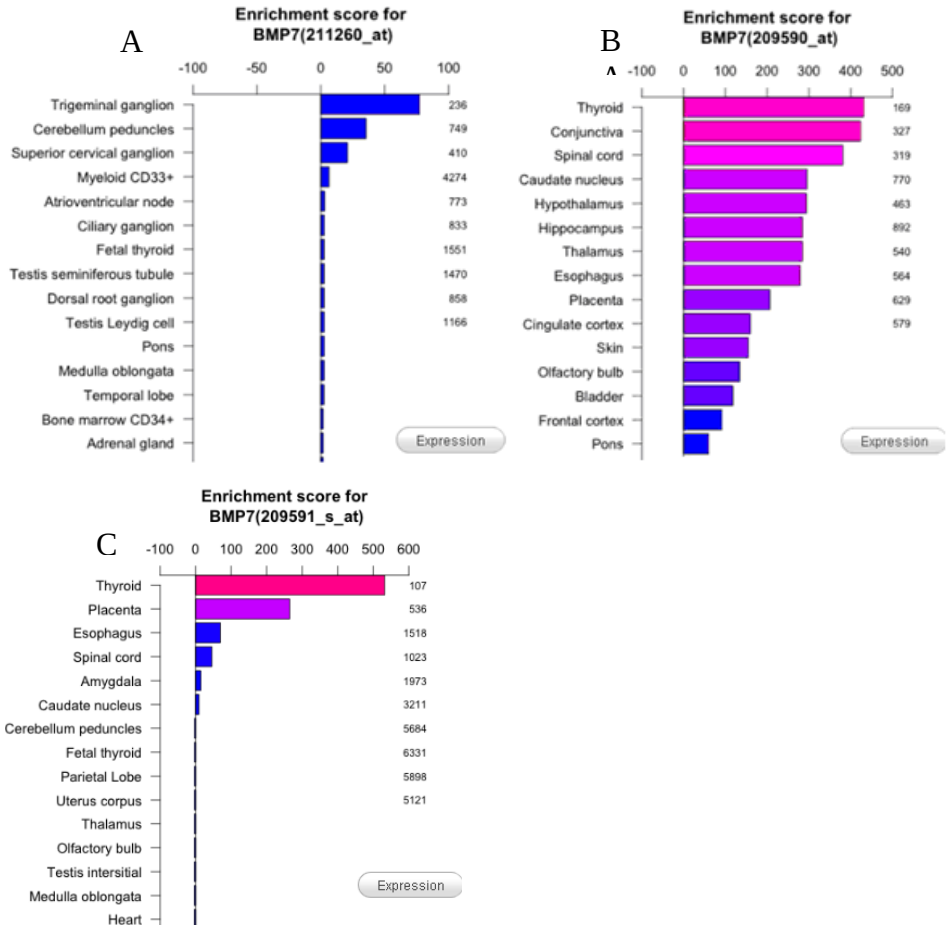



**Figure 3.6 Expression of significant *trans* effect (BMP7) is enriched in trigeminal ganglion**

The Affymetrix U133 Plus 2.0 Probe identifiers for Bmp7 (211260_at) was input into the gene enrichment profiler from the Center for Computational and Integrative Biology. It is a database with extensive expression profiles for 12,000 genes across 126 human tissues. Higher enrichment scores are related to higher tissue specificity. In figure (A) the significant BMP7 probe (211260_at) is enriched in the trigeminal ganglion, the cerebellum peduncles, and the superior cervical ganglion. In figures B and C the other probes for BMP7 (209590_at and 209591_s_at) shows specificity for the thyroid. The numbers on the right show the gene's rank in the tissue type.

**DISCUSSION:**

The 16p11.2 microduplicaton and microdeletion has been associated with an increased risk for the development of neurological disorders. In this study we used the genome wide expression data from lymphoblast cell lines to discover seven *trans* effects in which gene expression correlates with copy number at the 16p11.2 locus. Lymphoblast cell lines have been previously used to determine dysregulated genes in neurological disorders. As far as we are aware this is the first time lymphoblast cell lines were used to determine expression effects resulting from the 16p11.2 region. We applied stringent criteria to the analysis of our genome wide expression data, in addition to modifying SAM to randomly sample 18 samples for each run, and running 6000 iterations. We also applied a stringent FDR cut-off (9.88%). This only allowed genes which were positively correlated with copy number to be included in the analysis. Increasing the FDR did allow genes which were negatively correlated with copy number to be included as significant, but we were reluctant to do so, as the correlation of these probes was not as strong, and we did not want to capture probes which correlated with copy number due to noise. The first 24 probes with the highest SAM score corresponded to genes located in the 16p11.2 region. This made us confident that we were capturing probes which correlate between gene expression and dosage at the 16p11.2 region. All significant probes were analyzed for sequence similarity to genes within the 16p11.2 region. After the stringent criteria were applied the final set of significant *trans* effects were URB1, PTGS1 C6orf35, BMP7, ARL17, SIAE, c19orf36, and a cDNA; cDNA FLJ30565.

**Individual gene analysis of seven significant *trans* effects.**

In order to understand which of the seven significant *trans* effects from our genome wide expression analysis may be playing a role in the neuropsychiatric phenotype associated with the 16p11.2 region, we need to understand the these genes function. Little is known regarding the function of URB1. It is believed to be involved in ribosome biogenesis, is ubiquitously expressed through out tissues and has many predicted transcription factor binding sites. SIAE is an enzyme which removes a 9-O-acetyl group from Sialic acid, and codes for a lysosomal and cytosolic isoform [94]. ARL17 is in the ARF family, which are GTP binding proteins, and play a role in protein trafficking, vesicle budding, transport, and protein secretion [95]. ALR17 has been identified in a Region of Homozygosity (ROH) which was overrepresented in schizophrenia cases [96]. C6orf35 is a transmembrane protein of unknown function, with conservation in eukaryotes. This gene has been predicted to interact with MAPK1P1, a gene in the MAP kinase pathway, the same pathway of MAPK3 a candidate gene in the 16p11.2 region.

PTGS1(COX-1) also has a inducible isoform called COX2. PTGS1 is part of a family of cyclooxygenases which play many roles. PTGS1 is involved in prostaglandin biosynthesis. It plays a major role in the inflammatory pathway, where PTGS1 expression is inhibited by aspirin, or other anti-inflammatory drugs. Inhibiting this gene is the main reason for aspirin to be effective during cardiac

events. This gene also mediates chronic neuroinflamatory events [97]. In brains from patients of Creurzfeldt-Jakob disease PTGS1 had higher expression in diseased microglial cells/macrophages adjacent to neurons as compared to healthy brains [98]. The COX pathway functions in regulating blood flow in neonates. In in-vivo studies both COX1 and COX2 contributed to prostaglandin synthesis in neonatal human cerebral cortex, in both normal conditions, and under inflammatory stress [99]. This makes PTGS1 a good candidate in our study.

BMP7 belongs to the transforming growth factor (TGF-beta) super family of secreted signaling molecules which play several roles in embryonic development. BMPs play a role in formation of bone, and cartilage, but also play a role in neuronal and epidermal development [100-104]. Out of the seven significant *trans* effects BMP7 is among the most interesting because of its role in early development. BMP7 plays a role in limb formation, bone ossification, and also a role in the developing and injured brain [61, 77, 100]. Using a gene enrichment analyzer we found that the Affymetrix probe which corresponds to our significant *trans* effect BMP7 was enriched in three brain regions including the trigeminal ganglion, the cerebral penduncles and the superior cervical ganglion [93]. Although it is not clear as to how these brain regions (trigeminal ganglion) are associated with ASD or a schizophrenia phenotype, it is interesting that this specific probe has enriched expression in the brain.

**BMP7- a crucial gene in brain development and brain injury.**

BMP signaling pathways are crucial for the formation and maintenance of a functional brain [100, 104]. In the developing embryo BMP7 is expressed in the notocord, and the ectoderm. BMP7 receptors are expressed through out neural tube formation and cortical development [100, 105]. During brain development, BMP7 is expressed by the roof plate and acts as a heterodimer with growth/ differentiation factor 7 (GDF7) to play a direct role in axon guidance. The BMP7 gradient acts as a chemorepelent of commissural axon growth in the roof plate, which induces collapse of commissural axon growth cones [106]. BMP7 also regulates a neural cell adhesion molecule L1, in an in vitro study [107].

In several in vitro studies BMP7 has been shown to enhance dendrite growth in cortical, sympathetic and hippocampal neurons. This action was specific to dendrite length, and not axon length. Dendrite outgrowth is specific to BMP7, as similar results were not attained using BMP2 or BMP7 two other molecules in the BMP family [77, 108, 109].

BMP7 plays a protective role during neuronal cell death. Upon stroke, cerebral ischemia, or other brain trauma there is an increase in glutamate in the extracellular space to mili-molar concentrations, which causes dendrite retraction and ultimately neuron death [110]. BMP7 and its receptors are unregulated after CNS injury [100, 110]. Neuronal cell death has been shown to be ameliorated upon addition of BMP7 in a dosage depended manner in both in-vivo and in-vitro experiments [100, 110]. In one study a primary neuron culture treated with 100uM glutamate for 8 hours. Upon the addition of BMP7 in a dosage dependent manner it decreased glutamate induced cell death [110]. In another study a mild

cervical spinal cord injury was inflicted upon rats. BMP7 administered intracerebroventricularly by injection improves motor function in 2 weeks in the injured rats. [111] BMP7 has also been used as a treatment in neuronal regeneration after brain injury in mouse [100].

**BMP7 and MAPK3 a gene found in the 16p11.2 locus.**

A recent study has shown a possible link between MAPK3 a gene within the 16p11.2 region and BMP7. BDNF has been shown to induce the expression of BMP7 through the ERK/MAPK pathway [77]. Brain-derived neurotropic factor (BDNF) regulates the development of the late embryonic cerebral cortex. Upon in vivo expression of BDNF, BMP7 was up-regulated at day 14 in the mouse embryo. It was the only up-regulated TGF-beta gene out of 25 tested. ERK1(MAPK3) forms a complex with ERK2(MAPK1) to influence transcription of genes through its phosphorylation activity. In the BDNF/MAPK/ERK pathway ERK phosphorylates p53/p73 which negatively regulates it. This results in the induction of BMP7 transcription, and increased mRNA expression. [77] BMP7 has been previously shown to have a possible p53 response element on its first intron [112].

As a result of the BMP7 induced expression by BDNF/MAPK/ERK pathway, neuronal migration and synaptogenesis are affected in the developing cerebral cortex. Treatment of cortical cultures with BDNF transforms radial glia into astrocytes prematurely, which affects the radial migration of pyramidal neurons. This effect is neuronal specific as glial cells treated with BMP7 had no effect [77].

BDNF has been well characterized in both autism and schizophrenia. There have been conflicting reports as to the increase or decrease of BDNF in serum levels, lymphocytes, and brain in autism and schizophrenia patients. A study did show a significant decrease of BDNF in serum levels in autism patients [113]. Reciprocally BDNF protein was reported to be increased in the hippocampus and anterior cingulated cortex in schizophrenia patients [114]. Again there is still a debate as to the increases or decreases of BDNF in both autism, schizophrenia, and other psychiatric disorders including depression and attention deficit hyper activity disorder [115].

Interestingly a virus used to develop an animal model for autism has been shown to inhibit the phosphorylation of MAPK3/1 (ERK1/2). The Borna disease virus (BDV) has been associated with psychiatric disorders in humans [116, 117]. Mice infected with BDV have been used as an rat model for autism due to their tendencies to exhibit autistic characteristics [118]. BDV acts by impairing BDNF pathway induced phosphorylation of ERK1/2 (MAPK3/1) [119, 120]. In rat brains infected by BDV BDNF phosphorylation and activation of ERK1/2 were severely reduced as compared to controls. The infected neurons also did not show an increase in BDNF synaptic proteins, as compared to controls. The number of synapses was reduced, and clustered in the bouton of infected neurons [119]. Abnormalities in neonatal mice infected with Borna virus affects the cerebellum, reduction of serotonin in the hippocampus.

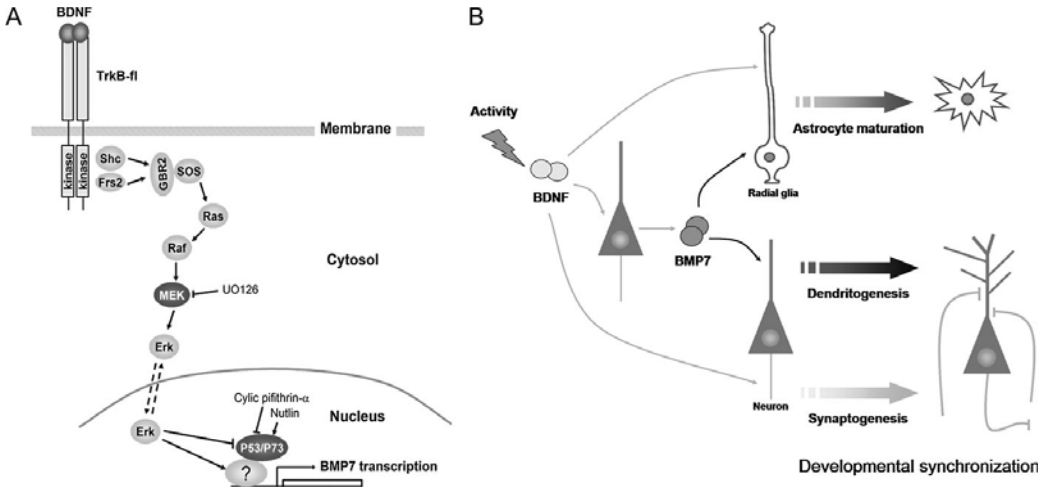



**Figure 3.7 BMP7 expression is induced by MAPK pathway.** Here is BDNF/MAPK/ERK pathway. BDNF activates the TrkB receptor which leads to the activation of the MAPK/ERK cascade. ERK transports from the Cytosol into the Nucleus where it negatively regulates p53/p73. This induces BMP7 expression to occur.

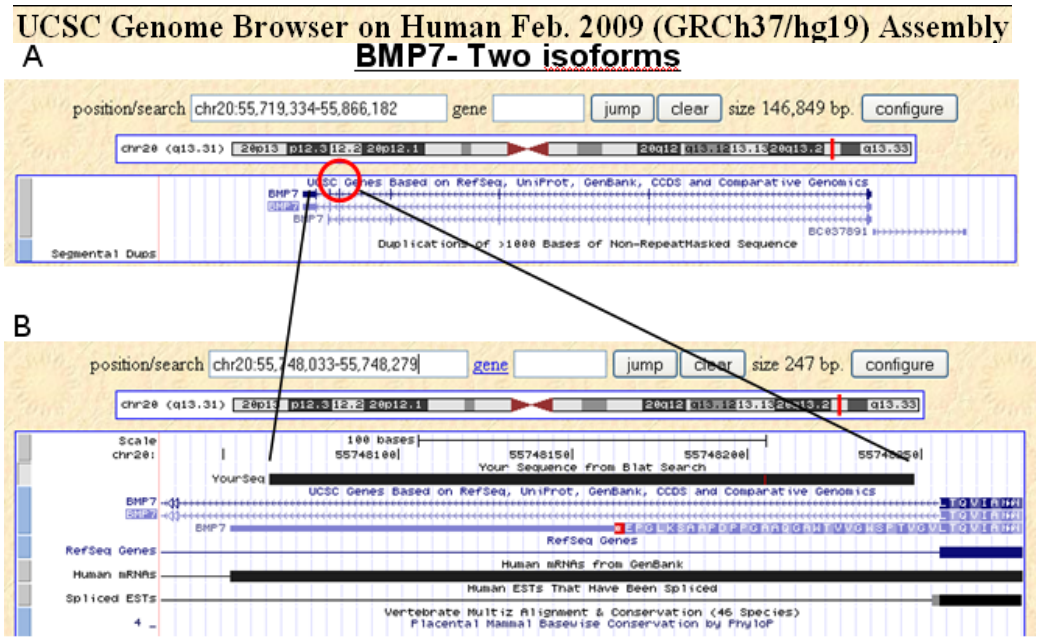
Image taken from Cereb Cortex. 2009 Dec 27. [Epub ahead of print] BDNF/MAPK/ERK-Induced BMP7 Expression in the Developing Cerebral Cortex Induces Premature Radial Glia Differentiation and Impairs Neuronal Migration.Ortega JA, Alcántara S. Unit of Cell Biology, Department of Experimental Pathology and Therapeutics, School of Medicine, University of Barcelona, 08907L'Hospitalet de Llobregat,Spain.

**BMP7 – a tale of two isoforms**

In our study the Affymetrix U133 Plus 2.0 probe which was correlated with expression was 211260_at, corresponds to a particular region in one BMP7 isoform. This isoform has 6 exons, compared to the larger isoform which has 7. This isoform codes for 29 amino acids in the last exon, exon 6 which are not found in other larger isoform.

It is interesting that the particular probe which was significant in our analysis using lymphoblasts from schizophrenic and autistic patients has enriched expression in regions in the brain however it is unclear how these brain tissues play a role in autism, schizophrenia or cognitive deficits. The other BMP7 probes which were not significant in our analysis had enriched expressed in other organs, and in particular the thyroid. It would be interesting to use bioinformatic tools

investigate the functional and molecular properties of BMP7 in order to define, such as pathway and co-expression analysis to determine a link between the clinical phenotype found in patients, and a clue to a gene in the 16p11.2 region which could be mediating its dysregulation.




C

```
>uc002xyc.2 (BMP7) length=412
MHVRSLRAAAPHSFVALWAPLFLLRSALADFSLDNEVHSSFIHRRLRSQERREMQREILS
ILGLPHRPRPHLQGKHNSAPMFMLDLYNAMAVEEGGGPGGQGFSYPYKAVFSTQGPPLAS
LQDSHFLTDADMVMSFVNLVEHDKEFFHPRYHHREFRFDLSKIPEGEAVTAAEFRIYKDY
IRERFDNETFRISVYQVLQEHLGRESDLFLLDSRTLWASEEGWLVFDITATSNHWVVNPR
HNLGLQLSVETLDGQSINPKLAGLIGRHGPQNKQPFMVAFFKATEVHFRSIRSTGSKQRS
QNRSKTPKNQEALRMANVAENSSSDQRQACKKHELYVSFRDLGWQDWIIAPEGYAAYYCE
GECAFPLNSYMNATNHAIVQTLVGVTPSWGVVTWAGQAAGPPDPAASKLGPE
```

**Figure 3.8 BMP7- a tale of two isoforms**

From the Genome browser. Examples of the two isoforms for BMP7. Figure A shows the two BMP7 isoforms in their entirety. A BLAT search in the UCSC genome browser determines sequence similarity. The black bar above the two BMP7 isoforms in Figure B shows the sequence from the BLAT search that the sequence for probe (211260_at) binds to. The probe which corresponds to the significant *trans* effect binds to the last exon of this particular BMP7 isoform which is unique. Figure C is the amion acid sequence for BMP7. Highlighted in blue is the sequence which is unique to the BMP7 isoform which was captured by our expression analysis.

From: UCSC Genome Browser: Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. *Genome Res.* 2002 Jun;12(6):996-1006.

## Chapter 4

## Bioinformatic analysis of significant *trans* effects, and 16p11.2 genes.

Mary Kusenda[1,2], Shane McCarthy[2], Vladimir Vacic[2], Pascal Grange[2].

Bioinformatic analysis using Ingenuity Pathway Analysis, InnateDB, ConsensusDB, UCLA gene expression tool UGET, and GATHER was performed by Mary Kusenda, with helpful advice from Shane McCarthy and Vladimir Vacic. Allan Brain Atlas co-expression data was generated by Pascal Grange.

[1.] Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[2.] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, NY 11724

**Summary:** Patients with the 16p11.2 microdeletion or microduplication are at an increased risk for developing neuropsychiatric disease. From our genome wide expression analysis using lymphoblast cell lines from 16p11.2 carriers, seven significant *trans* effects show positive correlation with dosage at the 16p11.2 variation. Three of these seven were analyzed using RT-PCR and validated. We hypothesized that there may be a gene or genes within the 16p11.2 variation which is influencing the expression of these significant *trans* effects. Molecular and functional interactions between the 16p11.2 genes and the seven *trans* effects would help us understand which of the 16p11.2 genes could be contributing to the dysregulation. Pathway analysis software and data bases which determine predicted transcription factor binding sites, molecular interactions, and co-expression can determine interactions, connections, and common pathways between our two gene sets. Through our analysis using pathway tools from Ingenuity Pathway Analysis (IPA), Innate DB, Consensus DB, co-expression data from UGET, GATHER and Allen Brain Atlas we have found many different indirect connections and shared pathways between the significant *trans* effects and genes on the 16p11.2 region.

**Introduction:**

The molecular pathways, interactions, and networks genes belong to are intricately complex. It is sometimes difficult to make links between genes which are not known to have a direct interaction. One form of gene analysis is Individuals Gene Analysis (IGA), in which genes in a gene set are individually analyzed to gain knowledge about a genes function [121]. This method can not be used to determine how long lists of genes function together [121]. Bioinformatics is a useful tool which integrates statistical and computing techniques to study biological data and has been applied to gene-expression studies [122]. These tools use reference data, systems biology, and knowledge bases, to give a more complete view of genes interactions with each other [122, 123].

There are different bioinformatic tools which analyze gene function, each with its advantages and disadvantages [82]. In order to understand the mechanisms behind a genes role in a pathway one can utilize reference data bases, systems biology, and knowledge bases. Gene ontology (GO) has been an initiative to unify all genes and terms across all species and databases into one standard set. This will enable bioinformatic tools to accurately determine over represented genes or pathways in a gene set [122]. Reference databases center around GO terms, and pathway analysis to determine a gene's role at a molecular level [123] [122]. Knowledge bases manage, and retrieve data which is then used to provide network and functional analysis based on GO terms. The broad and large amount of information from systems biology can be applied to gene expression studies which are gene-centric [123]. Systems biology uses knowledge bases and focuses in the integration of experimental data such as gene expression studies with computational tools designed to analyze and combine data [122, 124].

Pathway tools can manage large amounts of microarray data and find consistent pathways or functions which are enriched in a gene set [82, 123]. Pathway tools such as Ingenuity Pathway Analysis (IPA) are commercial and have an extensive knowledge base, while others such as GATHER, David, Panther, InnateDB, Consensus DB, are freeware [123] [82, 125]. Certain pathway tools use GO terms to determine over represented pathways in a set of genes. Ingenuity Pathway Analysis builds networks using direct and indirect interactions between genes found in its knowledgebase. All pathway tools reveal more about genes within the gene sets which enables better understanding of experimental data [82]. While pathway tools integrate functional and molecular data, meta-analysis combines sample data from similar studies.

We conducted a genome wide expression analysis using lymphoblast cell lines from 16p11.2 carriers. In our results seven genes outside of the mutation showed positive correlation with genotype. The final result of such a genome-wide expression study is a list of significant genes, with little knowledge regarding the interactions between those genes [123]. Analysis of the genes found through such a study is not an easy task [82, 121, 122], In IGA significant genes from an expression study which meet an cut-off value are individually analyzed to gain knowledge about a genes role [121]. Not much can be determined about how a list of dysregulated genes functions together, not to mention that the cut-off between significant and not significant genes is arbitrary [121]. Instead we used pathway tools from Ingenuity Pathway Analysis, Innate DB, Consensus DB, co-expression data from UGET, GATHER and Allen Brain Atlas to determine the interactions between the 16p11.2 genes and the seven significant *trans* effects.

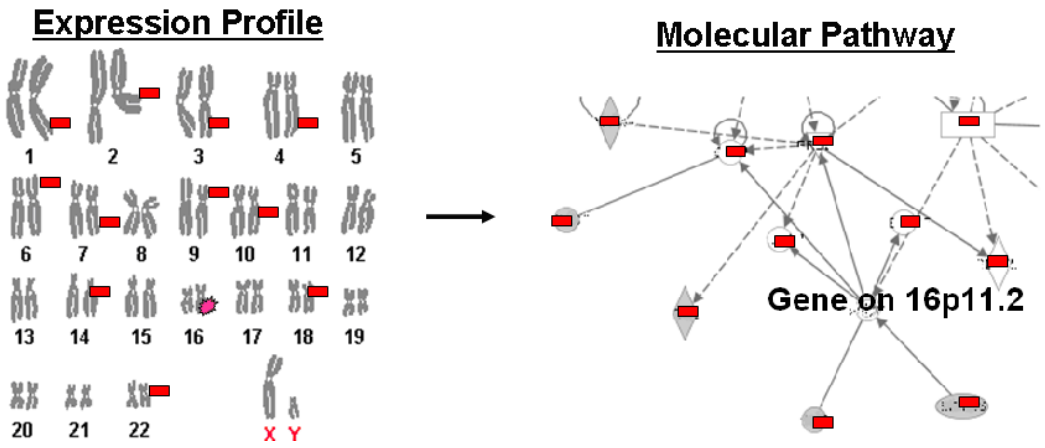



**Figure 4.1 Rationale behind using pathway tools to find links between 16p11.2 genes and the seven significant *trans* effects.**
Patients with a 16p11.2 microdeletion or microduplication have an abnormal neurological phenotype. On the left is a cartoon karyotype of a person with a 16p11.2 CNV. Such an individual would have *dosage* effects (purple star), or genes within the variation which are dysregulated, but we also expect to determine *trans* effects (red bars) or genes outside of the region which are also

dysregulated. We hypothesize that this dysregulated gene set will fit into a molecular pathway which may be associated with a gene on the 16p11.2 region.

**Materials and Methods:**

**Pathway analysis:**

One standard set of genes was created which comprised of the twenty seven 16p11.2 genes and seven significant *trans* effects from our genome wide expression analysis. This gene set was used in all Pathway Analysis and Co-expression analysis unless other wise specified.

**Ingenuity Pathway Analysis (IPA).** We used Ingenuity Pathway Analysis (Ingenuity Systems) a web based knowledge base with extensive biological pathways knowledge to create a network between the 16p11.2 genes and our seven significant *trans* effects. IPA was chosen due to its depth of biological knowledge, and its powerful computational analysis. We input all 27 genes and the seven significant *trans* genes into the IPA analysis, using direct and indirect pathway criteria, in network generation. We used data from the network, functions of pathway, and canonical pathway features.

**InnateDB.** (www.innatedb.ca) We used InnateDB to analyze molecular interactions. It is predominantly used to analyze pathways and functions involved in immunity, but it also has data on 100,000 other molecular interactions in both human and mouse. The *Interactions, Interactors, Genes, Pathways,* and *Predicted Transcription Factor Interactions* functions were used in the analysis.

**ConsensusDB. version 14** (http://cpdb.molgen.mpg.de/). ConsensusDB has data on molecular interactions from fourteen resources available to the public, which were gathered from open access databases, and are manually curated. We used the overrepresentation analysis tool.

**Gene Annotation Tool to Help Explain Relationships (GATHER).** **(**http://**gather**.genome.duke.edu/ ) We used this tool to analyze gene signatures from our list of genes against a large data base, to determine enriched functions [125].

**Co-expression Analysis:**

**UCLA Gene Expression tool (UGET)**

(http://genome.ucla.edu/~jdong/GeneCorr.html)

Co-expressed genes were determined using UGET, a database with co-normalized Affymetrix expression data. The seven *trans* effects were input into the analysis, and were analyzed in the context of the 16p11.2 genes.

**Allen Brain Atlas**. (http://www.brain-map.org.) This tool was used to determine tissue specific co-expression in the brain between the significant *trans* effect BMP7 and a gene on the 16p11.2 region MAPK3.

**RESULTS:**

We used bioinformatic tools including pathway tools, and co-expression analyzers to determine interactions between 16p11.2 genes and the seven significant *trans* effects. Determining interactions between these two gene sets may give us a clue as to which genes in the 16p11.2 region are influencing the

dysregulation of the *trans* effects, and may be contributing to the neurodevelopmental phenotype associated with this region.

Various pathway tools were used to learn about common pathways, connections and interactions between the 16p11.2 genes and the seven significant *trans* effects. Using Ingenuity Pathway Analysis, networks and canonical pathways formed through indirect and direct connections between these two gene sets were revealed. One significant network using IPA was Cellular Growth and Proliferation, Connective Tissue Development and Function. In this network BMP7, a *trans* effect was indirectly connected to ERK, a complex composed of MAKP3, and MAPK1. MAPK3 is a gene located within the 16p11.2 region. URB1 was directly connected to NCK1 which was then indirectly connected to MAPK3. The top pathways found using the *functions of pathway* feature were cell morphology, embryonic development, tissue development, cellular development, cancer and carbohydrate metabolism. Several *trans* effects were represented in these pathways including PTGS1, and BMP7. The top canonical pathways were sulfate metabolism, cysteine metabolism, chondroitin sulfate biosynthesis, and keratin sulfate biosynthesis, BMP signaling pathway, and TGF-beta signaling. The first four pathways are involved in sulfur metabolism, and biosynthesis which may be influenced by SULTA3/4, a pair of genes found in the segmental duplications which flank the 16p11.2 region.

Similar analysis using InnateDB also revealed common pathways. MAPK3 and PTGS1 are involved in a pathway involved in aspirin blocking signaling pathway involved in platelet activation. ALDOA, CDIPT, and PTGS1 are involved in Metabolic Pathways. BMP7 and MAPK3 are involved in the TGF-beta signaling pathway. Several transcription binding interactions were determined using the predicted transcription factor interactions from InnateDB. Using the interactions feature in InnateDB showed that none of the 16p11.2 genes and seven significant *trans* effects had any interactions with each other. The tool GATHER confirmed this, using its Protein binding feature.

We also looked at co-expression between 16p11.2 genes and the *trans* genes using a meta-analysis data base. Using the meta-analysis co-expression tool UGET there were a few genes which commonly co-expressed with the *trans* effects including BOLA2, SEZ6L2, KCTD13, PPP4C, and HIRIP3, with BOLA2 having the highest co-expression value. The co-expression database revealed that BOLA2, followed by MAZ, and SPN have the highest co-expression value compared to other 16p11.2 genes. Out of the seven significant *trans* effects BOLA2 has the highest co-expression value with ARL17, and c6orf35. MAZ has the highest co-expression value with URB1. SPN has the highest co-expression with PTGS1.

Co-expression between BMP7 a significant *trans* effect, and MAPK3 a gene on the 16p11.2 region was analyzed using Brain Atlas. The co-expression value between these two genes in the whole brain was 0.3304. The regions of the brain which shared highest levels of co-expression were dorsomedial nucleus of the hypothalamus, ventral part of the lateral geniculate complex and the pedunculopontine nucleus. The regions of the brain which have high co-expression do not play a clear role in autism spectrum disorders, or schizophrenia.

**Ingenuity Pathway Analysis**

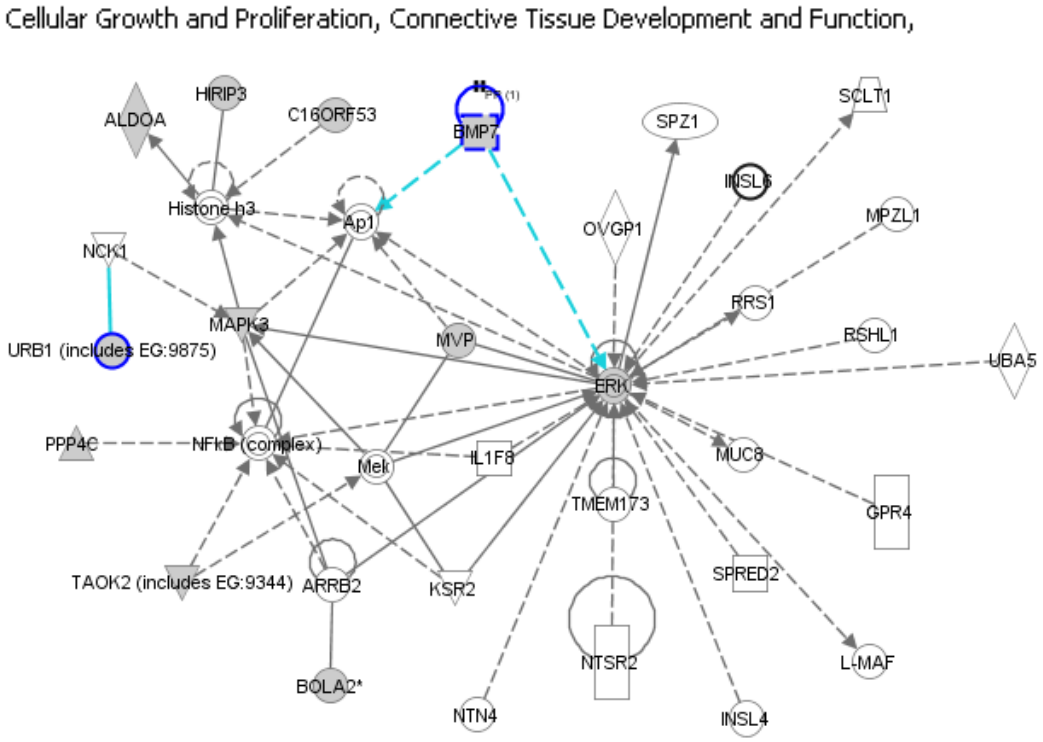



**Figure 4.2 Network between significant *trans* effects and 16p11.2 genes using Ingenuity Pathway Analysis**

Here is a network which represents Cellular Growth and Proliferation, Connective Tissue Development and Function which was built using Ingenuity Pathway Analysis Software. This network was generated using the 27 genes found in the 16p11.2 variation and the 7 *trans* effects. Webs of connections were found among these genes. The shapes in gray represent genes input into the software program for analysis, and genes in white are ones what connect them in order to make a web or a network. The two genes highlighted in blue represent two of the seven significant *trans* effects. As shown, BMP7 is connected to ERK. ERK is made up of ERK1 and ERK3. MAPK3 found within the 16p11.2 region is synonymous with ERK1. URB1 is also connected to MAPK3 through its binding to NCK1.

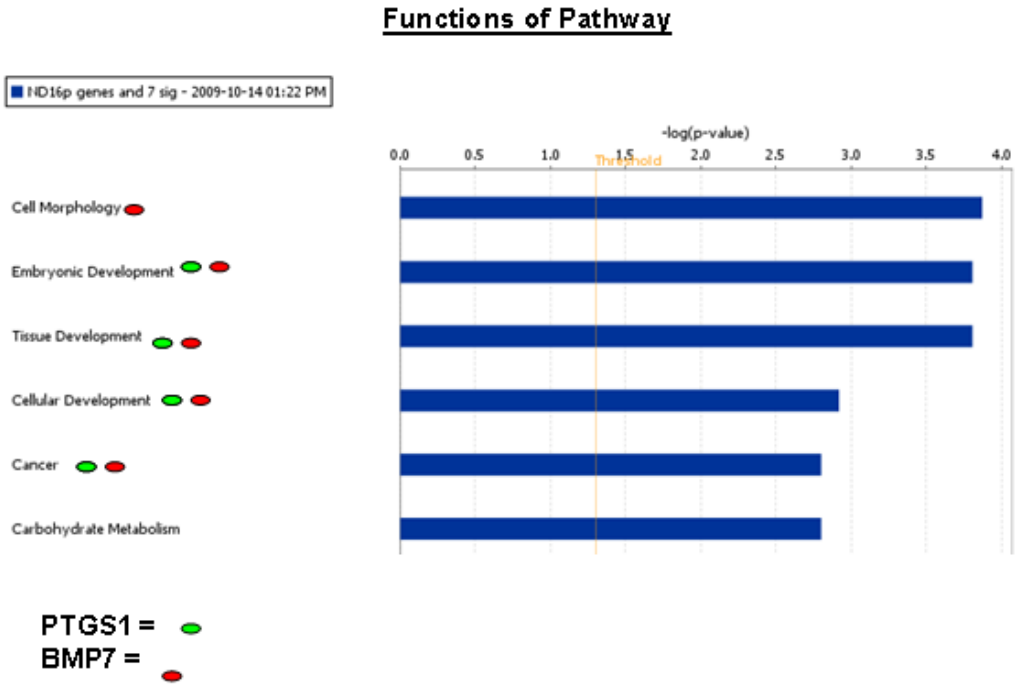



**Figure 4.3 List of top Functions from Ingenuity Pathway Analysis Ingenuity**

Here is a list of functions from Ingenuity Pathway Analysis using the same genes as in Figure 4.2. This is a list of the top 6 “functions of pathway” from the analysis. Some of our significant genes appear in these pathways. The green oval to represent PTGS1, and the red oval represents BMP7. The squares represent the p-value for the association of the genes used in the analysis and the functions of pathway.

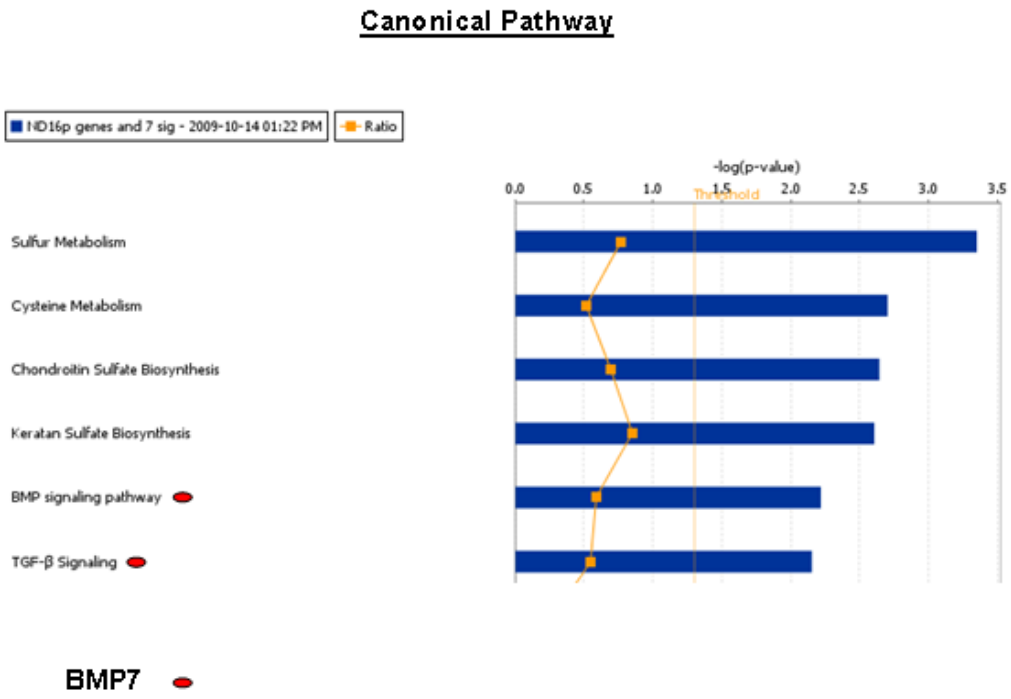



**Figure 4.4 List of top canonical pathways from Ingenuity Pathway Analysis**

Here are the top six canonical pathways as represented from our data. The 16p11.2 genes and the 7 significant *trans* effects were used in our analysis. The top 4 canonical pathways were unfortunately influenced by a sulfatransferase gene found on 16p11.2 BMP7, our significant *trans* effect is represented by the red oval. The squares represent the p-value for the association of the genes used in the analysis and the canonical pathway.

| Name | InnateDB Pathway ID | Source Database | Source Database ID | Pathway Name | Organism |
|---|---|---|---|---|---|
| MAPK3 | 3963 | PID_BIOCARTA | 100030_sppapathway | Aspirin blocks signaling pathway involved in platelet activation | Homo sapiens |
| PTGS1 | 3963 | PID_BIOCARTA | 100030_sppapathway | Aspirin blocks signaling pathway involved in platelet activation | Homo sapiens |
| ALDOA | 4373 | KEGG | hsa01100 | Metabolic pathways | Homo sapiens |
| CDIPT | 4373 | KEGG | hsa01100 | Metabolic pathways | Homo sapiens |
| PTGS1 | 4373 | KEGG | hsa01100 | Metabolic pathways | Homo sapiens |
| QPRT | 4373 | KEGG | hsa01100 | Metabolic pathways | Homo sapiens |
| BMP7 | 465 | KEGG | hsa04350 | TGF-beta signaling pathway | Homo sapiens |
| MAPK3 | 465 | KEGG | hsa04350 | TGF-beta signaling pathway | Homo sapiens |
| BMP7 | 304 | INOH | P4_TGF_beta_super_family_signaling_pathway_canonical__0 | TGF-betasuperfamilysignalingpathway(canonical) | Homo sapiens |

**Table 4.1 List of common pathways shared by significant *trans* effects, and 16p11.2 genes using InnateDB**

The 27 genes on the 16p11.2 region and seven significant *trans* effects were input into the pathway feature in InnateDB. Here are common pathways which are shared between genes in the 16p11.2 region, and our seven significant *trans* effects. *Trans* effects are highlighted in yellow. MAPK3 and PTGS1 are involved in a pathway involved in "asprin blocking signaling pathway involved in platelet activation." ALDOA, and CDIPT, and PTGS1 are involved in metabolic pathways, BMP7 and MAPK3 are involved in the TGF-beta signaling pathway.

| Transcription Factor Name | Gene Name | Gene Name | Gene Name | Gene Name | Gene Name | Gene Name |
|---|---|---|---|---|---|---|
| AIRE | C16orf54 | CDIPT | KCTD13 | SPN | URB1 | |
| CREB1 | ASPHD1 | C16orf54 | C16orf92 | TBX6 | URB1 | |
| E2F1 | ALDOA | FAM57B | GDPD3 | URB1 | | |
| NRF1 | C16orf53 | DOC2A | KCTD13 | TAOK2 | URB1 | YPEL3 |
| PDX1 | CDIPT | PTGS1 | URB1 | | | |
| POU1F1 | BMP7 | BOLA2B | CORO1A | GDPD3 | PTGS1 | TAOK2 |
| PPARA | C16orf54 | GDPD3 | OR51A4 | PRRT2 | URB1 | |
| RFX1 | C16orf54 | GIYD1 | SIAE | TBX6 | TMEM219 | |
| SP1 | BMP7 | C16orf92 | GDPD3 | KCTD13 | PRRT2 | URB1 |
| STAT5A | C16orf92 | KCTD13 | TMEM219 | SIAE | TAOK2 | |
| TFAP2A | CDIPT | DOC2A | GIYD1 | PRRT2 | PTGS1 | |
| TFDP1 | ALDOA | BMP7 | GIYD1 | | | |
| TP53 | GDPD3 | PTGS1 | YPEL3 | | | |
| XBP1 | ASPHD1 | URB1 | | | | |

**Table 4.2 List of common predicted transcription factor binding sites shared by significant *trans* effects, and 16p11.2 genes using InnateDB**

The 27 genes on the 16p11.2 region and seven significant *trans* effects were input into the *predicted transcription factor interactions* feature in InnateDB. This list shows all transcription factors which interacted with more that one gene on the 16p11.2 region or *trans* effect. The first column has the transcription factor name,

followed by the names of genes which are predicted to interact with it. The seven significant *trans* effects are highlighted in yellow.

| Query Name | Interaction Level | Interaction | Interactor Species | Interaction Type |
|---|---|---|---|---|
| ARL17P1;ARL17 | | No interactions | | |
| ARL17P1;ARL17 | | No interactions | | |
| BMP7 | Direct interaction | Acvr2a and BMP7 | Mus musculus/Homo sapiens | direct interaction |
| BMP7 | Direct interaction | Acvr2a interacts with BMP7 | Mus musculus/Homo sapiens | physical association |
| BMP7 | Direct interaction | BMP7 interacts with BMPR2 | Homo sapiens | physical association |
| BMP7 | Direct interaction | BMP7 interacts with Bmpr1b | Mus musculus/Homo sapiens | physical association |
| BMP7 | Direct interaction | BMP7 interacts with BMPR2 | Homo sapiens | unspecified |
| BMP7 | Direct interaction | BMP7 interacts with SOSTDC1 | Homo sapiens | unspecified |
| BMP7 | Direct interaction | BMP7 interacts with NOG | Homo sapiens | unspecified |
| BMP7 | Direct interaction | BMP7 interacts with ENG | Homo sapiens | unspecified |
| BMP7 | Direct interaction | ACVR2B interacts with BMP7 | Homo sapiens | unspecified |
| BMP7 | Direct interaction | ACVR1 interacts with BMP7 | Homo sapiens | unspecified |
| BMP7 | Direct interaction | BMP7 interacts with SMAD1 | Homo sapiens | unspecified |
| BMP7 | Direct interaction | BMP7 interacts with BMP7 | Homo sapiens | unspecified |
| BMP7 | Direct interaction | ACVR2A interacts with BMP7 | Homo sapiens | unspecified |
| BMP7 | Direct interaction | BMP7 interacts with BMPR1A | Homo sapiens | unspecified |
| BMP7 | Direct interaction | BMP7 interacts with BMPR1B | Homo sapiens | unspecified |
| C19orf36 | | No interactions | | |
| C6orf35 | | No interactions | | |
| PTGS1 | Direct interaction | NUCB1 interacts with PTGS1 | Homo sapiens | unspecified |
| PTGS1 | Direct interaction | PTGS1 interacts with PTGS1 | Homo sapiens | unspecified |
| SIAE | | No interactions | | |
| URB1 | Direct interaction | NCK1 interacts with URB1 | Homo sapiens | physical association |

**Table 4.3 List of interactions between significant *trans* effects, and 16p11.2 genes using InnateDB**

The seven significant *trans* effects were input into the *Interactions* feature in InnateDB. This list shows the interactions between the seven significant *trans* effects and other genes. There are no genes on the 16p11.2 region which interact with the seven significant *trans* effects through this analysis.

| | *p-value* | *q-value* | *radius* | *centers_names* | *set_sources* |
|---|---|---|---|---|---|
| 1 | 1.05E-08 | 3.21E-06 | 1 | Coiled coil domain containing 6 | HPRD; PIG; Spike |
| 2 | 4.41E-08 | 6.72E-06 | 1 | ETS-domain protein Elk-1-2 | BioCarta; Biogrid; CORUM; HPRD; INOH; IntAct-LS; IntAct-SS; MINT; NetPath; PID; PIG; Reactome; Spike |
| 3 | 3.66E-07 | 3.72E-05 | 1 | MEK1/2(MKK1/2) | INOH; Spike |
| 4 | 2.28E-06 | 0.000174 | 1 | DUSP4 | BioCarta; Biogrid; HPRD; PIG; Reactome; Spike |
| 5 | 7.01E-06 | 0.000427 | 1 | dcc_human | Biogrid; HPRD; IntAct-SS; PIG; Spike |
| 6 | 1.36E-05 | 0.000593 | 1 | 6-phosphofructokinase, muscle type | BioCarta; Biogrid; EHMN; HPRD; HumanCyc; IntAct-LS; IntAct-SS; KEGG; PIG; Reactome; Spike |
| 7 | 1.36E-05 | 0.000593 | 1 | Noggin | BioCarta; Biogrid; HPRD; IntAct-SS; PID; PIG; Reactome; Spike |
| 8 | 1.72E-05 | 0.000656 | 1 | Proteasome subunit beta type 5 | BioCarta; Biogrid; CORUM; DIP; HPRD; IntAct-LS; IntAct-SS; MINT; NetPath; PIG; Reactome; Spike |
| 9 | 2.06E-05 | 0.000699 | 1 | GEMIN4 | Biogrid |
| 10 | 2.34E-05 | 0.000714 | 1 | CHRD | BioCarta; Biogrid; HPRD; IntAct-SS; MINT; PID; PIG; Spike |

**Table 4.4 List of overrepresented neighborhood based sets between significant *trans* effects, and 16p11.2 genes using ConnsensusDB**
Here are the top 10 enriched neighborhood based sets using the over representation analysis in ConsensuDB. MEK1/2 appears on this list as a result of MAPK3 a gene on the 16p11.2 region.

| # | Annotation | ln(Bayes factor) | Genes |
|---|---|---|---|
| 1 | DCC: deleted in colorectal carcinoma | 6.4 | MAPK3 MAZ |
| 2 | PARP4: poly (ADP-ribose) polymerase family, member 4 | 3.59 | MVP |
| 3 | PPP4R2: protein phosphatase 4, regulatory subunit 2 | 3.59 | PPP4C |
| 4 | DUSP5: dual specificity phosphatase 5 | 3.59 | MAPK3 |
| 5 | FALZ: fetal Alzheimer antigen | 3.59 | MAZ |
| 6 | GPD2: glycerol-3-phosphate dehydrogenase 2 (mitochondrial) | 3.59 | ALDOA |
| 7 | PPP4R1: protein phosphatase 4, regulatory subunit 1 | 3.59 | PPP4C |
| 8 | SN: sialoadhesin | 2.91 | SPN |
| 9 | TH: tyrosine hydroxylase | 2.91 | MAPK3 |
| 10 | ARMC7: armadillo repeat containing 7 | 2.91 | KCTD13 |
| 11 | IER3: immediate early response 3 | 2.91 | MAPK3 |
| 12 | DUSP4: dual specificity phosphatase 4 | 2.5 | MAPK3 |
| 13 | GMFB: glia maturation factor, beta | 2.5 | MAPK3 |
| 14 | PTPN7: protein tyrosine phosphatase, non-receptor type 7 | 2.5 | MAPK3 |
| 15 | PTPRR: protein tyrosine phosphatase, receptor type, R | 2.5 | MAPK3 |
| 16 | RPS6KA2: ribosomal protein S6 kinase, 90kDa, polypeptide 2 | 2.5 | MAPK3 |
| 17 | HIST1H4A: histone 1, H4a | 2.5 | HIRIP3 |
| 18 | PTPN5: protein tyrosine phosphatase, non-receptor type 5 (striatum-enriched) | 2.5 | MAPK3 |
| 19 | ZMYND19: zinc finger, MYND domain containing 19 | 2.22 | KCTD13 |
| 20 | DUSP3: dual specificity phosphatase 3 (vaccinia virus phosphatase VH1-related) | 2.22 | MAPK3 |
| 21 | DUSP6: dual specificity phosphatase 6 | 2.22 | MAPK3 |
| 22 | TAL2: T-cell acute lymphocytic leukemia 2 | 2.22 | MAPK3 |
| 23 | TMSB4X: thymosin, beta 4, X-linked | 2.22 | PPP4C |
| 24 | NOG: noggin | 2.22 | BMP7 |
| 25 | DUSP1: dual specificity phosphatase 1 | 1.99 | MAPK3 |
| 26 | IGBP1: immunoglobulin (CD79A) binding protein 1 | 1.99 | PPP4C |
| 27 | ATP1A1: ATPase, Na+/K+ transporting, alpha 1 polypeptide | 1.99 | MAPK3 |
| 28 | RPS6KA3: ribosomal protein S6 kinase, 90kDa, polypeptide 3 | 1.99 | MAPK3 |
| 29 | SNCG: synuclein, gamma (breast cancer-specific protein 1) | 1.99 | MAPK3 |
| 30 | RPS6KA4: ribosomal protein S6 kinase, 90kDa, polypeptide 4 | 1.99 | MAPK3 |
| 31 | UNC13B: unc-13 homolog B (C. elegans) | 1.81 | DOC2A |
| 32 | POLD2: polymerase (DNA directed), delta 2, regulatory subunit 50kDa | 1.81 | KCTD13 |

| | | | |
|---|---|---|---|
| 33 | MAP2K2: mitogen-activated protein kinase kinase 2 | 1.81 | MAPK3 |
| 34 | MAP2K3: mitogen-activated protein kinase kinase 3 | 1.81 | MAPK3 |
| 35 | MAP2K1IP1: mitogen-activated protein kinase kinase 1 interacting protein 1 | 1.81 | MAPK3 |
| 36 | RPS6KA1: ribosomal protein S6 kinase, 90kDa, polypeptide 1 | 1.66 | MAPK3 |
| 37 | MYLK: myosin, light polypeptide kinase | 1.53 | MAPK3 |
| 38 | ATP6V1E1: ATPase, H+ transporting, lysosomal 31kDa, V1 subunit E isoform 1 | 1.53 | ALDOA |
| 39 | HIRA: HIR histone cell cycle regulation defective homolog A (S. cerevisiae) | 1.53 | HIRIP3 |
| 40 | GATA4: GATA binding protein 4 | 1.41 | MAPK3 |
| 41 | REL: v-rel reticuloendotheliosis viral oncogene homolog (avian) | 1.41 | PPP4C |
| 42 | BMPR1B: bone morphogenetic protein receptor, type IB | 1.41 | BMP7 |
| 43 | SPIB: Spi-B transcription factor (Spi-1/PU.1 related) | 1.41 | MAPK3 |
| 44 | TNFSF11: tumor necrosis factor (ligand) superfamily, member 11 | 1.41 | MAPK3 |
| 45 | ENG: endoglin (Osler-Rendu-Weber syndrome 1) | 1.31 | BMP7 |
| 46 | PJA1: praja 1 | 1.31 | KIF22 |
| 47 | SREBF2: sterol regulatory element binding transcription factor 2 | 1.31 | MAPK3 |
| 48 | STXBP1: syntaxin binding protein 1 | 1.31 | DOC2A |
| 49 | ABCF3: ATP-binding cassette, sub-family F (GCN20), member 3 | 1.21 | HIRIP3 |
| 50 | BMPR2: bone morphogenetic protein receptor, type II (serine/threonine kinase) | 1.21 | BMP7 |
| 51 | ELK1: ELK1, member of ETS oncogene family | 1.13 | MAPK3 |
| 52 | LGALS1: lectin, galactoside-binding, soluble, 1 (galectin 1) | 1.13 | SPN |
| 53 | HIST2H2BE: histone 2, H2be | 1.13 | HIRIP3 |
| 54 | ACVR2: activin A receptor, type II | 1.13 | BMP7 |
| 55 | ACVR2B: activin A receptor, type IIB | 1.13 | BMP7 |
| 56 | C6orf55: chromosome 6 open reading frame 55 | 1.05 | KCTD13 |
| 57 | PTEN: phosphatase and tensin homolog (mutated in multiple advanced cancers 1) | 1.05 | MVP |
| 58 | PEA15: phosphoprotein enriched in astrocytes 15 | 1.05 | MAPK3 |
| 59 | ACVR1: activin A receptor, type I | 1.05 | BMP7 |
| 60 | GTF2I: general transcription factor II, i | 0.97 | MAPK3 |
| 61 | MSN: moesin | 0.97 | SPN |
| 62 | BMPR1A: bone morphogenetic protein receptor, type IA | 0.97 | BMP7 |
| 63 | C1QBP: complement component 1, q subcomponent binding protein | 0.97 | MAPK3 |
| 64 | HIST3H3: histone 3, H3 | 0.97 | HIRIP3 |
| 65 | SCAM-1: vinexin beta (SH3-containing adaptor molecule-1) | 0.91 | MAPK3 |
| 66 | SNAP25: synaptosomal-associated protein, 25kDa | 0.91 | LOC112476 |
| 67 | MAP2K1: mitogen-activated protein kinase kinase 1 | 0.73 | MAPK3 |
| 68 | CASP9: caspase 9, apoptosis-related cysteine protease | 0.73 | MAPK3 |
| 69 | HDAC4: histone deacetylase 4 | 0.68 | MAPK3 |

| | | | |
|---|---|---|---|
| 70 | PLD2: phospholipase D2 | 0.63 | ALDOA |
| 71 | HIF1A: hypoxia-inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor) | 0.58 | MAPK3 |
| 72 | SUFU: suppressor of fused homolog (Drosophila) | 0.58 | QPRT |
| 73 | SIAH1: seven in absentia homolog 1 (Drosophila) | 0.41 | KIF22 |
| 74 | MYST2: MYST histone acetyltransferase 2 | 0.38 | KCTD13 |
| 75 | NCOA3: nuclear receptor coactivator 3 | 0.35 | BMP7 |
| 76 | EPOR: erythropoietin receptor | 0.31 | MAPK3 |
| 77 | XRN1: 5'-3' exoribonuclease 1 | 0.28 | ALDOA |
| 78 | FLJ22494: hypothetical protein FLJ22494 | 0.28 | KCTD13 |
| 79 | IMMT: inner membrane protein, mitochondrial (mitofilin) | 0.24 | KIF22 |
| 80 | THBS1: thrombospondin 1 | 0.18 | CORO1A |
| 81 | VIL2: villin 2 (ezrin) | 0.15 | SPN |
| 82 | GDF9: growth differentiation factor 9 | 0.05 | KIF22 |
| 83 | MAPK14: mitogen-activated protein kinase 14 | 0.02 | MAPK3 |
| 84 | INSR: insulin receptor | -0.03 | MAPK3 |
| 85 | LNX: ligand of numb-protein X | -0.09 | KCTD13 |
| 86 | DNCL1: dynein, cytoplasmic, light polypeptide 1 | -0.18 | ALDOA |
| 87 | NFKB1: nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 (p105) | -0.22 | PPP4C |
| 88 | RELA: v-rel reticuloendotheliosis viral oncogene homolog A, nuclear factor of kappa light polypeptide gene enhancer in B-cells 3, p65 (avian) | -0.35 | PPP4C |
| 89 | LYN: v-yes-1 Yamaguchi sarcoma viral related oncogene homolog | -0.37 | MAPK3 |
| 90 | PCNA: proliferating cell nuclear antigen | -0.38 | KCTD13 |
| 91 | KRTAP4-12: keratin associated protein 4-12 | -0.58 | QPRT |
| 92 | ESR1: estrogen receptor 1 | -0.7 | MVP |
| 93 | FYN: FYN oncogene related to SRC, FGR, YES | -0.75 | SPN |
| 94 | JUN: v-jun sarcoma virus 17 oncogene homolog (avian) | -0.75 | MAPK3 |
| 95 | PXN: paxillin | -1.18 | MVP |
| 96 | TP53: tumor protein p53 (Li-Fraumeni syndrome) | -1.29 | MAPK3 |

**Table 4.5 List of protein binding interactions between significant *trans* effects, and 16p11.2 genes using GATHER.**

The 27 genes on the 16p11.2 region and seven significant *trans* effects were input into the *Protein Binding* feature in GATHER. There was no protein which bound to more than one gene from the two gene sets in this list. Higher Bayes factors corresponds to stronger evidence for protein binding. Only two genes of the 16p11.2 region, MAPK3 and MAZ bind to the same protein (DCC, Deleted in Colerectal Cancer.

| | ARL17 | BMP7 | C19orf36 | C6orf35 | PTGS1 | SIAE | URB1 |
|---|---|---|---|---|---|---|---|
| | 1554245_x_at | 211260_at | 223744_s_at | 218453_s_at | 215813_s_at | 223744_s_at | 217633_at |
| 1 | BOLA2 | SEZ6L2 | LOC440356 | BOLA2 | SPN | MAPK3 | LOC100132767 |
| 2 | PPP4C | SEZ6L2 | FLJ25404 | BOLA2 | SPN | KCTD13 | LOC100132767 |
| 3 | PKMYT1 | SPN | KCTD13 | PRRT2 | MVP | DOC2A | TBX6 |
| 4 | QPRT | LOC283875 | KCTD13 | --- | SPN | SEZ6L2 | MAZ |
| 5 | HIRIP3 | TAOK2 | HIRIP3 | CDIPT | PPP4C | SEZ6L2 | SULT1A4 |
| 6 | RPS10 | FAM57B | PKMYT1 | ASPHD1 | LOC124446 | ASPHD1 | HIRIP3 |
| 7 | C16orf53 | TBX6 | DOC2A | HIRIP3 | ALDOA | SEZ6L2 | --- |
| 8 | BOLA2 | SPN | LOC124446 | LOC124446 | ALDOA | KCTD13 | BOLA2 |
| 9 | SPN | LOC100134408 | FAM57B | MAZ | GIYD1 | FLJ25404 | PKMYT1 |
| 10 | MAZ | BOLA2 | CCDC95 | ALDOA | CDIPT | TAOK2 | KREMEN2 |
| 11 | ALDOA | MAPK3 | BOLA2 | ASPHD1 | CORO1A | --- | TAOK2 |
| 12 | SULT1A4 | FLJ25404 | TAOK2 | C16orf53 | SULT1A4 | SEZ6L2 | --- |
| 13 | ALDOA | LOC100132767 | RPS10 | KCTD13 | LOC124446 | PAQR4 | FLJ25404 |
| 14 | GIYD1 | --- | HIRIP3 | SEZ6L2 | RPS10 | SPN | GIYD1 |
| 15 | ALDOA | TBX6 | --- | HIRIP3 | DKFZp547E087 | --- | --- |
| 16 | MAZ | KREMEN2 | PPP4C | SEZ6L2 | ALDOA | ALDOA | --- |
| 17 | QPRT | --- | LOC100133019 | LOC440356 | TAOK2 | BOLA2 | SEZ6L2 |
| 18 | HIRIP3 | MAZ | BOLA2 | ALDOA | QPRT | SPN | --- |
| 19 | CCDC95 | LOC100132767 | PRRT2 | BOLA2 | TNFRSF12A | LOC283875 | SPN |
| 20 | --- | --- | TAOK2 | DOC2A | LOC283875 | TBX6 | --- |

**Table 4.6 Co-expression between significant *trans* effects, and 16p11.2 genes using UGET a co-expression analyzer**

The seven significant *trans* effects were input into UGET a co-expression analyzer. Each probe from the significant *trans* effects was extracted from the data and ordered by co-expression value. Here are the top twenty genes from the 16p11.2 region which are co-expressed with each significant *trans* effect. There are no 16p11.2 genes which stand out, and are highly co-expressed through out the *trans* effects. There are a few which are common among the top 5 co-expressed genes including BOLA2, SEZ6L2, KCTD13, PPP4C, and HIRIP3.

| | Gene name | | | ARL17 | BMP7 | C19orf36 | C6orf35 | PTGS1 | SIAE | URB1 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | BOLA2 | 231500_s_ | 0.117131 | -0.1223 | -0.2291 | -0.0139 | 0.2144 | -0.0214 | -0.0463 | -0.1211 |
| 2 | BOLA2 | 241644_at | 0.112341 | -0.1383 | 0.3341 | -0.0438 | 0.0589 | -0.0565 | 0.0901 | 0.001 |
| 3 | MAZ | 228798_x_ | 0.106455 | -0.1606 | -0.0855 | -0.0195 | 0.1404 | -0.045 | -0.0605 | -0.1114 |
| 4 | --- | 234825_at | 0.101634 | -0.1844 | -0.1136 | -0.021 | 0.1776 | 0.0034 | -0.0388 | -0.0616 |
| 5 | SPN | 216981_x_ | 0.087766 | 0.0235 | 0.0344 | -0.0398 | -0.0859 | 0.1115 | -0.0739 | -0.0044 |
| 6 | SPN | 206057_x_ | 0.078921 | 0.0496 | 0.0453 | -0.0358 | -0.1028 | 0.1619 | -0.0739 | 0.0142 |
| 7 | SULT1A3 | 222094_at | 0.071279 | -0.087 | 0.0699 | -0.0621 | -0.0786 | 0.0091 | -0.0376 | 0.0132 |
| 8 | MAZ | 212064_x_ | 0.069686 | -0.0541 | -0.0981 | -0.0707 | 0.0396 | -0.0678 | -0.062 | -0.0871 |
| 9 | KCTD13 | 45653_at | 0.069617 | -0.1204 | 0.0373 | 0.2532 | 0.0889 | -0.0872 | 0.1562 | 0.0395 |
| 10 | LOC283901 | 1558525_a | 0.055114 | -0.1505 | -0.0156 | -0.0212 | -0.0518 | -0.0029 | -0.0795 | 7.00E-04 |
| 11 | DKFZp547E( | 215002_at | 0.054583 | -0.1407 | 0.0647 | -0.0231 | -0.0499 | 0.04 | -0.0735 | -0.1392 |
| 12 | --- | 1564130_x | 0.054207 | 0.1318 | 0.2412 | -0.0374 | -0.0963 | 0.0074 | 0.1415 | 0.0941 |
| 13 | C16orf53 | 231878_at | 0.051783 | -0.0141 | -0.2159 | -0.0556 | 0.0921 | -0.0153 | -0.1932 | -0.0189 |
| 14 | KCTD13 | 221889_at | 0.047772 | -0.1144 | 0.1087 | 0.2574 | 0.0431 | -0.0917 | 0.2819 | 0.0656 |
| 15 | LOC440356 | 240537_s_ | 0.04159 | -0.0379 | -0.0111 | 0.8754 | 0.0623 | -0.0463 | 0.0239 | 0.0113 |
| 16 | SPN | 1568964_x | 0.0394 | 0.2087 | 0.0562 | -0.022 | -0.066 | 0.1869 | -0.0766 | 0.0029 |
| 17 | HIRIP3 | 229697_at | 0.038548 | 0.161 | -0.0202 | 0.0046 | 0.144 | -0.1039 | -0.0965 | 0.1995 |
| 18 | FLJ25404 | 1557162_a | 0.037707 | -0.0847 | 0.3104 | 0.274 | -0.1478 | -0.0528 | 0.1499 | 0.1554 |
| 19 | YPEL3 | 232077_s_ | 0.032855 | -0.2 | -0.0367 | -0.0399 | -0.1045 | 0.0054 | -0.1197 | 0.0171 |
| 20 | --- | 1564128_a | 0.030741 | 0.0607 | 0.1175 | -0.0401 | -0.0146 | 0.0092 | -0.0151 | 0.0335 |

**Table 4.7 Top 20 genes which have highest co-expression with significant *trans* effects**

The seven significant *trans* effects were input into UGET a co-expression analyzer. Each 16p11.2 gene was extracted from the data and ordered by co-expression value with the seven significant *trans* effects combined. Here are the top twenty genes from the 16p11.2 region which are co-expressed with the combined significant *trans* effect. Two probes in BOLA2 stand out as being having the highest co-expression with the *trans* effects.

| Region ('fine' annotation) | BMP7 / MAPK3 similarity coefficient |
|---|---|
| 'Accessory olfactory bulb' | Na |
| 'Spinal nucleus of the trigeminal_ caudal part' | Na |
| 'Dorsomedial nucleus of the hypothalamus' | 0.9032 |
| 'Ventral part of the lateral geniculate complex' | 0.8415 |
| 'Pedunculopontine nucleus' | 0.8296 |
| 'Cerebellar nuclei' | 0.8251 |
| 'Ventral medial nucleus of the thalamus' | 0.7878 |
| 'Anterior hypothalamic nucleus' | 0.7753 |
| 'Epithalamus' | 0.7451 |
| 'Lateral dorsal nucleus of thalamus' | 0.7421 |
| 'Cuneiform nucleus' | 0.7283 |
| 'Posterior hypothalamic nucleus' | 0.7228 |
| 'Postpiriform transition area' | 0.7103 |
| 'Superior colliculus_ sensory related' | 0.6931 |
| 'Ventral tegmental area' | 0.6918 |
| 'Mediodorsal nucleus of thalamus' | 0.684 |
| 'Hypoglossal nucleus' | 0.6488 |
| 'Tuberal nucleus' | 0.6434 |
| 'Piriform-amygdalar area' | 0.6411 |
| 'Arcuate hypothalamic nucleus' | 0.639 |
| 'Zona incerta' | 0.6372 |
| 'Motor nucleus of trigeminal' | 0.6342 |
| 'Principal sensory nucleus of the trigeminal' | 0.6327 |
| 'Central amygdalar nucleus ' | 0.6307 |
| 'Superior central nucleus raphé' | 0.6292 |
| 'Mammillary body' | 0.6278 |
| 'Spinal nucleus of the trigeminal_ oral part' | 0.6236 |
| 'Intralaminar nuclei of the dorsal thalamus' | 0.5984 |
| 'Anteromedial nucleus' | 0.5933 |
| 'Fundus of striatum' | 0.5905 |
| 'Ventral group of the dorsal thalamus' | 0.587 |
| 'Dorsal part of the lateral geniculate complex' | 0.586 |
| 'Reticular nucleus of the thalamus' | 0.5817 |
| 'Midbrain raphé nuclei' | 0.5747 |
| 'Pallidum_ medial region' | 0.5621 |
| 'Nucleus of the lateral lemniscus' | 0.5549 |
| 'Lateral septal nucleus ' | 0.5501 |
| 'Pretectal region' | 0.5452 |
| 'Bed nuclei of the stria terminalis ' | 0.5441 |
| 'Superior olivary complex' | 0.5425 |
| 'Medial geniculate complex' | 0.5351 |
| 'Paragigantocellular reticular nucleus' | 0.531 |
| 'Parabrachial nucleus' | 0.5156 |
| 'Piriform area' | 0.5121 |
| 'Taenia tecta' | 0.5111 |
| 'Parafascicular nucleus' | 0.5086 |

| | |
|---|---|
| 'Retrohippocampal region' | 0.5062 |
| 'Lateral posterior nucleus of the thalamus' | 0.5017 |
| 'Pallidum_ dorsal region' | 0.499 |
| 'Subiculum' | 0.4949 |
| 'Hypothalamic lateral zone' | 0.49 |
| 'Vestibular nuclei' | 0.4857 |
| 'Anterior pretectal nucleus ' | 0.4821 |
| 'Ventromedial hypothalamic nucleus' | 0.4797 |
| 'Lateral group of the dorsal thalamus' | 0.4794 |
| 'Red Nucleus' | 0.4703 |
| 'Cerebellar cortex' | 0.4664 |
| 'Ammon's Horn' | 0.4642 |
| 'Olfactory tubercle ' | 0.458 |
| 'Magnocellular reticular nucleus' | 0.4546 |
| 'Substantia nigra_ reticular part' | 0.4532 |
| 'Anteroventral nucleus of thalamus ' | 0.4456 |
| 'Nucleus accumbens ' | 0.4352 |
| 'Hypothalamus' | 0.4118 |
| 'Midbrain reticular nucleus_ retrorubral area' | 0.4085 |
| 'Superior colliculus_ motor related' | 0.3934 |
| 'Ventral posterior complex of the thalamus' | 0.3892 |
| 'Pallidum_ ventral region' | 0.383 |
| 'Dentate gyrus' | 0.3799 |
| 'Inferior olivary complex' | 0.3699 |
| 'Periaqueductal gray' | 0.3678 |
| 'Tegmental reticular nucleus' | 0.3596 |
| 'Cortical amygdalar area' | 0.3567 |
| 'Dorsal column nuclei' | 0.3491 |
| 'Facial motor nucleus' | 0.3442 |
| 'Cochlear nuclei ' | 0.3436 |
| 'Nucleus of the lateral olfactory tract' | 0.3306 |
| 'Spinal nucleus of the trigeminal interpolarpart | 0.3113 |
| 'Nucleus of the solitary tract' | 0.3072 |
| 'Midbrain' | 0.3061 |
| 'Inferior colliculus' | 0.2958 |
| 'Medial amygdalar nucleus ' | 0.2946 |
| 'Cerebral cortex' | 0.2902 |
| 'Caudoputamen ' | 0.2858 |
| 'Medulla' | 0.2672 |
| 'Substantia nigra_ compact part' | 0.2522 |
| 'Anterior amygdalar area' | 0.2175 |
| 'Pons' | 0.1952 |
| 'Pontine gray' | 0.1926 |
| 'Medulla_ behavioral state related' | 0.1779 |
| 'Pontine central gray' | 0.1554 |
| 'Main olfactory bulb' | 0.136 |
| 'Anterior olfactory nucleus' | 0.1205 |
| 'Lateral reticular nucleus' | 0.046 |

**Table 4.8. (p75-76) List of brain tissue (fine annotation) with highest co-expression values between MAPK3, and BMP7 using Allen Brain Atlas**
Co-expression values or similarity co-efficient sorted from highest co-expression to lower co-expression in the regions of the brain (fine annotation) for BMP7 one of the significant *trans* effect and MAPK3 a gene on the 16p11.2 region.

| Region ('big12' annotation) | BMP7 / MAPK3 similarity coefficient |
|---|---|
| 'Retrohippocampal region' | 0.4483 |
| 'Cerebellum' | 0.4468 |
| 'Hypothalamus' | 0.4083 |
| 'Olfactory areas' | 0.3916 |
| 'Hippocampal region' | 0.3903 |
| 'Pallidum' | 0.3888 |
| 'Thalamus' | 0.3616 |
| 'Striatum' | 0.3556 |
| 'Cerebral cortex' | 0.2902 |
| 'Medulla' | 0.2801 |
| 'Midbrain' | 0.2629 |
| 'Pons' | 0.2118 |

**Table 4.9 List of brain tissue (big 12 annotation) with highest co-expression values between MAPK3, and BMP7 using Allen Brain Atlas**
Co-expression values or a similarity co-efficient sorted from highest co-expression to lower co-expression in brain tissue (big 12 annotation) for BMP7 a significant *trans* effect and MAPK3 a gene on the 16p11.2 region.

**DISSCUSSION**

Carriers of the 16p11.2 microdeletion and microdupliction are at an increased risk for the development of a neuropsychiatric phenotype. We used genome wide expression analysis to determine genes outside of the mutation which may be dysregulated due to dosage sensitive genes in the 16p11.2 region. We expect the dosage sensitive genes on the 16p11.2 region would interact with the genes they are dysregulating. Bioinformatic tools were used in our analysis to determine interactions between the 16p11.2 genes and the seven significant *trans* effects.

Pathways analysis, bioinformatic, and co-expression tools found no direct interactions between the seven significant *trans* effects and genes on the 16p11.2 locus. Our analysis did however shed light on networks, possible indirect connections, and interactions through predicted transcription binding sites. It also revealed genes in the two gene sets which were co-expressed with one another.

Ingenuity Pathway Analysis discovered several indirect connections between our two gene sets. The second network (Cellular Growth and

Proliferation, Connective Tissue Development and Function) formed by IPA created an indirect link between BMP7 and MAPK3. Molecules in the TGF-beta super family are involved in growth and proliferation, development, cellular differentiation, organogenesis cancer, and heart disease [101, 103, 104, 126]. It is unclear through IPA as to how BMP7 and MAPK3 interact with each other[103, 126]. We do know from other publications that BMP7 and MAPK3 interact via the BDNF/MAPK cascade [77]. IPA also showed an indirect connection between URB1 and MAPK3. MAPK3 is a gene involved in many signaling, and molecular pathways. In IPA it had thousands of direct and indirect interactions, which increases the probability that it would be indirectly linked to a significant *trans* effect. Caution must be used when analyzing interactions between MAPK3 or any other gene hub. Similar to IPA, InnateDB also links BMP7 and MAPK3 as both being a part of the TBF-beta signaling pathway. Pathway analysis using InnateDB revealed that BMP7 and MAPK3 are both part of the TGF-beta (Transforming Growth Factor) signaling pathway, as BMP7 is a TGF-beta molecule.

InnateDB's *predicted transcription factor interaction* feature found several transcription factors which interact with our gene set. Notably URB1 was found to interact with many transcription factors including AIRE, CREB1, E2F1, NRF1, PDX1, PPARA, SP1, and XBP1. Little is known about URB1, besides that it is ubiquitously expressed across tissues. This coupled with the many predicted transcription binding sites may suggest that the URB1 may be involved in regulation. Some of these transcription factors play a role in cancer, such as E2F1, and TP53. Five genes in the 16p11.2 region (c16orf53, DOC2A, KCTD13, TAOK2, YPEL3) and one significant *trans* effect (URB1) have a predicted transcription factor interaction with nuclear respiratory factor 1 (NRF1). This gene regulates respiration, heme biosynthesis, and the transcription of mitochondrial DNA, and interestingly, neurite outgrowth [127-129]. Neurite outgrowth was shown to be increased in mouse cortical neurons upon expression of NRF1 [129]. Genes associated with neurite outgrowth have been linked to both autism and schizophrenia [130-133]. InnateDB common pathway analysis also revealed that PTGS1 ALDOA, CDIPT, and QPRT play a role in various metabolic pathways.

The UGET a co-expression database was used to discover which genes on the 16p11.2 region had the highest co-expression with the seven significant *trans* effects. UGET revealed that BOLA2, followed by MAZ, and SPN have the highest co-expression value with the seven significant *trans* effects compared to other 16p11.2 genes. BOLA2 is a gene within the 16p11.2 segmental duplication, involved in proliferation. Out of the seven significant *trans* effects BOLA2 has the highest co-expression value with ARL17, and c6orf35. MAZ is a zinc finger protein which binds to c-MYC and has the highest co-expression value with URB1 [134]. SPN has the highest co-expression with PTGS1. It is unclear how these genes are linked.

Our bioinformatic analysis failed to reveal any direct interactions between our two gene sets, the genes in the 16p11.2 CNV and our significant *trans* effects. However we have found several indirect connections between the two gene sets. Genes do not necessarily need to have direct molecular interactions

to create an effect on each other. Certain genes may be needed to induce the formation of certain tissue types which then turn on tissue specific genes, yet those two genes have no molecular interaction [61]. Other genes may affect each other by being a part of the same signaling cascade [135]. Ingenuity Pathway Analysis and other network building programs require a large number of genes (>100) in order to get significant and meaningful results, which we do not have. Additionally MAPK3, a gene with many indirect and direct connections may create a bias due the great number of connections it has. All data bases are also limited to current knowledge in biology which may not represent all that exists in biology. Data bases frequently had no data on putative proteins and pseudogenes such as, LOC124446 a gene found in the 16p11.2 interval. Such genes may be involved in disease pathogenesis, but little data is available for them just yet. The regions of the brain which show co-expression between BMP7 and MAPK3 using Allen Brain Atlas do not correlate with regions of the brain previously associated with autism or schizophrenia.

Using the data generated using pathway analysis, InnateDB, and co-expression analysis we were able to understand more about common pathways, and functions between the 16p11.2 genes and the significant *trans* effects.

## Chapter 5
## Clinical Review of 16p11.2 microduplication and microdeletion carriers

Mary Kusenda [1,2], Shane McCarthy[2], Jonathan Sebat [2]

The literature review of all clinical publications was performed by Mary Kusenda. Hilde Peeters gave us clinical information on three microdeletion patients (UZ Leuven). Mary gathered head circumference data from the AGRE Autism Genetics Resourse Exchange data set. Shane McCarthy, Jonathan Sebat, gathered head circumference data from Tamim Shaikh, (personal communication), Childrens Hospital Pennsylvania (CHOP), Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE), University of Washington,: National Institute of Mental Health (NIMH), Marshfield Clinical Human Genetics (MCHG) WI, Childrens Hospital Boston, and performed the head circumference analysis.

[1] Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[2] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road

Cold Spring Harbor, NY 11724
**Summary:** There is great heterogeneity in symptoms observed in 16p11.2 microdeletion and microduplication carriers [29, 34]. If the neuropsychiatric phenotype observed in these carriers is caused by a mutation that is common amongst these patients, there should be other common phenotypes shared by patients with the same genotype. I hypothesized that there are potentially common developmental abnormalities among carriers, which can be found through analysis of clinical data. After meticulously analyzing medical and clinical data I found nine microdeletion carriers (n=59) and four microduplication carriers (n=27) reported bone abnormalities or scoliosis. Three patients reported severe abnormalities involving the kidney. Knockout mice of one of the significant *trans* effects BMP7, has also been associated with skeletal and kidney abnormalities [61]. We also determined that patients with the microdeletion were associated with a larger head circumference as compared to the average population, and as compared to microduplication carriers [3]. Discovery of these common developmental abnormalities coupled with the understanding of functional and molecular knowledge of the16p11.2 genes, and their significant *trans* effects will shed light on pathways which may be contributing to the developmental phenotype associated with the region.

**Introduction**

The 16p11.2 CNV is associated with multiple neuropsychiatric phenotypes [2, 3, 5]. The microdeletion is found in 0.1% of those with speech and language delay, 0.6% of obese patients, and 0.01% of the general population [3, 4]. The microduplication is associated with 0.3% of schizophrenia, 0.3% of obesity patients, 0.04% of those with speech and language delay, 0.04% of bipolar disorder, and is found in 0.03% of the general population [2-4, 136]. The microdeletion is more penetrant for language delay, autism, and obesity, and is rarer in the general population as compared to the microduplication. This single genetic locus is associated with several psychiatric disorders and phenotypes. Reading over clinical data should reveal phenotypic commonalities which are found among these carriers of the mutation.

Although there is a broad psychiatric phenotype associated with the 16p11.2 microdeletion and microduplication most carriers are reported to have developmental delays in speech and language [29]. These delays vary in severity. A child may begin to be verbal at age two, with phrase speech appearing around age three [137]. Mainly expressive language is affected as opposed to receptive language [137]. Once the patient begins to speak, there is continuous progress [36]. Inability to communicate may contribute to behavioral problems such as aggression and hyperactivity seen in these children [2, 36, 137]. Aggressive behavior and hyperactivity were initially reported to be associated with the mutation [2]. Since then, no clear behavioral pattern has been found unique to these patients.

Cognitive impairment can be found in 16p11.2 deletion patients [137]. Understanding abstract ideas may develop very late [36]. A patients IQ can range

from mildly retarded to normal [137]. Higher than average IQ was observed in a few microdeletion and microduplication cases [137].

Motor development delays are also present, but are not as common as speech and language delays [137]. Two out of three patients with the deletion have motor delay in varying severity [36]. Reports so far have shown that most patients walk, but may lack motor coordination [36]. Fine motor control, such as development of the "pincer grip" to grasp objects also may develop late [36]. A minority of patients also exhibit seizures, although these are well controlled with medication, and decrease in severity with age [29, 36, 79].

Estimates report that one third of those with the deletion have unusual. facial features that don't seem to be common to the syndrome [34, 36, 137]. Currently there is no facial dysmorphism found in 16p11.2 patients which would aid in diagnosis [137]. Common facial dysmorphisms reported in deletion cases include broad forehead, flat midface, hypertelorism, and micrognathia [137], [29, 34, 138]. Congenital abnormalities found in deletion cases include almost all organ systems including congenital heart defects, cleft palate, multicystic dysplastic kidney, pyloric stenosis and fusion of lower ribs. The only known malignancy in a 16p11.2 case was a female diagnosed with Wilms Tumor, (2 years 9 months) who is a 16p11.2 deletion carrier [35]. Deletion cases in particular had structural brain abnormalities which did not show any commonality between patients [29].

Duplication carriers have more severe facial dysmorphisms than deletion carriers, which are not common to carriers. Duplication cases had less severe abnormalities which include cleft lip, pectus excavatum, torticollis, pes planus, hypospadias and mild scoliosis [29]. Duplication carriers don't have as many reports of brain lesions and none that are common to the genotype [29].

Clinical reports which featured 16p11.2 carriers, gave detailed descriptions of their cognitive, neuropsychiatric, and congenital abnormalities. Shimojima *et al* reported a boy with the deletion, mental retardation and developmental delay [59]. This child had a severe phenotype, can not stand without support and has little verbal communication. His physical abnormalities include bone malformations such as hemivertebra of the thoracic vertebra (T10, T12) and lumbar vertebra (L3). The patient is also missing his right twelfth rib, and has hypoplasia of the left twelfth rib. The patient has other abnormalities not involving the skeleton which include has inguinal hernia, and astigmatism [59].

Another patient with a *de novo* 16p11.2 deletion inherited a 2 base pair deletion in CORO1A a gene in the 16p11.2 region, resulting in a premature stop codon. The child had ear infections, upper respirator tract infections, acid reflux, and pneumonia in her first year of life. She developed chickenpox after receiving a cutanious varicella vaccine at 13 months. It was discovered that the child had a decreased number of lymphocyte, and T cells, with a normal count of B, NK cells, and immunoglobulins. She had developed worsening infections until a blood transfusion, cured her. She has a low growth rate, language and motor delays, low to average cognitive function, and inattention [80].

To conclude, 16p11.2 patients vary in severity of phenotype. They have varying degrees of developmental delay, cognitive dysfunction, and congenital

abnormalities. Patients with the microdeletion exhibit a more severe phenotype as compared patients with the microduplication [3, 29, 34, 35, 59]. Several studies have done a clinical review of phenotypes associated with the 16p11.2 locus, but have not included all available sample data. By analyzing clinical information for 16p11.2 cases published in various sources, and from data reported to us from collaborators I will determine if there is a common phenotype found in 16p11.2 patients.

## Materials and Methods

### Acquirement and analysis of clinical phenotype data.

Phenotype data was gathered and analyzed from published manuscripts, posters presented at conferences, and the Unique Guide 2009 a website which has gathered phenotype data on 16p11.2 patients from parents and other family members. (http://www.rarechromo.org/html/home.asp) The sources used in the analysis include Weiss *et al* (2008); Kumar *et al* (2008); Gherbranious *et al* ( 2007); Bijlisma *et al* (2009); Fernandez *et al* (2009); Hernando *et al* (2002); McCarthy *et al* (2009); Sebat *et al* (2007); Shinawi *et al* (2009); Miller *et al* (2009) Unique (2009) and Shiow *et al* (2009). We also used unpublished data from the Autism Genetic Resource Exchange (AGRE) database for known 16p11.2 carriers. From AGRE, data from the patient's medical history, language questionnaire, physical exam, and peabody picture vocabulary test was used. Seven cases presented in a poster American College of Medical Genetics by Sacharow (2009) abstract number 132, were included in our analysis. We also included unpublished phenotypic data from Hilda Peters for three patients with the microdeletion.

All phenotype data was categorized, and tallied. Clinical data from various publications were gathered with the goal of defining common developmental defect which may be affected in 16p11.2 patients. All patients with common anomalies and all phenotype data on those patients were retained in our data. All of the anomalies were then investigated using the Bedside Dysmorphologist textbook and other literature [139]. Common abnormalities found were compared with its prevalence in the general population.

### Acquirement and analysis of head circumference data.

Existing clinical data for height, weight, and head circumference were gathered from previous publications (Weiss *et al* and Gherbanious *et al*) and from unpublished sources (Tamim Shaikh, personal communication, CHOP: Childrens Hospital Pennsylvania; CATIE: Clinical Antipsychotic Trials of Intervention Effectiveness; AGRE: Autism Genetics Resourse Exchange; UW: University of Washington; NIMH: National Institute of Mental Health; MCHG: Marshfield Clinical Human Genetics, WI; CHB : Childrens Hospital Boston). Patients known to be of Hispanic, Polynesisn, and African American descent were removed from our analysis due to insufficient reference data for those populations.

Height data (in cm) was available for 23 patients with the microdeletion and nine with the microduplication. Head circumference is measured near the glabella and near the occipital bone, and is called the Occipital-Frontal

Circumference (OFC) [3]. Head circumference data was available for nine of the microduplication patients and 23 of the microdeletion patients [3] . All data was provided at one time point except for one sample, CG20261 which had data for the ages of 25, and 5. Head circumference, and height measurements were converted to standardized (z) scores categorized by age and gender. The standard craniofacial normative database was used to compute the z scores and define macrocephaly and microcephaly.

**RESULTS:**

I analyzed existing phenotypic data for 59 patients with the 16p11.2 microdeletion, and 27 with the microduplication from various sources with the hope of defining a common phenotype. By reading clinical data on 16p11.2 patients I was able to determine that 7/59 deletion carriers, and 4/27 duplication carriers have a skeletal abnormality. Three microdeletion carriers had kidney abnormalities. We also analyzed head circumference data to discover that individuals with the microdeletion had a larger head circumference, than individuals with the microduplication. There was an abundance of skeletal abnormalities which centered on the ribs, and vertebra found in 16p11.2 carriers. Deletion carriers suffer from more severe abnormalities as compared to those with the duplication. Although not common, three deletion carriers had kidney abnormalities which were severe enough to be reported.

**Skeletal Abnormalities**

One common developmental anomaly found amongst 16p11.2 patients are skeletal abnormalities which presented in variable severity. Microdeletion carriers suffer from more severe abnormalities as compared to those with the duplication. Several microdeletion patients have rib, and vertebral abnormalities, at a higher prevalence rate than reported in the general population. Two deletion carriers have hemivertebrae of T10, a thoracic vertebra [140]. One of the two patients has additional abnormalities including hemivertebra of T12 and Lumbar Vertebra (L3), missing his right twelfth rib, and has hypoplasia of the left twelfth rib [59]. One microdeletion patient is reported to have fusion of the lower ribs, parietal foramina, along with speech delay, and a low normal IQ [29]. Another microdeletion patient had vertebral abnormalities, along with mild developmental delay [141]. Three microduplication cases and one deletion case had scoliosis [29, 34, 35]. Two of the microduplication cases with scoliosis also had pes planus or flat feet [29, 34, 35]. Micrognathia was also found in two microdeletion and one microduplication case [29]. One microdeletion patient has polydactyly along with micrognathia [29] [142]. Our genome-wide expression study contains data from lymphoblast cell lines on one of the samples which had Hemivertebra. This child had the second lowest expression level for BMP7 in all our collected data. However we did not use this child in the final genome-wide expression SAM analysis because the sample's expression data had more variation than average.

**Kidney Abnormalities**

Three 16p11.2 microdeletion cases had kidney abnormalities. One has multicystic dysplastic kidney. One infant died shortly after birth due to severe respiratory distress. Upon ultrasound it was discovered that the child did not have

any kidneys. This infant also had minor anomalies and heart disease [141]. The only malignancy ever reported in a 16p11.2 case was Wilms tumor, a tumor of the kidney [35]. Wilms Tumor 1 is a tumor suppressor involved in the developing kidney. Patients with a mutation or reduced expression of Wilms Tumor 1 have a higher susceptibility to develop unilateral or bilateral tumor of the kidney. The child with the microdeletion and Wilms Tumor made us hypothesis that expression of Wilms Tumor 1 gene may show correlation with genotype. Wilms Tumor did not reach statistical significance in our SAM analysis, but ranked number 421 out of 56,000 by its SAM score. Plotting the expression value against genotype showed a slight correlation with genotype. The expression of the probe was also low, and the correlation may be due to noise.

**Head Circumference**

We investigated whether the 32 microdeletion and microduplication carriers were associated with head circumference. The average OCF values were 1.25 for microdeletion, and -.28 for the microduplication which reached statistical significance. (two tailed Wilcoxin rank sum test (p=0.0007)). Deletion individuals also have a greater head circumference than the average population (P=0.0001). This was not observed in patients with the microduplication (p=0.29) [3].

| Source | microdeletion | microduplication |
|---|---|---|
| Shinawi et al (2009) | 16 | 10 |
| Sacharaw (2009) | 7 | 0 |
| Fernandez et al (2009) | 3 | 3 |
| Sebat et al (2007) | 1 | 0 |
| Weiss et all (2008) | 8 | 4 |
| Gherbranious et al ( 2007) | 2 | 0 |
| Bijlisma et al (2009) | 14 | 0 |
| Peeters (2009) | 3 | 0 |
| AGRE data base | 5 | 10 |

**Table 5.1 Sources of clinical data for 16p11.2 carriers.**
Total number of 16p11.2 carriers with clinical phenotype data from various sources, 59 patients with the microdeletion, 27 with the microduplication. (86 total)

Table A

| | CASE A | CASE B | CASE C | CASE D | CASE E | CASE F | CASE G |
|---|---|---|---|---|---|---|---|
| Reference | Shimojima et al. (2009) | Fernandez et al (2009) | Shinawi et al (2009) | Sacharow (2009) ACMG poster #132 | Bijlsma et al. (2009) | Shinawi et al (2009) | Shinawi et al (2009) |
| 16p11.2 | Deletion | Deletion | Deletion | Deletion | Deletion | Deletion | Deletion |
| Anomaly | Hemivertabrae of T10, T12, L3, Missing twelfth rib, hypoplasia of left 12th rib | Hemivertabrae of T10 | Parietal formamina, fusion of lower ribs, | Vertebral abnormalities | Scoliosis, growth retardation | Polydactyly, Mild Micrognathia | Micrognathia |
| Autism | No | YES | No | No | No | NA | Yes |
| Dev. Delays | Non verbal | Language regression | speech delay | Mild developmental delay | psycomotor delays | Mild motor and speach delay | Motor and speech delay, MR |
| Other | Astigmatism, Microcephaly | Severly Dysmorphic | patent foramen ovale, (PFO) Abn heart rhythm (resolved) | cleft lip and palate, Chiari I malformation, hyperpigmented macules | Clinodactyly of fingers. | Hypertelorism, PFO, PDA, congenital diaphragmatic hernia, cleft plate | MRI showed prominence of extra CSF space |

Table B

| Case A | CASE B | CASE C | CASE D |
|---|---|---|---|
| Shinawi et al (2009) | Fernandez et al (2009) | Shinawi et al (2009) | Shinawi et al (2009) |
| Duplication | Duplication | Duplication | Duplication |
| Mild Scoliosis, pes planus | Scoliosis | Micrognathia | Pes Planus, Phagiocephaly, telecanthus, mild scoliosis |
| No | YES | No | No |
| Mild stutter, MR | Language regression | Mild motor delay, Mild MR, speech delay. | Mild motor and speech delay, stutters |
| frontal bossing, prominent ears, deepset eyes | Hypertelorism | hypertelorism, facial asymettry | Asthma |

**Table 5.2 16p11.2 carriers with reported skeletal abnormalities**
Out of the 59 patients with the microdeletion, and 27 with the microduplication 7 deletion carriers, and 4 duplication carriers have some sort of bone abnormality. The table is split into two parts. Part A for deletion carriers and part B for duplication carriers. It includes descriptions of the bone abnormality as described from various publications, and sources. Along with this information is whether the carrier has been diagnosed with autism, a description of the patient's developmental delays, and other congenital abnormalities not involving bone structure. Patients with the deletion showed more severe abnormalities as compared to those with the duplication.

| | CASE A | Case B | Case C |
|---|---|---|---|
| Reference | Shinawi et al 2009 | Sacharow (2009) ACMG poster #132 | Bijisma (2009) |
| 16p11.2 | Deletion | Deletion | Deletion |
| Anomaly | Multicystic dysplastic Kidney | Ultrasound showed No kidneys | Wilms tumor |
| Autism | too young to be tested | N/a | No |
| Dev. Delays | Motor Delay | N/a | None |
| Other | overfolded helicies, Mildly prominent 3rd and lateral ventricles, seizures | Severe respiratory distress. Dies shortly after birth | Muscular Hypotonia |

**Table 5.3 16p11.2 microdeletion carriers with reported kidney abnormalities**
This table features three deletion cases which have severe abnormalities of the kidney. It also includes a description of developmental delays, and other abnormalities.

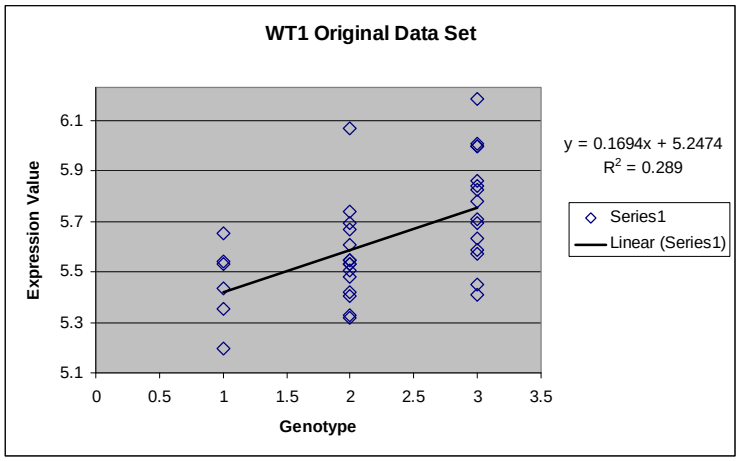



**Figure 5.1 Expression data for Wilms Tumor 1.**
This is a graph from the genome wide expression analysis on the lymphoblast cell lines from 16p11.2 carriers for the gene Wilms Tumor 1 (WT1). (Probe 216953_s_at) WT1 was 421 out of 56,000 probes. Expression level is plotted on the y axis, and genotype is on the x axis. There is a slight positive correlation of expression of WT1 with genotype.

| Study | Patient | Source[a] | 16p11.2. Genotype | Gender | Ethnicity[b] | Primary Diagnosis[c] | Age at Examination | WISC Performance IQ | Ravens IQ | Estimated IQ | Weight (kg) | Height (cm) | Height Z-score | OFC (cm) | OFC Z-Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Referal | CHOP 1 | CHOP | Deletion | Female | EA | DD | 3.25 | | | | 12.40 | 86.30 | -2.59 | 50 | -0.17 |
| Referal | CHOP 3 | CHOP | Deletion | Male | EA | DD | 19 | | | | | | | 58.5 | 0.42 |
| Referal | CHOP 4 | CHOP | Deletion | Male | EA | DD | 14.75 | | | | 74.40 | 178.70 | 2.17 | 58 | 1.99 |
| Current | 03C15581 | CATIE | Duplication | Male | EA | SCZ | 62 | | | | 99.79 | 185.42 | 1.09 | | |
| Current | 03C15896 | CATIE | Duplication | Male | EA | SCZ | 44 | | | | 58.06 | 182.88 | 0.78 | | |
| Current | 03C18520 | CATIE | Deletion | Male | EA | SCZ | 23 | | | | 96.16 | 175.26 | -0.17 | | |
| Current | AU002903 | AGRE | Duplication | Male | EA | ASD | 15.46 | | 90 | | 69.40 | 179.07 | 0.86 | 54.8 | -0.99 |
| Current | AU002904 | AGRE | Duplication | Female | EA | - | 13.36 | | 94 | | 49.44 | 156.21 | -0.13 | 53.3 | -0.35 |
| Current | AU002905 | AGRE | Duplication | Female | EA | ASD | 11.94 | | | | 39.46 | 153.67 | 0.97 | 53.4 | 0.26 |
| Current | AU011004 | AGRE | Duplication | Male | EA | ASD | 11.04 | | 110 | | | | | 53.5 | -0.14 |
| Current | AU032704 | AGRE | Duplication | Male | EA | ASD | | | 75 | | | | | | |
| Current | AU032705 | AGRE | Duplication | Male | EA | ASD | | | 50 | | | | | | |
| Current | AU032706 | AGRE | Duplication | Male | EA | ASD | | | 107 | | | | | | |
| Current | AU032707 | AGRE | Duplication | Male | EA | - | | | 107 | | | | | | |
| Current | AU041905 | AGRE | Deletion | Male | EA | ASD | 7.96 | | 108 | | 24.49 | 124.46 | -0.02 | 54 | 1.32 |
| Current | AU0938301 | AGRE | Deletion | Male | EA | ASD | 14 | | | | | | | 56.5 | 1.06 |
| Current | CG20261 | Columbia | Deletion | Female | EA | ASD | 5 | | | | | 157.48 | -0.76 | 51.5 | -0.18 |
| Current | MC235 | UW | Duplication | Male | EA | SCZ | | | | 80-90 | | | | | |
| Current | Rap-2011 | NIMH | Duplication | Female | EA | SCZ | 15 | 81 | | | 48.10 | 152.40 | -1.44 | 54 | -0.13 |
| Current | Rap-676 | NIMH | Duplication | Female | EA | SCZ | 13 | 72 | | | 34.50 | 157.50 | 0.02 | 52.5 | -0.94 |
| Ghebranious et al | Twin1 | MCHG | Deletion | Male | - | DD | 28 | | | | | | | 59.5 | 1.11 |
| Ghebranious et al | Twin2 | MCHG | Deletion | Male | - | DD | 28 | | | | | | | 57.3 | -0.42 |
| Weiss et al | CHBDel1 | CHB | Deletion | Male | - | DD/MR/ASD | 6y6m | | | | 19.00 | 108.10 | -1.65 | 51 | -0.60 |
| Weiss et al | CHBDel2 | CHB | Deletion | Male | - | DD/MR/ASD | 2y9m | | | | | | | 52 | 1.32 |
| Weiss et al | CHBDel3 | CHB | Deletion | Male | - | DD/MR/ASD | 17m | | | | 9.00 | 74.20 | -1.42 | 48.5 | -0.53 |
| Weiss et al | CHBDel4 | CHB | Deletion | Male | - | DD/MR/ASD | 9y2m | | | | 65.40 | 143.30 | 0.89 | 54.5 | 1.17 |
| Weiss et al | CHBDel5 | CHB | Deletion | Male | - | DD/MR/ASD | 9y2m | | | | 71.90 | 150.00 | 1.73 | 56 | 2.25 |
| Weiss et al | CHBDup1 | CHB | Duplication | Male | - | DD/MR/ASD | 14m | | | | 11.00 | | | | 0.84 |
| Weiss et al | CHBDup2 | CHB | Duplication | Female | - | DD/MR/ASD | 3y3m | | | | 14.70 | 97.20 | -0.45 | 46.5 | -2.91 |
| Weiss et al | CHBDup4 | CHB | Duplication | Female | - | DD/MR/ASD | 9y9m | | | | 33.50 | | | 51 | -0.90 |
| Weiss et al | ICEDel1 | Iceland | Deletion | Female | EA | ASD/MR | 5y2m | | | | 18.00 | 106.00 | -1.87 | 52.5 | 1.71 |
| Weiss et al | ICEDel2 | Iceland | Deletion | Male | EA | Atypical Autism | 10y6m | | | | 69.00 | 152.00 | 1.76 | 54 | 0.38 |
| Bijlsma | Case 1 | - | Deletion | Male | - | DD | 44 | | | | 172 | 84.9 | -0.57 | | |
| Bijlsma | Case 2 | - | Deletion | Male | - | DD | 17yr 2m | | | | 180 | 130.5 | 0.52 | 60 | 3.81 |
| Bijlsma | Case 3 | - | Deletion | Female | - | DD | 8yr 2m | | | | 130 | 45 | 0.17 | 54.5 | 1.90 |
| Bijlsma | Case 4 | - | Deletion | Male | - | ADHD | 13yr | | | | | | 2.00 | | |
| Bijlsma | Case 5 | - | Deletion | Male | - | ASD | 3yr | | | | | | 0.00 | | -0.80 |
| Bijlsma | Case 6 | - | Deletion | Female | - | DD | 7yr | | | | 122 | 25 | -0.28 | 52.5 | 0.67 |
| Bijlsma | Case 7 | - | Deletion | Male | - | DD | 18m | | | | 72 | 8.2 | -1.85 | 47.5 | -1.43 |
| Bijlsma | Case 8 | - | Deletion | Female | - | DD | 11yr | | | | 146 | 42.8 | -0.04 | | |
| Bijlsma | Case 10 | - | Deletion | Female | - | DD | 8yr | | | | 128 | 24 | -0.14 | 53 | 0.85 |
| Bijlsma | Case 11 | - | Deletion | Male | - | DD | 4yr | | | | 101.5 | 15 | -1.67 | 51.5 | -0.23 |
| Bijlsma | Case 12 | - | Deletion | Female | - | Dysmorphism | 11m | | | | 77.5 | 8.4 | 1.79 | 45.5 | 0.21 |
| Bijlsma | Case 13 | - | Deletion | Male | - | DD | 4yr 6m | | | | 112 | 21 | 0.47 | | |

**Table 5.4 Head circumference data from various publications**

Table 5.5 contains data collected from various publications (Gherbranious *et al*, Weiss *et al*, Bijisma *et al*) and other sources[a]. Available data on IQ, weight, height, and OFC is presented.

[a] CHOP: Childrens Hospital Pennsylvania; CATIE: Clinical Antipsychotic Trials of Intervention Effectiveness; AGRE: Autism Genetics Resource Exchange; UW: University of Washington; NIMH: National Institute of Mental Health; MCHG: Marshfield Clinical Human Genetics, WI; CHB : Children's Hospital Boston

[b] European Ancestry (EA); Ancestry Unknown (-)

c Autism Spectrum Disorder (ASD): Global Developmental Delay (DD); schizophrenia (SCZ); Mental Retardation (MR); No diagnosis (-).

Table taken from: McCarthy *et al*. Microduplications of 16p11.2 are associated with schizophrenia Nature Genetics 41, 1223 - 1227 (2009) Published online: 25 October 2009 | doi:10.1038/ng.474

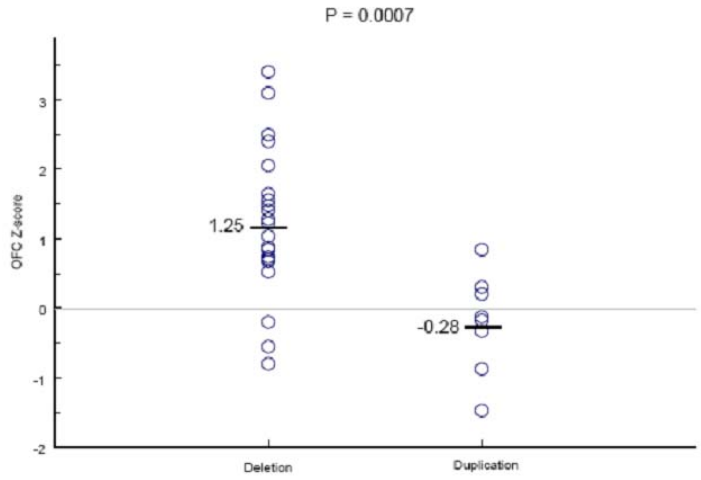



**Figure 5.2 Head circumference and genotype**
Here are the Z scores of Occipital-Frontal Circumference (OFC). On the left are z-scores for all individuals with the deletion, and on the right are all individuals with the duplication. The mean head circuference z score for deletions was 1.25, and the mean z-score for duplications was -0.28.

**Discussion**

Clinical phenotype data from 16p11.2 carriers were analyzed with the intention of defining common phenotypes among patients as carriers with the same genetic mutation (microdeletion or microduplication) should share common phenotypes. Both microdeletion and microduplication patients have skeletal abnormalities at a higher than expected frequency. Three microduplication patients were observed to have severe kidney abnormalities. Research into pathways and genes involved in rib, vertebral, and kidney development, along with our knowledge of dysregulated genes found in 16p11.2 carriers can shed light on pathways which may be responsible for the common phenotype seen in these patients.

**Congenital Skeletal Abnormalities in 16p11.2 cases:**

Clinical data on 16p11.2 carriers determined out of 13 of the 86 carriers have some version of bone abnormalities. This common phenotypic anomaly could give us a clue in elucidating the molecular mechanisms behind this abnormality, and ultimately insight into which genes on the 16p11.2 region could be contributing to the neurodevelopmental phenotype. Several patients have rib, and vertebral abnormalities, at a higher prevalence rate than reported in the general population. Especially of interest were the two separate patients which had hemivertebrae of T10, a thoracic vertebrae [34, 59]. Hemivertabrae is a malformation of the developing vertebra. and is the most common cause of congenital scoliosis [140]. Hemivertebra exists in the general population at a low frequency (5-10 in 10,000), and for both patients to have the same vertebra affected is astounding. This abnormality may suggest a developmental defect

which may be caused by a dosage effect resulting from the 16p11.2 region. Four cases in our clinical review also exhibited scoliosis (3-5 in 1000 in general population). One deletion carrier had fused lower ribs which is reported to be rare in the general population [139]. That same carrier also had patent foramen ovale (PFO) (25% of population) or a hole between the left and right atria of the heart, and parietal foramina [139]. Parietal foramina is a defect in the skull bone with a prevalence of 1:15000 to 1:25000 [143]. Micrognathia, an underdeveloped mandible, (1 in 4000 births) was also common among patients, and is found in the two deletions, and one duplication carrier [144]. Micrognathia is commonly associated with genetic syndromes, such as trisomy 18, Pierre Robin syndrome, and DiGeorge Syndrome [139, 145, 146]. Interestingly DiGeorge syndrome is associated with a CNV on chromosome 22, which contains MAPK1, a gene which forms a complex with MAPK3 a gene on the 16p11.2 region. One of the patients which had micrognathia also had polydactyly (1.73 in 1,000 births) [142]. Polydactyly is a condition where the individual has more than five fingers or toes, and is common in Greig syndrome, acrocallosal syndrome, and oro-facio-digital (OFD) syndromes [139]. Caregivers describe dental abnormalities are found in three in five 16p11.2 microdeletion cases [36].

By further studying genes involved in the development of skeletal formation, particularly of the ribs and vertebrae, we can elucidate which gene or pathway may be involved in skeletal abnormalities seen in 16p11.2 patients. Many genes in the Bone Morphogenic Protein (BMP) family are involved in early development. Genes in the BMP family play a role in formation of bone, and cartilage, but also play a role in neuronal and epidermal development [101, 105, 147]. The BMP pathway has been associated with formation of the skull bones associated with parietal foramina, an abnormality seen in one microdeletion carrier [148, 149]. Mutant mice for Msx2 a target of the BMP pathway, are shown to have skull abnormalities which may be related to the pathogenesis of parietal foramina [149]. BMP4, a gene in the BMP family is involved in mandible formation and micrognathia a condition seen in four 16p11.2 carriers [150].

Bone Morphogenic Protein 7 (BMP7), a significant *trans* effect ranked number 35, out of 56,000 probes, and is a gene involved in early development. This gene is a prime candidate for further study.

**BMP7- a significant *trans* effect from our genome wide expression analysis**

BMP7 belongs to the transforming growth factor (TGF-beta) super family of secreted signaling molecules which play several roles in embryonic development. There is crosstalk between BMPs with Wnt/ B catenin signaling, ISmads, Smads, Smurfs, and other Smad interacting molecules [101, 104]. Several of the BMPs have redundancies and can be compensated for each other during initial stages of development. This can sometimes explain the variable phenotype associated with animal models of BMP gene knockouts [147].

BMP7 is a gene shown to be involved in early development of the skeleton, limbs, eye, and kidney in the mouse. BMP7 expression can be detected in the heart, bone, brain, kidney, and eye. BMP7 heterozygous mice are normal while BMP7 knockout mice show defects in eye, skeletal, and kidney

development [61, 62, 151]. Ultimately BMP7 knockout mice have very small kidneys, and die due to renal failure. Polymorphisms found in BMP7 have not been associated with any human disease. (Online Inheritance in Man (OMIM) database)

**Skeletal abnormalities found in Bmp7 null mutant mice**.

BMP7 plays a role in bone ossification, and skeletal patterning during development. Bmp7 heterozygous mice do not display skeletal abnormalities. Bmp7 null mice exhibit specific skeletal abnormalities limited to the skull, hind limb polydactyly, and ribs [62]. Mice prevalently exhibited skull abnormalities which included smaller or fusion of the basosphenoid bones, and reduction of the pterygoid bones [62]. Rib cage abnormalities included asymmetric rib pairing, rib fusion, missing ribs, reduction of posterior pairs of ribs, xiphoid process malformations were seen in varying severity in these mice [61, 62]. Interestingly patients with the 16p11.2 mutation have exhibited similar rib, and vertebral abnormalities, such as fusion or the ribs, vertebral abnormalities, and hemivertebra. Bone mineralization was also reported to be slower in Bmp7 newborn null mice in vertebra, proximal ribs, and bones of the leg, including the fibula, tibia, and femur [61]. No difference in bone mineralization was seen in 12 day old knockout mice [61].

Unusual dental disruptions or development, and micrognathia has been reported by family member of children with the 16p11.2 deletion [29, 36]. BMP7 is involved in dental development. In the mouse, Bmp7 expression becomes progressively stronger in during odontoblast differentiation bell stage of the first molar tooth germs, incisor, second molar, surrounding osteogenic area, whiskers, and hair follicles. Dental tissues develop normally which may be due to maternal transfer of the BMP7, or functional redundancy of other BMPs [106].

The most prevalent Bmp7 null mice skeletal defect was hind limb polydactyly (82%), a sixth digit in the foot [62]. One of the 16p11.2 deletion patients is recorded to have polydactyly [29]. 2-3 syndactyly has been observed in a handful of 16p11.2 cases, but this same observation was seen in siblings without the 16p11.2 mutation, and this commonalty was disregarded.

**Kidney abnormalities found in Bmp7 null mutant mice**.

BMP7 homozygous null mice die shortly after birth due to severe kidney underdevelopment [62]. Knockout mice have small kidneys with little development in the glomerulous in the cortical region. Unilateral or bilateral polycystic kidneys are observed in newborn Bmp7 knockout mice. Any mutant mice which survive after birth develop bilateral or unilateral kidneys [61]. Wilms tumor suppressor (WT1) an important gene involved in early kidney development and is expressed in podocytes of mature glomeruli (day 14.5 dpc), and metanephric mesenchymal cells of the kideny (12.5 dpc). BMP7 is important for the induction of a tissue type in the kidney, the metanephric mesenchyme cells, which is necessary for induction of WT1. As a result of the reduced expression of BMP7, WT1 expression is greatly reduced or absent in the cortical region and ureteric, where the metanphric mesenchyme should have been formed [62].

**Congenital kidney abnormalities observed in 16p11.2 cases.**

Although not significantly frequent, severe congenital abnormalities of the kidney were seen in three microdeletion carriers. One particularly striking case was of an infant with the microdeletion, who did not have any kidneys, and died shortly after birth. Another case had multicystic dysplastic kidney, in which cysts develop in the kidney. It is the most common cause of abdominal tumors in children (1:2400 to 1:4300 prevalence rate in general population) [152]. Another deletion carrier had Wilms tumor, a rare tumor of the kidney (1 in 10,000 US CDC), and to date is the only malignancy reported in a patient with a known 16p11.2 variation. Wilms tumor suppressor 1, located on the short arm of chromosome 11, is the gene associated with Wilms Tumor. Individuals with a mutation, haploinsufficiency, or decreased expression in this tumor supressor have a higher susceptibility to developing unilateral or bilateral kidney tumors [153-156]. In our genome-wide expression study there was a slight correlation in expression with 16p11.2 genotype in one microarray probe for Wilms Tumor suppressor 1. This association did not reach statistical significance.

Interestingly expression of WT1 is reduced in the developing kidney in BMP7 null mice as a result of decreased expression of BMP7. In our study BMP7 expression is reduced in patients with the deletion. Is the reduction of BMP7 causing WT1 expression to be decreased in 16p11.2 patients, resulting in this one 16p11.2 microdeletion carrier to develop Wilms Tumor? Unfortunately our genome wide expression study did not include this case, so we can not determine the expression level of either BMP7, or WT1 in this patient. The number of 16p11.2 cases which have kidney abnormalities, is very few, but those which exist are severe. It can not be concluded whether BMP7, or genes in the 16p11.2 region are responsible for these congenital abnormalities, but the similarities between these cases and BMP7 mice merit further investigation. Any conclusions can not be made as to whether the 16p11.2 deletion was in fact contributing to this congenital abnormality, and more cases have to be brought to light in order to make any assumption.

**Bmp7 and brain development**

BMP7 has been well characterized in its mechanism in limb formation, during embryogenesis. The pathway by which BMP7 plays a role in the brain is not as well understood. BMP7 is involved in early brain development, and in recovery from brain injury. In several in vitro studies BMP7 has been shown to enhance dendrite growth in cortical, sympathetic and hippocampal neurons. This action was specific to dendrite length, and not axon length. Dendrite outgrowth is specific to BMP7, as similar results were not attained using BMP2 or BMP7 two other molecules in the BMP family [77, 108, 109]. During brain development, BMP7 is expressed by the roof plate and acts as a heterodimer with growth/ differentiation factor 7 (GDF7) to play a direct role in axon guidance. Neuronal cell death has been shown to be ameliorated in upon addition of BMP7 in a dosage depended manner in both in-vivo and in-vitro experiments [100, 110].

A recent finding shows that BMP7 is induced by Brain Derived Growth Factor (BDNF) through the BDNF/MAPK/ERK pathway. Upon in-vivo expression of BDNF BMP7 was the only up-regulated gene from the TGF-beta family at day 14 in the mouse embryo. BDNF induces the MAPK/ERK signaling cascade. ERK1/2, made from a complex of ERK1 (MAPK3) and ERK2(MAPK1) phosphorylate p53/p63/p73. MAPK3 is a kinase within the 16p11.2 region. This phosphorylation negatively regulated p53/p63/p73, which activates the transcription of BMP7 [77].

Dysregulation of BDNF induced transcription of BMP7 causes brain abnormalities to occur. As a result of the increased BMP7 expression in the brain, radial glial cells were shown to prematurely develop into astrocytes which lead to abnormal radial migration of pyramidal neurons [77]. This causes the intricate layering of cells in the cortex to be altered.

**Links between BMP7 null mice and congenital abnormalities in 16p11.2 carriers.**

We have provided evidence for a dosage depended effect of BMP7 from our genome-wide expression study using lymphoblast cell lines from 16p11.2 carriers. There is an overrepresentation of skeletal defects in 16p11.2 patients, and certain types of severe kidney abnormalities. These skeletal defects and kidney abnormalities are similar to ones see in BMP7 mutant mice. Since the congenital abnormalities seen in 16p11.2 carriers are so similar to those seen in Bmp7 null mice, and BMP7 is dysregulated in 16p11.2 carriers we can hypothesize that BMP7 may be contributing to the congenital abnormalities seen in 16p.11.2 carriers.

**Problems with Gathering Phenotypic data**

Finally to conclude, gathering phenotypic data was very challenging. Phenotypic data differs for each patient, by site, publication, and by who was gathering phenotypic data. Recently more phenotypic data has been published on cases with the 16p11.2 mutation, as interest in the region has increased. Data for all patients is not complete for all categories.

Existing clinical data on 16p11.2 microdeletion and microduplication patients had to have been analyzed with some caution. Clinical dysmorphisms, such as facial features, sometimes are familial rather than due to the mutation.[2, 79] In addition to the lack of existing data, clinical data on siblings related to the proband were also limited. Data on siblings proved very valuable, and helped in elucidating which clinical features were specific to the proband with the mutation.

# CHAPTER 6

# Conclusion

**Disruption of the BDNF/MAPK/ERK pathway, which involves a gene on 16p11.2 CNV may lead to dysregulation of BMP7, a significant gene in our study.**

The 16p11.2 microdeletion and microduplication has been associated with neuropsychiatric disease [2, 3]. Deletion carriers are at an increase risk for developing autism and developmental delay, while carriers of the reciprocal duplication are at an increased risk for developing schizophrenia, bipolar disorder and developmental delay [2, 3, 157]. We wanted to determine dysregulated genes resulting from the alteration in genomic dosage at the 16p11.2 locus. This would reveal disrupted pathways which could point us to a clue as to which gene on the 16p11.2 region is mediating the dysregulation, and ultimately contributing to the neurodevelopmental phenotype associated with the CNV. Using the lymphoblast cell lines from cases (copy number 1, and 3) and controls (copy number 2) we have analyzed the genome wide expression profile and determined a set of seven genes outside of the 16p11.2 region which dosage positively correlates with copy number. BMP7 plays a role in during early embryogenesis, in brain development and plays a neuro-protective role during brain injury, and is a great candidate gene from our study [61, 77, 100, 111].

A group in Cold Spring Harbor Laboratory led by Alea Mills created a 16p11.2 mouse model which they used to look at gene expression using mouse brains with a copy number 1, 2 and 3. Two significantly dysregulated genes from their study overlapped between our study and their study, one of which was BMP7. It is very encouraging that there is some concordance between human lymphocyte and mouse brain expression data.

The BMP7 Affymetrix probe's sequence which was significant in our analysis was unique to one specific human isoform. Using a gene enrichment analyzer it was found that this specific probe had enriched expression for a region in the brain, the trigeminal ganglion. This is not to say that there is no expression of this Affymetrix probe in lymphocytes, the tissue used in our genome-wide expression experiment, as there is. This specific probe has enriched expression for the trigeminal ganglion as compared to all other tissues. The other two BMP7 probes which were not significant in our genome wide expression data have enriched expression in other tissues (thyroid). The trigeminal ganglion is a nerve bundle responsible for innervation and feeling in the face. The observation that the BMP7 probe which was significant in our genome-wide expression analysis has enrichment in the brain is interesting to our study on a variant associated with an increased risk for neuropsychiatric disease.

BMP7 is a gene in the TGF-beta family, is involved in development [101]. In the mouse BMP7 is involved in skeletal, kidney, and eye development [61]. BMP7 knockout mice have rib, vertebral abnormalities, and polydactyly at varying penetrance. The kidneys of these animals were underdeveloped, and a key

gene Wilms Tumor 1, a tumor suppressor was dysregulated as a result of decreased BMP7 expression [61]. 16p11.2 carriers tend to have skeletal abnormalities at an increased frequency that expected. (See figure 5.2) Skeletal abnormalities mimic those seen in BMP7 mouse models, and include fused ribs, hemivertebra, vertebral abnormalities, and polydactyly [29, 34, 59, 141]. The only malignancy every reported in a 16p11.2 carrier was Wilms tumor. It was found in child at 2 years 9 months with the 16p11.2 deletion [35].

BMP7 has been associated with MAPK3 a gene on the 16p11.2 region. Brain Derived Neurotropic Factor (BDNF) activates the MAPK/ERK signaling cascade which results in BMP7 expression in embryonic neurons. ERK1/2(MAPK3/MAPK1) induces expression by phosphorylating p53. After p53/p63/p73 is phosphorylated it stops repressing BMP7 which increases BMP7 expression. p53/p63/p73 has been previously predicted to bind to BMP7. In an in-vivo experiment observing a developing mouse at midgestation, an increase in BMP7 expression had been shown to causes havoc in the developing brain. Cortical neurons turn into astrocytes prematurely, and glutamatergic neurons which were migrating to upper layers of the cortical layers stop prematurely [77]. Borna disease virus (BDV) acts by impairing the BDNF pathway through inhibited phosphorylation of ERK1/2 (MAPK3/1) [119]. Borna virus has been used as a rat model for autism [118].

BDNF has been well characterized in both autism and schizophrenia and other psychiatric illnesses [113-115, 158-161]. There have been contradicting reports as to the increase or decrease of BDNF in serum levels, lymphocytes, and brain in autism and schizophrenia patients [113-115, 158, 160, 161]. A study did show a significant decrease of BDNF in serum levels in autism patients[113]. Reciprocally BDNF protein was reported to be increased in the hippocampus and anterior cingulate cortex in schizophrenia patients[114].

Basal levels of BDNF have been found in lymphoblast cell lines, the tissue type used in our study [162]. In light of the association of BMP7 and MAPK3 by induction by BDNF, we can hypothesize that maybe an increase or decrease of MAPK3 in the developing embryo may cause BMP7 expression to be dysregulated contributing to abnormal brain development. BMP7 has also been shown to be indirectly linked through our bioinformatic analysis using Ingenuity Pathway Analysis (See figure 4.2). BMP7 and MAPK3 were also found to be in the same pathway using InnateDB (See figure 4.5).

The RAS/MAPK pathway has already been associated with developmental syndromes. A class of syndromes called the “Rasopathies” are associated with mutations in genes in the RAS/MAPK pathway The Neuro-cardio-facial-cutanious syndromes (NCFC) include Noonan, Costello, LEOPARD [70]. NCFC have been associated with mutations in activation of ERK1/2. These patients exhibit a high rate of cognitive and psychiatric disease. Patients with a 1MB deletion in 22q which includes MAPK1 have neurodevelopmental deficits, and

macrocephaly [163]. This shows mutations, or haploinsufficiency of genes in the RAS/MAPK pathway contribute to developmental, and psychiatric disease [70].

Together, the fact that BMP7 correlates with dosage in the 16p11.2 locus, and its dysregulation through the BDNF/MAPK/ERK pathways causes abnormal brain development we can hypothesis that MAPK3 a gene found in the 16p11.2 region is contributing to the neurodevelopmental phenotype associated with the region. The theory that MAPK3 is causing the dysregulation of BMP7 in lymphoblast cell lines can be tested in different ways. We have attempted to use shRNA constructs in normal lymphoblast cell lines to knockdown the expression of various 16p11.2 genes. MAPK3 was one of the genes tested. Unfortunately the transfection efficiency of lymphoblast cell lines is very low, and we were unsuccessful in our attempts. Borna virus, which has been used to develop a autism animal model, is known to block ERK1/2 (MAPK3/1) phosphorylation, causing abnormalities in neuronal activity and synaptogenesis [119, 120]. We can use this system to test if BMP7 expression is also dysregulated due to ERK1/2 inhibition.

What about the other significant *trans* effects and their plausible effect on the neuropsychiatric phenotype associated with the 16p11.2 region genes? There were other significant *trans* effects, which were indirectly linked to 16p11.2 genes through our bioinformatic analysis (Chapter 4). There was a draw back to out pathway analysis. Pathway analysis is known to work best when the number of genes analyzed is large (>100), which we did not have. However, the known function of the other significant *trans* effects did not play as great a role in development as BMP7 does.

Finally, if it is proven that 16p11.2 carriers have a dysregulation of the BDNF/MAPK/ERK pathway as a result of MAPK3, many drugs already exist on the market which cause a decrease in BDNF such as haloperidol and risperidone [164]. Giving drugs which target the BDNF/MAPK/ERK pathway to 16p11.2 carriers in an effort to alleviate symptoms, could be a form of personalized medicine in the future.

## Supplementary Chapter A.

## Determining whether TBX6, a gene on the 16p11.2 region is contributing to the dysregulation of BMP7, a significant *trans* effect.

Mary Kusenda[1,2], Deborah Chapman[3] Pascal Grange[2],

Tbx6 knockout mice were engineered by Deborah Chapman. The Chapman lab generously donated Tbx6 knockout mice, and controls to our study. RNA extraction and RT-PCR of Tbx6 knockout mice, and bioinformatic analysis was performed by Mary Kusenda. Pascal Grange analyzed the co-expression of BMP7 with TBX6 using The Allen Brain Atlas.

[1]Program in Genetics
Stony Brook University
Stony Brook, NY 11794

[2] Cold Spring Harbor Laboratory
Watson School of Biological Sciences
1 Bungtown Road
Cold Spring Harbor, NY 11724

[3]Department of Biological Sciences
University of Pittsburgh
101 Life Sciences Annex
Fifth & Ruskin Avenues
Pittsburgh, PA 1526

**Summary:** Our genome wide expression analysis implicates BMP7, a gene involved in BMP signaling, as a dysregulated gene in our study. This is interesting in light of the skeletal abnormalities observed in 16p11.2 carriers [29, 34, 59, 136, 141]. We wanted to test whether BMP7 was under control of TBX6 a transcription factor in the 16p11.2 region which is associated with similar bone malformations [60, 165, 166]. I performed RT-PCR to determine if there is reduced expression of Bmp7 in Tbx6 knockout mice. Expression of Bmp7 is not significantly reduced in Tbx6 null mice, as compared to controls. Additionally we were not able to discover a possible transcription binding site for TBX6 on BMP7. It is safe to assume that TBX6, a gene on the 16p11.2 region is not contributing to the dysregulation of BMP7, a dysregulated gene in the lymphoblasts of 16p11.2 carriers.

**Introduction:**

We conducted a genome wide expression analysis using lymphoblast cell lines from 16p11.2 carriers. We determined a list of dysregulated genes in which expression correlates with dosage at the 16p11.2 locus. BMP7 is a dysregulated *trans* effects from our genome wide expression study which gene expression levels positively correlate with genotype at the 16p11.2 locus. Bmp7 knockout mice have been characterized with having vertebral defects, as a result of decreased Bmp7 expression [61]. TBX6, a transcription factor found within the 16p11.2 interval is a gene involved in early development [167]. Tbx6 knockout mice also exhibit vertebral abnormalities, resulting from improper alignment of somites [60]. The skeletal abnormalities found in Tbx6 knockout mice seem similar to that seen in Bmp7 knockout mice. Similarly, 16p11.2 carriers exhibit skeletal abnormalities of the rib and vertebra.

Two publications reported sequencing the TBX6 gene in 16p11.2 deletion cases [35, 59]. Shimojima *et al* sequenced of the coding regions of TBX6 from 16p11.2 deletion carrier with rib and vertebral abnormalities [59]. Bijlisma *et al* also sequenced TBX6 and two other genes, ALDOA and SPN [35]. No mutation in TBX6 was found in either sequence analysis. Since no mutation was found through sequence analysis of patients with the 16p11.2 deletion, perhaps haploinsufficiency of TBX6 is enough to produce skeletal abnormalities in varying degrees of penetrance in 16p11.2 patients. Since Bmp7 and Tbx6 knockout mice have similar phenotypes we hypothesize that Tbx6 null mice will have lower expression of Bmp7 as compared to controls. We will use Tbx6-/- mice to test the reduced expression of bmp7 from both head and tail bud of 10.5 days post conception (dpc) embryos.

Tbx6 is part of the T-box transcription factor family, which plays a role during development [167]. Mutations in certain genes within the family have been associated with developmental syndromes. TBX5 is associated with Holt-oram syndrome which is characterized by abnormalities of the upper limbs and heart [168]. Patients with DiGeorge syndrome have haploinsufficiency of Tbx1 resulting from a microdeletion in 22q11.2 [169]. These patients have heart defects, and immune deficiencies, with a 20% that have autism or autism spectrum disorders [170].

Genes in the T-box transcription factor family are involved in cell fate, patterning, and organogenesis during vertebrate embryogenesis [167]. Tbx6 mediate other developmental genes through a large DNA binding domain with a consensus sequence of "AGGTGT". During embryogenesis the embryo is divided into three main sections, the ectoderm, mesoderm and the endoderm. TBX6 is responsible for directing regions of the mesoderm to develop Somites. Somites which do not develop properly result in fused ribs and hemivertabra [60, 166].

The gene responsible for patterning and specification of the somites is TBX6 [60]. TBX6 is first expressed in the primitive streak, and paraxial mesoderm. It is then down-regulated as somites develop. Mutant mouse models which are heterozygous mice for Tbx6 are normal, while homozygous mice exhibit defects in the paraxial mesoderm. In these mice, tissues which would be expected to differentiate into paraxial mesoderm instead differentiate into two ectopic neural tubes as a result of Tbx6 dysregulation. Mice with reduced Tbx6 expression levels have defects in somite patterning, fused ribs, fused vertebrae, and disrupted skeletal muscle patterning [60]. Tbx6 mutant mice have a phenotype of a short stature, and short tail. Adding back low levels of Tbx6 to these mutant mice, alleviates the severity of the rib and vertebrae fusions, suggesting this phenotype is dosage dependent on tbx6 [60].

BMPs and T-box genes both play a role in development and at certain time points mediate each others expression [101]. BMP signals target T-box proteins downstream in molecular pathways [101]. For example chick BMP2 promotes expression of Tbx2 and Tbx3. Indirect signaling between BMPs and T-box genes through indirect protein-protein interactions has also been observed. In one such instance Smad6 expression is upregulated by BMP, which then increases ubiquination of Tbx6 thereby reducing expression, and inhibiting myogenic genes. Here Smad6 is indirectly mediating signals between BMP and T-box proteins [101, 104]. We have not been able to discover a direct interaction between BMP7, one of the significant *trans* effects in our gene expression study and Tbx6. In light of these findings regarding crosstalk between BMPs and T-box genes, there may be an unknown signaling mediator between these two genes [101].

**Material and Methods:**

**Obtaining Tbx6 null knockout mice**

Frozen 10.5 dpc old mouse embryos were obtained from Deborah Chapman in the Department of Biological Sciences from the University of Pittsburgh. All samples were on a mixed 129SvJ/C57Bl6 background, and were a mixture of either head samples or tailbud samples. These samples included six Tbx6-/- mice along with seven control samples which were either heterozygous (Tbx6-/+) or homozygous normal (Tbx6+/+). (See table 6.1 for sample information)

**RNA extraction of Mouse embryo tissues.**

Embryonic mouse tissue samples are kept frozen on liquid nitrogen, and crushed into a powder using a pestle in a 2.5ml microcentrifuge tube. Tissue was pelleted down, and homogenized using Qiagen's Qiashredder. RNA extraction is performed using the RNeasy Mini kit following the protocol for RNA extraction

of frozen tissue exactly (Qiagen, Valencia, CA). RNA concentration and preliminary quality control is performed by Nanodrop ND-1000 spectrophotometer (Nanodrop technologies, Willmington DE).

**Real-time PCR of Tbx6-/- null knockout and control mice.**

RNA extracted from Tbx6 null knockout mice and control mice which passed quality control is used in real-time PCR. RNA is converted to cDNA using the Taqman RT-PCR reagents (Applied Biosystems). Real-time PCR Primers were designed for Bmp7, Wnt3a as a positive control, and Dll1, a gene not expressed in Tbx-/- mice as our negative control. Primers for Bmp7 include (5'AAGACGCCAAAGAACCAAGA 3') and reverse (5'TCAGGTGCAATGATCCAGTC 3'), Dll1 (5'AGCAACCCTGGCAGTGTAA 3') and reverse (5' TGCACGGCTTATGGTGAGTA 3'), and finally for Wnt3a (5' TTTGGAGGAATGGTCTCTCG 3') and reverse (5'CTTGAGGTGCATGTGACTGG 3'). 96-well optical plates were used for the Real-time PCR in triplicate for each gene tested. Real-time PCR reaction was carried out using 2X SYBR Green PCR master mix, 200 reactions (Applied Biosystems) in a 10ul volume. Real-time PCR was run using Real-time PCR ABI Prism 7000 Sequence Detection System (Applied Biosystems). The Real-time PCR cycling parameters were as follows: 48°C for 30 min, 95°C for 10 min, 40 cycles of (95°C for 15 sec, and 60°C for 30 sec, and 72°C for 30 sec) followed by a disassociation step of (95°C for 15 sec, and 60°C for 15 sec, and 95°C for 15 sec.) Results were processed using ABI Prism 7000 SDS software (Applied Biosystems). At the end of each run, the baseline and Ct was adjusted to ideal linearity for each standard curve. Final mRNA expression level was produced using the equation (2^(Ct test gene- Ct endogenous control)). (Comparative Method).

**Search for TBX6 consensus binding site in BMP7 promoter.**

In order to determine if the BMP7 promoter sequence has a transcription binding site for the TBX6 protein, I obtained the promoter sequence for BMP7 (5000 nucleotides upstream from the BMP7 start site) and searched for the TBX6 consensus site "AGGTGT"

**InnateDB** (www.innatedb.ca) This database is used analyze molecular interactions, predominantly for analysis of pathways and functions involved in immunity, but it also contains data for 100,000 other molecular interactions in both human and mouse. *Predicted Transcription Factor Interactions* were used in the analysis.

**Allen Brain Atlas**. (http://www.brain-map.org.) This tool was used to determine tissue specific co-expression in the brain between the significant *trans* effect BMP7 and a gene on the 16p11.2 region TBX6.

**RESULTS:**

Tbx6 is a gene on the 16p11.2 region which is associated with rib and vertebral malformation. 16p11.2 carriers exhibit an increased risk for skeletal abnormalities which center on the rib and vertebra. BMP7 is a gene which responds to the dosage at 16p11.2, and has been associated with rib and vertebral

mutations. Using Tbx6 null mice we wanted to determine if BMP7 expression was reduced in Tbx6 mice. The mean and median expression value for Tbx6 null mice was (2.8223; 2.2459) respectively (n=4). The mean and median value for control mice was (2.5883; 2.6558) respectively (n=5). There was greater variance of expression in Tbx6 null mice. These values showed that the expression of BMP7 was not significantly reduced in Tbx6 null mice as compared to controls. (Wilcoxon rank-sum test p-value = 0.5556)

The presence of a TBX6 consensus sequence on the BMP7 promoter would determine whether TBX6 is predicted to bind to BMP7 and activate transcription. We searched for the T-box consensus binding site "AGGTGT" 5kb upstream of the BMP7 start site the closest T-box consensus sequence was found ~3,300bp upstream. Using InnateDB's *Predicted Transcription Factor Interactions* feature we found that BMP genes did not have a predicted binding site to any genes in the T-box family.

Co-expression between BMP7 a significant *trans* effect, and MAPK3 a gene on the 16p11.2 region was analyzed using the Allen Brain Atlas. The co-expression value between these two genes in the whole brain was quite low (0.0986). The regions of the brain which shared highest levels of co-expression are the epithalmus, parafascicular nucleus and the nucleus of the solitary tract.

| Chapman Sample Name | Sample name | Genotype | Mouse Background (mixed) | Head/tail | Bmp7 | Wnt3a | 2^(Ct Bmp7 - Ct Wnt3a |
|---|---|---|---|---|---|---|---|
| T6Mut | 2 | Tbx6-/- | 129SvJ/C57Bl6 | 2 heads | 19.28555 | 18.03937 | 2.372130423 |
| T6Mut | 3 | Tbx6-/- | 129SvJ/C57Bl6 | 2 tailbuds | 19.7301 | 19.067 | 1.583481487 |
| T6Mut | 5 | Tbx6-/- | 129SvJ/C57Bl6 | 2 heads | 18.698033 | 17.6141 | 2.119807609 |
| T6Mut | 6 | Tbx6-/- | 129SvJ/C57Bl6 | 2 tailbuds | 20.175067 | 17.79263 | 5.214154505 |
| Control | 7 | Tbx6+/+ or Tbx6+/- | 129SvJ/C57Bl6 | 2 tailbuds | 18.461033 | 17.05187 | 2.655837113 |
| Control | 10 | Tbx6+/+ or Tbx6+/- | 129SvJ/C57Bl6 | 1 tailbud | 20.729067 | 19.5917 | 2.199791313 |
| Control | 11 | Tbx6+/+ or Tbx6+/- | 129SvJ/C57Bl6 | 2 tailbuds | 19.356433 | 18.1784 | 2.262681204 |
| Control | 12 | Tbx6+/+ or Tbx6+/- | 129SvJ/C57Bl6 | 2 heads | 19.181933 | 17.64607 | 2.899625671 |
| Control | 13 | Tbx6+/+ or Tbx6+/- | 129SvJ/C57Bl7 | 2 heads | 18.4234 | 16.87543 | 2.924047336 |

**Figure 6.1 mRNA expression and sample data for Tbx6 null mice, and controls**

The mRNA expression of Bmp7 and Wnt3a for Tbx6 null mice, and a mixture of controls (Tbx6+/+ or Tbx6+/-) by RT-PCT for each sample can be found in figure 6.1. The numbers given for BMP7 and Wnt3a represent Ct scores. mRNA expression was generated using the Comparative Method. (2^(Ct test gene- Ct endogenous control))

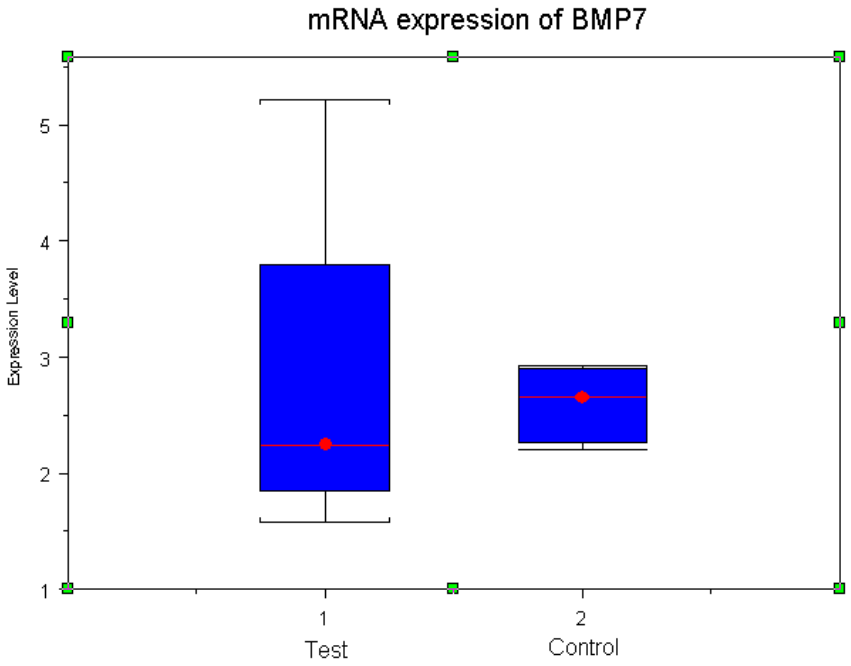



**Figure 6.1 Box whisker plot showing mRNA expression for BMP7 in Tbx6 null mice and controls.**
Box Whisker plot of mRNA expression for Tbx6-/- (n=4) and controls (Tbx6+/+ or Tbx6+/-) (n=5). The blue top and bottom boxes are the 75$^{th}$ and 25$^{th}$ percentile respectively. The red dot and line represent the 50$^{th}$ percentile or the median. There is no significant difference in Bmp7 expression between Tbx6 null mice and control samples (Wilcoxon rank-sum test p-value = 0.5556)

Sequence of BMP7 (see below)

ctgaacgtgcactgatgcttaaaggcctgggctgtgtagacagatgtggc
tttggatcctggctccattacagattagattggctgtctgactctggaca
aggtgctctcgggatgttggtcctcctctgcaatctgaggagcagaacta
gcctctccatgcagtggggagttgcgggtgatgtgggcttgggccgtgca
ggcctagcaggtgcccagcagaagcaagtgctcccaacaggtgacaggtc
ggccgctcccctgtcacgttttggaaggaggaaagggctacctctagtgc
tgaaccgaggatagttagtgctcaaaaccataccagatttcctccaatca
aggaagaaataacagccctgataataataaacaaaaccactgcttccccc
tagctggcttccagatcctagaaatcctgcatggtatcagctttctctag
agtgtgtgtgtcgtctgtattttccatttggaactgtggccactgccatg
tgtacttaagactgatggaagacgtttaatctcttattctgctgctttta
aagttgtgcaaagaaaatctcagagtgggcaagcgtgaggagtccaagcc
tcccgtgcaatgagagatctgcgtgggaaacaaaatttcaccac**aggtgt**
gctctaataatttcctctgcagggtccacctctgcacacaacgattttca
agcctgtgcaggaggcagctcttggccctctattccttgttggggttgga
ggtccaggtcatctggcctttgccaggtggtgtggacagagagagagcag
cccactggcttccttggctgtgctccccagggagcttccaggccagctga
gcctcctctagggcagatggatgaagatgaggtccccacctcgctgaaca
atgtgtgtgtcagcagaattcttttcttcttttactgccttagagagac

atgctgacttggtcaaaatcacttcagcaagatgtgactgatgatgtagg
aaatgaaaactcctggcccagcactgggaggccagcaggcgaggggcgcg
gacactgggaccgggccacccgacacaataacacccttacaattccagg
ccgtctttcatccaggcaaaggcggcccc**aggtgt**ggagcagactggggc
agatcagattacggtcctggagcccagtgctgggtgccctggccagggga
accacaactgggggtgtgtgggttgggggctagctgctggggaaatgaga
taggacccagggaagtccctgggccccagctggcctgcaggggcctagg
cacaatgtgaggattggaatgcagtgatgacctttcatctctcatgttc
ccttccacttcactctctcttttgctctggtacttttgctcttgctctca
gagtcaggatacatctgcttttctctgcttggccaaaaagcagaccatca
tgaaaagttttgttgttattttgtaactcaggaggctttcctgttaccat
attttccttttcatccagatggtaacacaataacatttatcaagggtttt
ctctgtgttccaggcactgtgcaaagccctttccatgaattaagttctgc
aacctcacagcgatcccagaagacaggcgctatcagtcaccccattttac
agatgggggaactg**aggtgt**agagaggttaagtcccttgcccatggtgca (3,300 bp)
cagctggaagagacagagctggagtgtgaatgcggctgggcaggctccag
tgcccaggctacctcctccacacaagacttgccctcggcaatctcaaagc
cttttctggtggtgggctcagctcccaaactggcatcggatgcactccca
gccagatatttcttgcttgccggttttcattcattcattcattcattcat
tcattcattcattcggcattcactgagggcctggtgcagtcttggggatg
cctctcggggaggaaacagggaaaagaaagacccccccaccaagcatgga
tcacagaaaagataaggctaaatgggggtttgtgggacttcagaggaaac
cttatctcttgaggtcttggatatgaagagcatgttgtctcttcctcttg
catgagaaaagatggcgtctcagaggaagggttgctggggtgagggatct
gggagatgccttagcttggcgcctgcacagtcagccctcagtcaaccggt
ctctttaggttttggctgtgcttattactattcattcaacaggtactaat
tgagcacctgctgtgtgccaggctcagaataggctcaggtgagatgcaca
aagaagggtaaactagaatccttgcttagacactgacggatcagttgttt
catatgtaaattgtagcaccaagacctgctgcccctgcccccagcctcac
ctgcttgtgaagatccctccaaaagatttgagagtagataaaaagcagag
actactactgaagaacagggctgctttggctccttattatttcagacttt
ggaagaaaatgacctccttttttctctactggcactggaggtggcatagcc
tgtccctagcaagccagcgctggagggcgtgtgcagggctggggaccgag
cctggtttctgttccctgctctgcagggctcaagcacttgctgttcctcc
acctgggatgcctttcccctggaaaagcctgtctctttcttgtctttcagg
actcaggtcagtggcatctcctccaaaaactcccccttcccaccctccatc
acctcaccctgtttatctgcgcccccgcccccactgcctgtcacttattg
caggctgaagtgacccaggctctccagttgtacactctcagatggaccct
ggacgactgtggcactcctgcaatttccccagtctccctggggtaggatt
cctgcttgccaggatgcccacctttccttctccctcctgcatgtcctcct
ctgcctggcttctgaattgtttccagagagagtgatagacaagatctgcc
tctccttcagtccctgaatcttatttaaggctcttgctttgcttccctgg
cctggaggccggctccttgatggagtctgccatgtgggttcgctcatggc
catgtcttcctgcccagcatggtgcttggccctgggactggccacataat
atctgggccaggtgcaaaattagtacggggcaggggtactttgttcata
ggtgattcagaaccacatatggtgacctcagagtaggaaaccaagtgtgg
ggcccttaagagctgggggccctgtacgactgtccaggttgcaggcccc
acagctcgcctcctgatatcctgtgctccatgcttgtctgttgaaggaag
gagtgaatggatgaagagcaggtggtgggggtggtttgagggccttgcct
ggtgggtgggtagaggcccctccctggcatggggctcaagacctgttcca
tcccacagcctggggcctgtgtgtaaatggccaggacctgcaggctggca
tttttctgctccttgcctggcctctggcctccccttctccacccatgtg
gcccctcaggctgccatctagtccaaaagtccccaagggagacccagagg
gccacttggccaaactacttctgctccagaaaactgtagaagaccataat
tctcttccccagctctcctgctccaggaaggacagccccaaagtgaggct
tagccagagcccctcccagacaagcgcccccgcttccccaacctcagccc

```
ttcccagttcatcccaaaggccctctggggacccactctctcacccagcc
ccaggaggggaaggagacaggatgaacttttaccccgctgccctcactgc
cactctgggtgcagtaattcccttgagatcccacaccggcagagggaccg
gtgggttctgagtggtctggggactccctgtgacagcgtgcatggctcgg
tattgattgagggatgaatggatgaggagagacaggagaggaggccgatg
gggaggtctcaggcacagacccttggaggggaagaggatgtgaagaccag
cggctggctccccaggcactgccacgaggagggctgatgggaagccctag
tggtggggctggggtgtctggtctcaggctgaggggtggctggaaagata
cagggccccgaagaggaggaggtgggaagaacccccccagctcacacgca
gttcacttattcactcaacaaatcgtgactgcgcagctacagtggctacc
aggcgctgggttcaaggcactgcgggtaccagaggtgcggagaagatcgc
tgatccgggccccagtgctctgggtgtctagcgggggtaagaaggcaata
aagaaggcacggagtaactcaaacagcaattccagacagcaagagaaact
acaggaaagaaaacaaacgtgcgaggggcgaggcgaggaaacaacctcag
cttggcaggtcttggaggtctctgggaggagaaagcagcgtctgatgggg
gcgggaggtggtgagtggggagaggtccaggcggagggaatggcgagcgc
agagacaggctggcaacggcttcagggaggcgcggaggggtcagcgtggc
tggcttaaaaggatacagggactgaggggcaagaccggctcaagggtcac
cgcttccaggaagccttctatttccgcgccacctccgcgctcccccaact
tttcccaccgcggtccgcagcccacccgtcctgctcgggccgccttcctg
gtccggaccgcgagtgccgagagggcagggccggctccgattcctccagc
cgcatccccgcgacgtcccgccaggctctaggcaccccgtgggcactcag
taaacatttgtcgagcgctctagagggaatgaatgaacccactgggcaca
gctgggggggagggcggggccgagggcaggtgggaggccgccggcgcggga
ggggcccctcgaagcccgtcctcctcctcctcctcctccgcccaggcccc
0 (start site)
```

**Figure 6.2 Search for TBX6 consensus binding site in BMP7 promoter.** In order to determine if BMP7 has a transcription binding factor for TBX6, I acquired BMP7's promoter sequence and searched for "AGGTGT" the TBX6 consensus sequence. Three consensus sequences were found. The closest one to the promoter was 3,300bp upstream. This occurrence may be random. Additionally the sequence is too far away from the start site. Bioinformatic tools which look for transcription binding sites did not identify the TBX6 binding site in the sequence, nor did they find a binding site for any T-box protein.

| Query Xre | Query Name | Transcription Factor |
|---|---|---|
| 655 | BMP7 | TF SMAD1 is predicted to bind BMP7 gene |
| 650 | BMP2 | TF SMAD4 is predicted to bind BMP2 gene |
| 649 | BMP1 | TF SOX10 is predicted to bind BMP1 gene |
| 653 | BMP5 | TF SOX10 is predicted to bind BMP5 gene |
| 655 | BMP7 | TF SP1 is predicted to bind BMP7 gene |
| 27302 | BMP10 | TF SRF is predicted to bind BMP10 gene |
| 27302 | BMP10 | TF SRY is predicted to bind BMP10 gene |
| 652 | BMP4 | TF STAT1 is predicted to bind BMP4 gene |
| 653 | BMP5 | TF STAT1 is predicted to bind BMP5 gene |
| 27302 | BMP10 | TF STAT5A is predicted to bind BMP10 gene |
| 652 | BMP4 | TF STAT5A is predicted to bind BMP4 gene |
| 653 | BMP5 | TF STAT5A is predicted to bind BMP5 gene |
| 27302 | BMP10 | TF TEF is predicted to bind BMP10 gene |
| 27302 | BMP10 | TF ETV7 is predicted to bind BMP10 gene |
| 654 | BMP6 | TF GTF2I is predicted to bind BMP6 gene |
| 654 | BMP6 | TF USF is predicted to bind BMP6 gene |
| 27302 | BMP10 | TF USF1 is predicted to bind BMP10 gene |
| 654 | BMP6 | TF USF1 is predicted to bind BMP6 gene |
| 654 | BMP6 | TF XBP1 is predicted to bind BMP6 gene |
| 653 | BMP5 | TF YY1 is predicted to bind BMP5 gene |
| 654 | BMP6 | TF ZEB1 is predicted to bind BMP6 gene |

**Table 6.2 Predicted transcription binding sites between T-box, and BMP genes using InnateDB**

InnateDB predicted transcription factor (TF) interaction function was used in our analysis. We input BMP genes into the database to determine if any of the genes them interacted with a t-box gene. No interactions between the two gene families were found. (TF binding factors are alphabetical, showing only snapshot near T-box)

| Region ('fine' annotation) | BMP7 / TBX6 similarity coefficient |
|---|---|
| 'Epithalamus' | 0.8775 |
| 'Parafascicular nucleus' | 0.787 |
| 'Nucleus of the solitary tract' | 0.7628 |
| 'Dorsomedial nucleus of the hypothalamus' | 0.7579 |
| 'Motor nucleus of trigeminal' | 0.7346 |
| 'Cerebellar nuclei' | 0.7111 |
| 'Anteroventral nucleus of thalamus ' | 0.6975 |
| 'Ventromedial hypothalamic nucleus' | 0.6835 |
| 'Nucleus of the lateral olfactory tract' | 0.6773 |
| 'Spinal nucleus of the trigeminal_ oral part' | 0.6495 |
| 'Anterior pretectal nucleus ' | 0.6319 |
| 'Midbrain raphé nuclei' | 0.6094 |
| 'Hypothalamic lateral zone' | 0.6067 |
| 'Dorsal column nuclei' | 0.5921 |
| 'Superior colliculus_ sensory related' | 0.5838 |
| 'Arcuate hypothalamic nucleus' | 0.5827 |
| 'Facial motor nucleus' | 0.5702 |
| 'Medial amygdalar nucleus ' | 0.5614 |
| 'Inferior olivary complex' | 0.5589 |
| 'Mediodorsal nucleus of thalamus' | 0.5406 |
| 'Intralaminar nuclei of the dorsal thalamus' | 0.5325 |
| 'Tuberal nucleus' | 0.5282 |
| 'Posterior hypothalamic nucleus' | 0.5266 |
| 'Mammillary body' | 0.5259 |
| 'Anteromedial nucleus' | 0.5222 |
| 'Pedunculopontine nucleus' | 0.4986 |
| 'Anterior hypothalamic nucleus' | 0.4813 |
| 'Cerebellar cortex' | 0.4809 |
| 'Ventral part of the lateral geniculate complex' | 0.4807 |
| 'Hypoglossal nucleus' | 0.4602 |
| 'Vestibular nuclei' | 0.46 |
| 'Lateral reticular nucleus' | 0.4319 |
| 'Cochlear nuclei ' | 0.4295 |
| 'Pontine gray' | 0.4199 |
| 'Lateral group of the dorsal thalamus' | 0.4163 |
| 'Nucleus of the lateral lemniscus' | 0.4163 |
| 'Central amygdalar nucleus ' | 0.409 |
| 'Superior central nucleus raphé' | 0.4086 |
| 'Ventral medial nucleus of the thalamus' | 0.406 |
| 'Zona incerta' | 0.3989 |
| 'Pretectal region' | 0.3725 |
| 'Ventral tegmental area' | 0.37 |
| 'Tegmental reticular nucleus' | 0.3685 |
| 'Lateral septal nucleus ' | 0.3659 |
| 'Spinal nucleus of the trigeminal_ caudal part' | 0.3529 |

| | |
|---|---|
| 'Superior colliculus_ motor related' | 0.3477 |
| 'Piriform-amygdalar area' | 0.3395 |
| 'Subiculum' | 0.3393 |
| 'Inferior colliculus' | 0.3363 |
| 'Parabrachial nucleus' | 0.3349 |
| 'Principal sensory nucleus of the trigeminal' | 0.3127 |
| 'Periaqueductal gray' | 0.2993 |
| 'Bed nuclei of the stria terminalis ' | 0.2987 |
| 'Caudoputamen ' | 0.2953 |
| 'Ventral group of the dorsal thalamus' | 0.2876 |
| 'Superior olivary complex' | 0.2854 |
| 'Lateral posterior nucleus of the thalamus' | 0.2842 |
| 'Accessory olfactory bulb' | 0.2809 |
| 'Cuneiform nucleus' | 0.2699 |
| 'Paragigantocellular reticular nucleus' | 0.2694 |
| 'Medulla' | 0.2677 |
| 'Reticular nucleus of the thalamus' | 0.2673 |
| 'Substantia nigra_ reticular part' | 0.2671 |
| 'Dentate gyrus' | 0.2614 |
| 'Hypothalamus' | 0.2558 |
| 'Taenia tecta' | 0.254 |
| 'Midbrain reticular nucleus_ retrorubral area' | 0.2494 |
| 'Nucleus accumbens ' | 0.2472 |
| 'Lateral dorsal nucleus of thalamus' | 0.2446 |
| 'Ammon's Horn' | 0.2421 |
| 'Olfactory tubercle ' | 0.2354 |
| 'Cortical amygdalar area' | 0.2283 |
| 'Medulla_ behavioral state related' | 0.2125 |
| 'Pallidum_ dorsal region' | 0.2012 |
| 'Pallidum_ medial region' | 0.1985 |
| 'Dorsal part of the lateral geniculate complex' | 0.1948 |
| 'Spinal nucleus of the trigeminal_ interpolar part' | 0.1941 |
| 'Anterior amygdalar area' | 0.1813 |
| 'Fundus of striatum' | 0.1801 |
| 'Cerebral cortex' | 0.1797 |
| 'Anterior olfactory nucleus' | 0.176 |
| 'Pallidum_ ventral region' | 0.1757 |
| 'Substantia nigra_ compact part' | 0.1735 |
| 'Medial geniculate complex' | 0.1732 |
| 'Midbrain' | 0.1659 |
| 'Pons' | 0.1629 |
| 'Red Nucleus' | 0.1572 |
| 'Magnocellular reticular nucleus' | 0.1531 |
| 'Ventral posterior complex of the thalamus' | 0.1309 |
| 'Main olfactory bulb' | 0.1176 |

**Table 6.3 List of brain tissue (fine annotation) with highest co-expression values between TBX6, and BMP7 using Allen Brain Atlas**

Co-expression values or similarity co-efficient from highest co-expression to lower co-expression in the regions of the brain (fine annotation) for BMP7 one of the significant *trans* effect and Tbx6 a gene on the 16p11.2 region.

| Region ('big12' annotation) | BMP7 / TBX6 similarity coefficient |
|---|---|
| 'Cerebellum' | 0.4794 |
| 'Hypothalamus' | 0.2422 |
| 'Hippocampal region' | 0.2393 |
| 'Medulla' | 0.2289 |
| 'Thalamus' | 0.2091 |
| 'Midbrain' | 0.2046 |
| 'Striatum' | 0.1903 |
| 'Pons' | 0.1864 |
| 'Cerebral cortex' | 0.1797 |
| 'Retrohippocampal region' | 0.0938 |
| 'Pallidum' | 0.0902 |
| 'Olfactory areas' | 0.0794 |

**Table 6.4 List of brain tissue (big 12 annotation) with highest co-expression values between TBX6, and BMP7 using Allen Brain Atlas**
Co-expression values or similarity co-efficient from highest co-expression to lower co-expression for BMP7 a significant *trans* effect and MAPK3 a gene on the 16p11.2 region.

**DISCUSSION:**

After conducting a genome-wide expression analysis using lymphoblast cell lines from 16p11.2 carriers we determined that BMP7 is a gene outside of the mutation which expression also correlates with genotype at the 16p11.2 locus. Animal models of both TBX6, a gene on the 16p11.2 region, and BMP7 exhibit skeletal abnormalities involving the ribs and vertebra. This was particularly interesting because 16p11.2 carriers have rib and vertebral abnormalities at a rate that is higher than seen in the general population. We wanted to determine if TBX6 was mediating the dysregulation of BMP7, and contributing to the skeletal abnormalities associated with 16p11.2 cases. Using RT-PCR we wanted to detect the mRNA expression levels of Bmp7 in Tbx6 null mice. Our results show that there is no significant difference in Bmp7 expression between Tbx6 mutant mice and controls. This shows that TBX6 is probably not the gene on 16p11.2 region which is affecting the dysregulation of BMP7.

We also used InnateDB to determine if any gene in the BMP family had a transcription factor binding site for a gene in the T-box family. There was no interaction. We scanned the DNA 5kb upstream of the human BMP7 gene start site to find the T-box consensus sequence. A consensus sequence was found

~3,300bp upstream of the BMP7 start site. This was considered to far to predict that a t-box gene binds there.

Co-expression results from Allen Brain Atlas showed that BMP7 and TBX6 are not as co-expressed in the brain as compared to other genes. (MAPK3, see figure 4.7 and 4.8)

We were not able to determine an interaction between BMP7 and TBX6. Additionally, we were not able to determine a decrease in BMP7 mRNA expression in TBX6 mice. Therefore, we do not believe that TBX6 is mediated in the dysregulation of BMP7 in lymphoblast cell lines in 16p11.2 carriers.

**List of Literature cited:**


1. Mefford, H.C. and E.E. Eichler, *Duplication hotspots, rare genomic disorders, and common disease.* Curr Opin Genet Dev, 2009. **19**(3): p. 196-204.
2. Kumar, R.A., et al., *Recurrent 16p11.2 microdeletions in autism.* Hum Mol Genet, 2008. **17**(4): p. 628-38.
3. McCarthy, S.E., et al., *Microduplications of 16p11.2 are associated with schizophrenia.* Nat Genet, 2009. **41**(11): p. 1223-7.
4. Mefford, H.C., et al., *A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease.* Genome Res, 2009. **19**(9): p. 1579-85.
5. Weiss, L.A., et al., *Association between microdeletion and microduplication at 16p11.2 and autism.* N Engl J Med, 2008. **358**(7): p. 667-75.
6. Ionita-Laza, I., et al., *Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis.* Genomics, 2009. **93**(1): p. 22-6.
7. Merikangas, A.K., A.P. Corvin, and L. Gallagher, *Copy-number variants in neurodevelopmental disorders: promises and challenges.* Trends Genet, 2009. **25**(12): p. 536-44.
8. Wu, X. and H. Xiao, *Progress in the detection of human genome structural variations.* Sci China C Life Sci, 2009. **52**(6): p. 560-7.
9. Lee, J.A. and J.R. Lupski, *Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders.* Neuron, 2006. **52**(1): p. 103-21.
10. van Ommen, G.J., *Frequency of new copy number variation in humans.* Nat Genet, 2005. **37**(4): p. 333-4.
11. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.
12. Cook, E.H., Jr. and S.W. Scherer, *Copy-number variations associated with neuropsychiatric conditions.* Nature, 2008. **455**(7215): p. 919-23.
13. Ionita-Laza, I., et al., *On the analysis of copy-number variations in genome-wide association studies: a translation of the family-based association test.* Genet Epidemiol, 2008. **32**(3): p. 273-84.
14. Xu, B., et al., *Strong association of de novo copy number mutations with sporadic schizophrenia.* Nat Genet, 2008. **40**(7): p. 880-885.
15. Boerkoel, C.F., et al., *Charcot-Marie-Tooth disease and related neuropathies: mutation distribution and genotype-phenotype correlation.* Ann Neurol, 2002. **51**(2): p. 190-201.
16. Nelis, E., et al., *Estimation of the mutation frequencies in Charcot-Marie-Tooth disease type 1 and hereditary neuropathy with liability to pressure palsies: a European collaborative study.* Eur J Hum Genet, 1996. **4**(1): p. 25-33.

17. Patel, P.I., et al., *Isolation of a marker linked to the Charcot-Marie-Tooth disease type IA gene by differential Alu-PCR of human chromosome 17-retaining hybrids.* Am J Hum Genet, 1990. **47**(6): p. 926-34.
18. *Genetics of Complex Human Diseases: A Laboratory Manual*, ed. A. Al-Chalabi. 2009, Cold Spring Harbor: Cold Spring Harbor Press.
19. Lupski, J.R., *Genomicorders: structural features of the genome can lead to DNA rearrangements and human disease traits dis.* Trends Genet, 1998. **14**(10): p. 417-22.
20. Dierssen, M., Y. Herault, and X. Estivill, *Aneuploidy: from a physiological mechanism of variance to Down syndrome.* Physiol Rev, 2009. **89**(3): p. 887-920.
21. Patterson, D., *Molecular genetic analysis of Down syndrome.* Hum Genet, 2009. **126**(1): p. 195-214.
22. *Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002.* MMWR Surveill Summ, 2007. **56**(1): p. 12-28.
23. Brandon, N.J., et al., *Subcellular targeting of DISC1 is dependent on a domain independent from the Nudel binding site.* Mol Cell Neurosci, 2005. **28**(4): p. 613-24.
24. Burdick, K.E., et al., *Elucidating the Relationship between DISC1, NDEL1, and NDE1 and the Risk for Schizophrenia: Evidence of Epistasis and Competitive Binding.* Hum Mol Genet, 2008.
25. Hennah, W., et al., *Genes and schizophrenia: beyond schizophrenia: the role of DISC1 in major mental illness.* Schizophr Bull, 2006. **32**(3): p. 409-16.
26. Bailey, J.A., et al., *Recent segmental duplications in the human genome.* Science, 2002. **297**(5583): p. 1003-7.
27. Sharp, A.J., et al., *Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome.* Nat Genet, 2006. **38**(9): p. 1038-42.
28. Stankiewicz, P. and J.R. Lupski, *Genome architecture, rearrangements and genomic disorders.* Trends Genet, 2002. **18**(2): p. 74-82.
29. Shinawi, M., et al., *Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioral problems, dysmorphism, epilepsy, and abnormal head size.* J Med Genet, 2009.
30. Marshall, C.R., et al., *Structural variation of chromosomes in autism spectrum disorder.* Am J Hum Genet, 2008. **82**(2): p. 477-88.
31. Shen, Y. and B.L. Wu, *Microarray-based genomic DNA profiling technologies in clinical molecular diagnostics.* Clin Chem, 2009. **55**(4): p. 659-69.
32. Feuk, L., et al., *Structural variants: changing the landscape of chromosomes and design of disease studies.* Hum Mol Genet, 2006. **15 Suppl 1**: p. R57-66.
33. Addington, A.M. and J.L. Rapoport, *The genetics of childhood-onset schizophrenia: when madness strikes the prepubescent.* Curr Psychiatry Rep, 2009. **11**(2): p. 156-61.
34. Fernandez, B.A., et al., *Phenotypic Spectrum Associated with De Novo and Inherited Deletions and Duplications at 16p11.2 in Individuals Ascertained for Diagnosis of Autism Spectrum Disorder.* J Med Genet, 2009.

35. Bijlsma, E.K., et al., *Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals.* Eur J Med Genet, 2009. **52**(2-3): p. 77-87.
36. Unique.guide, *16p11.2microdeletions and 16p proximal deletions* 2009.
37. Palmen, S.J., et al., *Neuropathological findings in autism.* Brain, 2004. **127**(Pt 12): p. 2572-83.
38. Torrey, E.F. and M.R. Peterson, *Schizophrenia and the limbic system.* Lancet, 1974. **2**(7886): p. 942-6.
39. Tam, G.W., et al., *Confirmed rare copy number variants implicate novel genes in schizophrenia.* Biochem Soc Trans. **38**(2): p. 445-51.
40. Bailey, A., et al., *Autism as a strongly genetic disorder: evidence from a British twin study.* Psychol Med, 1995. **25**(1): p. 63-77.
41. Abrahams, B.S. and D.H. Geschwind, *Advances in autism genetics: on the threshold of a new neurobiology.* Nat Rev Genet, 2008. **9**(5): p. 341-355.
42. Levy, S.E., et al., *Autism spectrum disorder and co-occurring developmental, psychiatric, and medical conditions among children in multiple populations of the United States.* J Dev Behav Pediatr. **31**(4): p. 267-75.
43. Amir, R.E., et al., *Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.* Nat Genet, 1999. **23**(2): p. 185-8.
44. Levy, S.E., D.S. Mandell, and R.T. Schultz, *Autism.* Lancet, 2009. **374**(9701): p. 1627-38.
45. Yeargin-Allsopp, M., et al., *Prevalence of Autism in a US Metropolitan Area.* JAMA, 2003. **289**(1): p. 49-55.
46. Lakhan, S.E. and K.F. Vieira, *Schizophrenia pathophysiology: are we any closer to a complete model?* Ann Gen Psychiatry, 2009. **8**: p. 12.
47. Owen, M.J., H.J. Williams, and M.C. O'Donovan, *Schizophrenia genetics: advancing on two fronts.* Curr Opin Genet Dev, 2009. **19**(3): p. 266-70.
48. van Os, J. and S. Kapur, *Schizophrenia.* Lancet, 2009. **374**(9690): p. 635-45.
49. Devon, R.S., et al., *Identification of polymorphisms within Disrupted in Schizophrenia 1 and Disrupted in Schizophrenia 2, and an investigation of their association with schizophrenia and bipolar affective disorder.* Psychiatr Genet, 2001. **11**(2): p. 71-8.
50. James, R., et al., *Disrupted in Schizophrenia 1 (DISC1) is a multicompartmentalized protein that predominantly localizes to mitochondria.* Mol Cell Neurosci, 2004. **26**(1): p. 112-22.
51. Millar, J.K., et al., *Disruption of two novel genes by a translocation co-segregating with schizophrenia.* Hum Mol Genet, 2000. **9**(9): p. 1415-23.
52. Bleuler, M. and R. Bleuler, *Dementia praecox oder die Gruppe der Schizophrenien: Eugen Bleuler.* Br J Psychiatry, 1986. **149**: p. 661-2.
53. Crespi, B., P. Stead, and M. Elliot, *Evolution in health and medicine Sackler colloquium: Comparative genomics of autism and schizophrenia.* Proc Natl Acad Sci U S A. **107 Suppl 1**: p. 1736-41.
54. Russell-Smith, S.N., M.T. Maybery, and D.M. Bayliss, *Are the Autism and Positive Schizotypy Spectra Diametrically Opposed in Local Versus Global Processing?* J Autism Dev Disord.

55. Carroll, L.S. and M.J. Owen, *Genetic overlap between autism, schizophrenia and bipolar disorder.* Genome Med, 2009. **1**(10): p. 102.
56. Almeida, R.A., et al., *A new step towards understanding Embedded Figures Test performance in the autism spectrum: the radial frequency search task.* Neuropsychologia. **48**(2): p. 374-81.
57. Yan, J., et al., *Neurexin 1alpha structural variants associated with autism.* Neurosci Lett, 2008. **438**(3): p. 368-70.
58. Beckmann, J.S., X. Estivill, and S.E. Antonarakis, *Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability.* Nat Rev Genet, 2007. **8**(8): p. 639-46.
59. Shimojima, K., et al., *A familial 593-kb microdeletion of 16p11.2 associated with mental retardation and hemivertebrae.* Eur J Med Genet, 2009. **52**(6): p. 433-5.
60. White, P.H., et al., *Defective somite patterning in mouse embryos with reduced levels of Tbx6.* Development, 2003. **130**(8): p. 1681-90.
61. Jena, N., et al., *BMP7 null mutation in mice: developmental defects in skeleton, kidney, and eye.* Exp Cell Res, 1997. **230**(1): p. 28-37.
62. Luo, G., et al., *BMP-7 is an inducer of nephrogenesis, and is also required for eye development and skeletal patterning.* Genes Dev, 1995. **9**(22): p. 2808-20.
63. Kishi, H., et al., *Human aldolase A deficiency associated with a hemolytic anemia: thermolabile aldolase due to a single base mutation.* Proc Natl Acad Sci U S A, 1987. **84**(23): p. 8623-7.
64. Kreuder, J., et al., *Brief report: inherited metabolic myopathy and hemolysis due to a mutation in aldolase A.* N Engl J Med, 1996. **334**(17): p. 1100-4.
65. Orita, S., et al., *Doc2: a novel brain protein having two repeated C2-like domains.* Biochem Biophys Res Commun, 1995. **206**(2): p. 439-48.
66. Sakaguchi, G., et al., *Doc2alpha is an activity-dependent modulator of excitatory synaptic transmission.* Eur J Neurosci, 1999. **11**(12): p. 4262-8.
67. Bedoyan, J.K., et al., *Duplication 16p11.2 in a child with infantile seizure disorder.* Am J Med Genet A. **152A**(6): p. 1567-74.
68. Kumar, R.A., et al., *Association and mutation analyses of 16p11.2 autism candidate genes.* PLoS One, 2009. **4**(2): p. e4582.
69. Crepel. *Narrowing the critical deletion region for autism spectrum disorders in chromosome 16p11.2*. in *World Congress of Psychatric Genetics 2009*. 2009. San Diego California US.
70. Tidyman, W.E. and K.A. Rauen, *The RASopathies: developmental syndromes of Ras/MAPK pathway dysregulation.* Current Opinion in Genetics & Development, 2009. **19**(3): p. 230-236.
71. Mazzucchelli, C., et al., *Knockout of ERK1 MAP kinase enhances synaptic plasticity in the striatum and facilitates striatal-mediated learning and memory.* Neuron, 2002. **34**(5): p. 807-20.
72. Chwang, W.B., et al., *ERK/MAPK regulates hippocampal histone phosphorylation following contextual fear conditioning.* Learn Mem, 2006. **13**(3): p. 322-8.
73. Stavridis, M.P., et al., *A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification.* Development, 2007. **134**(16): p. 2889-94.

74. Di Cristo, G., et al., *Requirement of ERK activation for visual cortical plasticity.* Science, 2001. **292**(5525): p. 2337-40.
75. Geschwind, D.H., *DNA microarrays: translation of the genome from laboratory to clinic.* Lancet Neurol, 2003. **2**(5): p. 275-82.
76. Mantripragada, K.K., et al., *Genomic microarrays in the spotlight.* Trends Genet, 2004. **20**(2): p. 87-94.
77. Ortega, J.A. and S. Alcantara, *BDNF/MAPK/ERK-Induced BMP7 Expression in the Developing Cerebral Cortex Induces Premature Radial Glia Differentiation and Impairs Neuronal Migration.* Cereb Cortex, 2009.
78. Bochukova, E.G., et al., *Large, rare chromosomal deletions associated with severe early-onset obesity.* Nature. **463**(7281): p. 666-70.
79. Ghebranious, N., et al., *A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation.* Am J Med Genet A, 2007. **143**(13): p. 1462-71.
80. Shiow, L.R., et al., *Severe combined immunodeficiency (SCID) and attention deficit hyperactivity disorder (ADHD) associated with a Coronin-1A mutation and a chromosome 16p11.2 deletion.* Clin Immunol, 2009. **131**(1): p. 24-30.
81. Bowden, N.A., et al., *Preliminary investigation of gene expression profiles in peripheral blood lymphocytes in schizophrenia.* Schizophr Res, 2006. **82**(2-3): p. 175-83.
82. Curtis, R.K., M. Oresic, and A. Vidal-Puig, *Pathways to the analysis of microarray data.* Trends Biotechnol, 2005. **23**(8): p. 429-35.
83. de Vries, B.B., et al., *Diagnostic genome profiling in mental retardation.* Am J Hum Genet, 2005. **77**(4): p. 606-16.
84. Hu, V.W., et al., *Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes.* BMC Genomics, 2006. **7**: p. 118.
85. Whitney, A.R., et al., *Individuality and variation in gene expression patterns in human blood.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1896-901.
86. Yoon, H., et al., *Gene expression profiling of isogenic cells with different TP53 gene dosage reveals numerous genes that are affected by TP53 dosage and identifies CSPG2 as a direct target of p53.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15632-7.
87. Ooi, L. and I.C. Wood, *Regulation of gene expression in the nervous system.* Biochem J, 2008. **414**(3): p. 327-341.
88. Gregg, J.P., et al., *Gene expression changes in children with autism.* Genomics, 2008. **91**(1): p. 22-9.
89. Baron, C.A., et al., *Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism.* J Autism Dev Disord, 2006. **36**(8): p. 973-82.
90. Yang, M.S. and M. Gill, *A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence.* International Journal of Developmental Neuroscience, 2007. **25**(2): p. 69-85.
91. Nishimura, Y., et al., *Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways.* Hum Mol Genet, 2007. **16**(14): p. 1682-98.

92. Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5116-21.
93. Benita, Y., et al., *Gene enrichment profiles reveal T cell development, differentiation and lineage specific transcription factors including ZBTB25 as a novel NF-AT repressor.* Blood.
94. Takematsu, H., et al., *Lysosomal and cytosolic sialic acid 9-O-acetylesterase activities can Be encoded by one gene via differential usage of a signal peptide-encoding exon at the N terminus.* J Biol Chem, 1999. **274**(36): p. 25623-31.
95. Katayama, T., et al., *Role of ARF4L in recycling between endosomes and the plasma membrane.* Cell Mol Neurobiol, 2004. **24**(1): p. 137-47.
96. Lencz, T., et al., *Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia.* Proc Natl Acad Sci U S A, 2007. **104**(50): p. 19942-7.
97. Matousek, S.B., et al., *Cyclooxygenase-1 mediates prostaglandin E elevation and contextual memory impairment in a model of sustained hippocampal interleukin-1beta expression.* J Neurochem.
98. Deininger, M.H., et al., *Cyclooxygenase-1 and -2 in brains of patients who died with sporadic Creutzfeldt-Jakob disease.* J Mol Neurosci, 2003. **20**(1): p. 25-30.
99. Parfenova, H., et al., *COX-1 and COX-2 contributions to basal and IL-1 beta-stimulated prostanoid synthesis in human neonatal cerebral microvascular endothelial cells.* Pediatr Res, 2002. **52**(3): p. 342-8.
100. Bani-Yaghoub, M., et al., *Neuroregenerative strategies in the brain: emerging significance of bone morphogenetic protein 7 (BMP7).* Biochem Cell Biol, 2008. **86**(5): p. 361-9.
101. Chen, D., M. Zhao, and G.R. Mundy, *Bone morphogenetic proteins.* Growth Factors, 2004. **22**(4): p. 233-41.
102. Christian, J.L., *BMP, Wnt and Hedgehog signals: how far can they go?* Curr Opin Cell Biol, 2000. **12**(2): p. 244-9.
103. Hong, C.C. and P.B. Yu, *Applications of small molecule BMP inhibitors in physiology and disease.* Cytokine Growth Factor Rev, 2009. **20**(5-6): p. 409-18.
104. von Bubnoff, A. and K.W. Cho, *Intracellular BMP signaling regulation in vertebrates: pathway or network?* Dev Biol, 2001. **239**(1): p. 1-14.
105. Dudley, A.T. and E.J. Robertson, *Overlapping expression domains of bone morphogenetic protein family members potentially account for limited tissue defects in BMP7 deficient embryos.* Dev Dyn, 1997. **208**(3): p. 349-62.
106. Helder, M.N., et al., *Bone morphogenetic protein-7 (osteogenic protein-1, OP-1) and tooth development.* J Dent Res, 1998. **77**(4): p. 545-54.
107. Perides, G., et al., *Regulation of neural cell adhesion molecule and L1 by the transforming growth factor-beta superfamily. Selective effects of the bone morphogenetic proteins.* J Biol Chem, 1994. **269**(1): p. 765-70.
108. Lee-Hoeflich, S.T., et al., *Activation of LIMK1 by binding to the BMP receptor, BMPRII, regulates BMP-dependent dendritogenesis.* Embo J, 2004. **23**(24): p. 4792-801.
109. Podkowa, M., et al., *Microtubule stabilization by bone morphogenetic protein receptor-mediated scaffolding of c-Jun N-terminal kinase promotes dendrite formation.* Mol Cell Biol. **30**(9): p. 2241-50.

110. Esquenazi, S., et al., *BMP-7 and Excess Glutamate: Opposing Effects on Dendrite Growth from Cerebral Cortical Neurons in Vitro.* Experimental Neurology, 2002. **176**(1): p. 41-54.
111. de Rivero Vaccari, J.P., et al., *Neuroprotective effects of bone morphogenetic protein 7 (BMP7) treatment after spinal cord injury.* Neurosci Lett, 2009. **465**(3): p. 226-9.
112. Yan, W. and X. Chen, *Targeted repression of bone morphogenetic protein 7, a novel target of the p53 family, triggers proliferative defect in p53-deficient breast cancer cells.* Cancer Res, 2007. **67**(19): p. 9117-24.
113. Hashimoto, K., et al., *Reduced serum levels of brain-derived neurotrophic factor in adult male patients with autism.* Prog Neuropsychopharmacol Biol Psychiatry, 2006. **30**(8): p. 1529-31.
114. Takahashi, M., et al., *Abnormal expression of brain-derived neurotrophic factor and its receptor in the corticolimbic system of schizophrenic patients.* Mol Psychiatry, 2000. **5**(3): p. 293-300.
115. Zorner, B., et al., *Forebrain-specific trkB-receptor knockout mice: behaviorally more hyperactive than "depressive".* Biol Psychiatry, 2003. **54**(10): p. 972-82.
116. Rackova, S., L. Janu, and H. Kabickova, *Borna disease virus circulating immunocomplex positivity and psychopathology in psychiatric patients in the Czech Republic.* Neuro Endocrinol Lett, 2009. **30**(3): p. 414-20.
117. Lebain, P., et al., *Borna disease virus and psychiatric disorders.* Schizophr Res, 2002. **57**(2-3): p. 303-5.
118. Pletnikov, M.V., et al., *Developmental brain injury associated with abnormal play behavior in neonatally Borna disease virus-infected Lewis rats: a model of autism.* Behav Brain Res, 1999. **100**(1-2): p. 43-50.
119. Hans, A., et al., *Persistent, noncytolytic infection of neurons by Borna disease virus interferes with ERK 1/2 signaling and abrogates BDNF-induced synaptogenesis.* Faseb J, 2004. **18**(7): p. 863-5.
120. Planz, O., S. Pleschka, and T. Wolff, *Borna disease virus: a unique pathogen and its interaction with intracellular signalling pathways.* Cell Microbiol, 2009. **11**(6): p. 872-9.
121. Nam, D. and S.Y. Kim, *Gene-set approach for expression pattern analysis.* Brief Bioinform, 2008. **9**(3): p. 189-97.
122. Khatri, P. and S. Draghici, *Ontological analysis of gene expression data: current tools, limitations, and open problems.* Bioinformatics, 2005. **21**(18): p. 3587-95.
123. Ganter, B., et al., *Pathway analysis tools and toxicogenomics reference databases for risk assessment.* Pharmacogenomics, 2008. **9**(1): p. 35-54.
124. Kahlem, P. and E. Birney, *Dry work in a wet world: computation in systems biology.* Mol Syst Biol, 2006. **2**: p. 40.
125. Chang, J.T. and J.R. Nevins, *GATHER: a systems approach to interpreting genomic signatures.* Bioinformatics, 2006. **22**(23): p. 2926-33.
126. Ripamonti, U., et al., *Transforming growth factor-beta isoforms and the induction of bone formation: implications for reconstructive craniofacial surgery.* J Craniofac Surg, 2009. **20**(5): p. 1544-55.
127. Baar, K., et al., *Skeletal muscle overexpression of nuclear respiratory factor 1 increases glucose transport capacity.* Faseb J, 2003. **17**(12): p. 1666-73.

128. Patti, M.E., et al., *Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8466-71.
129. Chang, W.T., et al., *A novel function of transcription factor alpha-Pal/NRF-1: increasing neurite outgrowth.* Biochem Biophys Res Commun, 2005. **334**(1): p. 199-206.
130. Floricel, F., et al., *Antisense suppression of TSC1 gene product, hamartin, enhances neurite outgrowth in NGF-treated PC12h cells.* Brain Dev, 2007. **29**(8): p. 502-9.
131. Hattori, T., et al., *DISC1 regulates cell-cell adhesion, cell-matrix adhesion and neurite outgrowth.* Mol Psychiatry.
132. Lepagnol-Bestel, A.M., et al., *SLC25A12 expression is associated with neurite outgrowth and is upregulated in the prefrontal cortex of autistic subjects.* Mol Psychiatry, 2008. **13**(4): p. 385-97.
133. Piton, A., et al., *Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia.* Mol Psychiatry.
134. Bossone, S.A., et al., *MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination.* Proc Natl Acad Sci U S A, 1992. **89**(16): p. 7452-6.
135. Shen, Y., et al., *Activation of the Jnk signaling pathway by a dual-specificity phosphatase, JSP-1.* Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13613-8.
136. Ballif, B.C., et al., *Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2.* Nat Genet, 2007. **39**(9): p. 1071-3.
137. Miller DT, N.R., Sobeih MM, Shen Y, Wu BL, Hundley R, Hanson E., *16p11.2 Deletion Syndrome.* Gene Reviews, 2009.
138. Bailey, J.A., et al., *Segmental duplications: organization and impact within the current human genome project assembly.* Genome Res, 2001. **11**(6): p. 1005-17.
139. Reardon, W., *The Bedside Dysmorphologist Classic Clinical Signs in Human Malformation Syndromes and their diagnostic Significance*. 2008: Oxford University Press.
140. Chan, G. and J.P. Dormans, *Update on congenital spinal deformities: preoperative evaluation.* Spine (Phila Pa 1976), 2009. **34**(17): p. 1766-74.
141. Stephanie Sacharow1, B.X., Klaas Wierenga2, Paul Benke3, Deborah Barbouth1, Parul Jayakar1,4. *16p11.2 Microdeletion Syndrome case series, including parent-to-child transmission, endocrinologic*
*disorder, and an anephric infant*. in *American Clinical Genetics Meeting*. 2009. Tampa Florida.
142. Perez-Molina, J.J., et al., *[Polydactyly in 26,670 consecutive births. The clinical characteristics, prevalence and risk factors]*. Bol Med Hosp Infant Mex, 1993. **50**(11): p. 803-8.
143. Mupparapu, M., R.E. Binder, and F. Duarte, *Hereditary cranium bifidum persisting as enlarged parietal foramina (Catlin marks) on cephalometric radiographs.* American Journal of Orthodontics and Dentofacial Orthopedics, 2006. **129**(6): p. 825-828.
144. MacKenzie, B., et al., *Twisted gastrulation limits apoptosis in the distal region of the mandibular arch in mice.* Dev Biol, 2009. **328**(1): p. 13-23.

145. Cooper-Brown, L., et al., *Feeding and swallowing dysfunction in genetic syndromes.* Dev Disabil Res Rev, 2008. **14**(2): p. 147-57.
146. Olasoji, H.O., P.J. Ambe, and O.A. Adesina, *Pierre Robin syndrome: an update.* Niger Postgrad Med J, 2007. **14**(2): p. 140-5.
147. Hogan, B.L., *Bone morphogenetic proteins: multifunctional regulators of vertebrate development.* Genes Dev, 1996. **10**(13): p. 1580-94.
148. Maxson, R. and M. Ishii, *The Bmp pathway in skull vault development.* Front Oral Biol, 2008. **12**: p. 197-208.
149. Ishii, M., et al., *Msx2 and Twist cooperatively control the development of the neural crest-derived skeletogenic mesenchyme of the murine skull vault.* Development, 2003. **130**(24): p. 6131-42.
150. Wang, Y.H., et al., *Effects of BMP-7 on mouse tooth mesenchyme and chick mandibular mesenchyme.* Dev Dyn, 1999. **216**(4-5): p. 320-35.
151. Murtaugh, L.C., J.H. Chyung, and A.B. Lassar, *Sonic hedgehog promotes somitic chondrogenesis by altering the cellular response to BMP signaling.* Genes Dev, 1999. **13**(2): p. 225-37.
152. Krzemien, G., et al., *Urological anomalies in children with renal agenesis or multicystic dysplastic kidney.* J Appl Genet, 2006. **47**(2): p. 171-6.
153. Schumacher, V., et al., *Correlation of germ-line mutations and two-hit inactivation of the WT1 gene with Wilms tumors of stromal-predominant histology.* Proc Natl Acad Sci U S A, 1997. **94**(8): p. 3972-7.
154. Ton, C.C., et al., *Smallest region of overlap in Wilms tumor deletions uniquely implicates an 11p13 zinc finger gene as the disease locus.* Genomics, 1991. **10**(1): p. 293-7.
155. Varanasi, R., et al., *Fine structure analysis of the WT1 gene in sporadic Wilms tumors.* Proc Natl Acad Sci U S A, 1994. **91**(9): p. 3554-8.
156. Haber, D.A., et al., *An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor.* Cell, 1990. **61**(7): p. 1257-1269.
157. Fernandez, B.A., et al., *Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder.* J Med Genet. **47**(3): p. 195-203.
158. Davies, A.M., H. Thoenen, and Y.A. Barde, *The response of chick sensory neurons to brain-derived neurotrophic factor.* J Neurosci, 1986. **6**(7): p. 1897-904.
159. Lu, B. and K. Martinowich, *Cell biology of BDNF and its relevance to schizophrenia.* Novartis Found Symp, 2008. **289**: p. 119-29; discussion 129-35, 193-5.
160. Pillai, A., *Brain-derived neurotropic factor/TrkB signaling in the pathogenesis and novel pharmacotherapy of schizophrenia.* Neurosignals, 2008. **16**(2-3): p. 183-93.
161. Tsai, S.J., *Is autism caused by early hyperactivity of brain-derived neurotrophic factor?* Med Hypotheses, 2005. **65**(1): p. 79-82.
162. Tseng, M., et al., *BDNF protein levels are decreased in transformed lymphoblasts from lithium-responsive patients with bipolar disorder.* J Psychiatry Neurosci, 2008. **33**(5): p. 449-53.

163. Samuels, I.S., S.C. Saitta, and G.E. Landreth, *MAP'ing CNS development and cognition: an ERKsome process.* Neuron, 2009. **61**(2): p. 160-7.
164. Angelucci, F., S. Brene, and A.A. Mathe, *BDNF in schizophrenia, depression and corresponding animal models.* Mol Psychiatry, 2005. **10**(4): p. 345-52.
165. Barnby, G., et al., *Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT.* Am J Hum Genet, 2005. **76**(6): p. 950-66.
166. White, P.H. and D.L. Chapman, *Dll1 is a downstream target of Tbx6 in the paraxial mesoderm.* Genesis, 2005. **42**(3): p. 193-202.
167. Chapman, D.L., et al., *Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development.* Dev Dyn, 1996. **206**(4): p. 379-90.
168. Li, Q.Y., et al., *Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family.* Nat Genet, 1997. **15**(1): p. 21-9.
169. Merscher, S., et al., *TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome.* Cell, 2001. **104**(4): p. 619-29.
170. Ensenauer, R.E., et al., *Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients.* Am J Hum Genet, 2003. **73**(5): p. 1027-40.