

THE PENNSYLVANIA STATE UNIVERSITY

SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

RAMIFICATIONS OF COPY NUMBER VARIANTS IN AUTISM

SU JEN KHOO

MAY 2012

A thesis
submitted in partial fulfillment
of the requirements for a baccalaureate degree
in Biotechnology with honors in Biotechnology

Reviewed and approved by the following:

Maria Krasilnikova
Research Assistant Professor of Biochemistry and Molecular Biology
Thesis Supervisor

Scott B. Selleck
Professor and Head
Department of Biochemistry and Molecular Biology
Thesis Co-Advisor

David Gilmour
Professor of Molecular and Cell Biology
Honors Advisor

Wendy Hanna-Rose
Associate Department Head for Undergraduate Studies
Department of Biochemistry and Molecular Biology

*Signatures are on file with the Schreyer Honors College

ABSTRACT

Autism has emerged as an important childhood concern regarding one's social and cognitive developmental condition. To date, many uncertainties manifest when questions are raised regarding one single explanation behind this developmental impairment. Various resources have emphasized the roles of environment [10, 83], epigenetics [83] and genetics [3, 4, 6] in determining the potential risk for autism. Symptoms and degree of penetrance of autism varies in every child that is observed in terms of social communication impairments [89], repetitive behavioral patterns [19] and restrained interests [19, 89]. Recent interest in this area of research has inspired collaboration between the Biochemistry and Molecular Biology Department in Pennsylvania State University, Department of Genome Sciences from University of Washington and the Medical Investigation of Neurodevelopmental Disorders (MIND) team from University California Davis. The collaborative work has incorporated a combination of technological advances that include detection of copy number variants (CNVs), implementing high density customized oligonucleotide arrays and clinical assessments to elicit better understanding regarding factors that can lead to autism. The study is comprised of conducting microarray scans on 554 blood DNA samples provided by the MIND Institute, in which 274 samples were obtained from autistic individuals while another 280 originated from normally developing individuals. Results obtained from the microarray analysis enabled investigators to compare the structural variants of individuals and detect dysregulation of genes in respective chromosomal regions that could cause neurological disorders [69]. Additional clinical assessments, primarily the Mullen Scales of Learning and Vineland Adaptive Behavior Scales provided the assessment of individuals' susceptibility to neurodevelopmental disorders based on overt phenotypes shown in their socialization and living skills [25]. Eleven significant cases of pathogenic deletion and duplication events found in autistic and normally developing individuals were discussed in this dissertation.

These events constituted large affected chromosomal areas in the genome and were further analyzed to cause a loss or gain of genes involved in neurodevelopmental pathways that could possibly lead to autism.

TABLE OF CONTENTS

| | |
|---|-----|
| Abstract..... | i |
| List of Figures..... | v |
| List of Tables..... | vi |
| Abbreviations..... | vii |
| Acknowledgements..... | ix |
| Chapter 1. INTRODUCTION..... | 1 |
| Chapter 2. MATERIALS AND METHODS..... | 5 |
| Ethics statement..... | 5 |
| Sample collection and ascertainment..... | 5 |
| Design of array Comparative Genomic Hybridization (aCGH)..... | 7 |
| Labelling and hybridization..... | 8 |
| Analysis of the aCGH..... | 9 |
| Chapter 3. RESULTS..... | 12 |
| Novel, rare CNV events found in autistic individuals..... | 18 |
| Novel CNV events in the typically developing group..... | 23 |
| Significant CNV events in both autistic and typically developing cohort..... | 25 |
| Chapter 4. DISCUSSION..... | 30 |
| A novel deletion in chromosome region 4q32.3 detected in an autistic individual..... | 30 |
| An atypical duplication at region of 4q13.1 in autistic individual of Asian ethnicity..... | 31 |
| Unique region of duplication at 6p22.31 identified in autistic individual..... | 31 |
| 1q21.1 deletion detected in autistic individual..... | 32 |

| | |
|--|----|
| Deletion events at region 7q11.22-23 in autistic and typically developing individuals..... | 32 |
| Significant 15q11.2 duplication in three autistic individuals..... | 33 |
| Novel deletion at region of 6q23.2 detected in a typically developing subject..... | 33 |
| Atypical deletion of 7q11.23 encountered in a typically developing subject..... | 34 |
| Duplication event at 22q11.22 in a typically developing individual..... | 34 |
| Occurrence of duplications involving 15q13.3 in autistic and typically developing individuals..... | 35 |
| Deletion and duplication events at 16p11.2 occurred in both autistic and typically developing individuals..... | 35 |
| Chapter 5. CONCLUSIONS AND FUTURE DIRECTIONS..... | 37 |
| References..... | 39 |
| Academic vitae..... | 47 |

LIST OF FIGURES

| | | |
|------------------|---|----|
| Figure 1 | A Scheme of the sample labeling and hybridization used in array experiments | 9 |
| Figure 2 | Detailed explanation of figure output utilized for analysis obtained from the UCSC genome browser | 11 |
| Figure 3 | A large deletion of 5Mbp at 4q32.3 in AUMA_14 sample | 18 |
| Figure 4 | A 2Mbp duplication at 14q13.1 in AU_AS_13 sample | 19 |
| Figure 5 | A 2Mbp duplication at 6q22.31 in AU_W_67 sample | 20 |
| Figure 6 | A 1Mbp deletion at 1q21.1 in AU_W_01 sample | 21 |
| Figure 7 | A 2Mbp deletion at 7q11.22 in AU_W_118 sample | 22 |
| Figure 8 | A 5Mbp duplication at 15q11.2 in AU_MIX_10 sample | 22 |
| Figure 9 | A 1Mbp deletion at 6q23.2 in TD_W_33 sample | 23 |
| Figure 10 | A 1Mbpdeletion at 7q11.23 in TD_W_104 sample | 24 |
| Figure 11 | A 1.5Mbp duplication at 22q11.22 in TD_W_23 sample | 25 |
| Figure 12 | A 500kbp duplication at 16p11.2 in AU_MIX_04 sample | 26 |
| Figure 13 | Deletion and duplication events found respectively in TD_AA_05 and TD_W_29 | 26 |
| Figure 14 | a) A 500kbp duplication in AUMA_060 at 15q13.3 b) A 500kbp deletion is observed in AU_W_28 at 15q3.3 | 28 |
| Figure 15 | a) A 500kbp duplication in TDMA_55 at 15q13.3 b) A 500kbp duplication in TDMA_71 at 15q13.3 | 29 |

LIST OF TABLES

| | | |
|----------------|---|----|
| Table 1 | Summary of the CHARGE study showing distribution of samples based on diverse ethnicity backgrounds | 7 |
| Table 2 | The ethnical distribution of the AU and TD samples that passed QC and HMM measures | 12 |
| Table 3 | Summary of events found for both AU and TD samples | 14 |
| Table 4 | Summary of Mullen Early Scales of Learning and Vineland Adaptive Behavior scores of the autistic (AU) and typically developing (TD) individuals | 16 |

ABBREVIATIONS

| | |
|--------|---|
| aCGH | Array comparative genomic hybridization |
| ADHD | Attention deficit hyperactivity disorder |
| ADI-R | Autism Diagnostic Interview Revised |
| ADOS | Autism Diagnostic Observation Schedule |
| APP | Amyloid protein precursor |
| ASD | Autism spectrum disorder |
| AU | Autistic |
| AU_AA | Autistic individual of African American ethnicity |
| AU_AS | Autistic individual of Asian ethnicity |
| AU_MIX | Autistic individual of Mixed race ethnicity |
| AU_W | Autistic individual of Caucasian ethnicity |
| AUMA | Autistic individual of Hispanic ethnicity |
| CHARGE | Childhood Autism Risks from Genetics and Environment |
| CNV | Copy number variant |
| DNA | Deoxyribonucleic acid |
| DSM-IV | Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition |
| GTPase | Guanosine triphosphate hydrolase |
| HMM | Hidden Markov Model |
| HMMSeg | Hidden Markov Model segmentation |
| IQ | Intelligence quotient |
| MIND | Medical Investigation of Neurodevelopmental Disorders |
| nAChR | neuronal nicotinic acetylcholine receptor |
| NAHR | Non-allelic homologous recombination |
| ODDD | Oculodentodigital dysplasia |
| QC | Quality control |

| | |
|---------|---|
| Ras | Rat sarcoma protein |
| Rho | Rhodopsin |
| SNP | Single nucleotide polymorphism |
| TAR | Thrombocytopenia with absent radius syndrome |
| TD | Typically developing |
| TD_AA | Typically developing individual of African American ethnicity |
| TD_AS | Typically developing individual of Asian ethnicity |
| TD_MIX | Typically developing individual of Mixed race ethnicity |
| TD_W | Typically developing individual of Caucasian ethnicity |
| TDMA | Typically developing individual of Hispanic ethnicity |
| UCSC | University of California, Santa Cruz |
| z-score | Standard score |

ACKNOWLEDGEMENTS

First and foremost, I would like to extend my gratitude and appreciation to Dr. Maria Krasilnikova, my thesis advisor for her never-failing support and determination. For two years, Dr. Krasilnikova has patiently guided me in conducting various laboratory skills in molecular biology that encapsulated basic bench work such as bacteria transformation and gel electrophoresis to more convoluted computational analysis work. Thank you for the confidence and courage you have shown and given me especially for participation in various scientific seminars and undergraduate exhibitions. I am very thankful for your advices and encouragement and deeply respect your willingness to devote time and effort for our projects.

Special thanks to Dr. Scott B. Selleck, our Professor and Head of Biochemistry and Molecular Biology Department for providing me an utmost opportunity to learn and work as an intern in a collaborative autism project between Pennsylvania State University, University of California, Davis and University of Washington. I deeply appreciate your willingness and enthusiasm to guide me in various analytical work.

Sincere appreciation goes to Dr. David Gilmour, my honors advisor for his support, guidance and assistance with my studies and thesis completion.

I would like to extend cordeal acknowledgement to Dr. Santhosh Girirajan and Carl Baker from the Genome Science Department at University of Washington for their endless devotions in guiding me for various laboratory, analytical and presentation skills during the internship.

I wish to thank you Dr. Evan Eichler and his laboratory team members for welcoming me to their laboratory during the summer internship and willingness to guide and assist me in adapting to the working environment in University of Washington.

Chapter 1. INTRODUCTION

Tremendous increase of exposure to novel scientific discovery has drawn more attention from parents about the influence of genetic predisposal and environmental effects towards the social and cognitive development of their children. Recent active area of neurodevelopmental studies has raised the issue of autism that leaves researchers with many questions about the role of nature and nurture. Autism, usually apparent by the age of 3 years [50], is categorized with three main symptoms that include ritualistic repetitive behavior, impaired social interaction and impaired communication as well as language skills [19]. Autism cases are diagnosed in 3 per 1000 individuals a year and the number rises to 6 per 1000 if all forms of autism spectrum disorders (ASD) are included. The risk of autism is relatively higher in males than females, approximately 4:1 ratio. Twin and family studies indicate greater impact of genetic basis for autism; recurrence risk for siblings to have autism comes to an account of 5-10% [49, 50, 58].

Several types of evidence indicate that genetic variation contributes significantly to autism. Karyotyping methods led to detection of 5-7% of autism cases based on large scale chromosomal anomalies [19, 57]. The emergence of DNA microarray technology has enabled detection of intermediate sized copy number variants. These intermediate-sized variants that include events of duplications and deletions in the human genome are termed copy number variants (CNV). This approach has been statistically and scientifically proven to detect more genetic variants among individuals, resulting in 100-fold difference when compared with single nucleotide polymorphisms (SNP) [3]. However, not all CNVs cause genomic disorders. Many common ones are also observed in healthy individuals and in fact, found to be contributing factors in studying human genome evolution [69]. Most cases of CNVs occur due to nonallelic homologous recombination (NAHR) between segmental duplications that are of 95 percent sequence identity. Genomic disorders such as Williams syndrome [54],

Smith-Magenis syndrome [9] and DiGeorge syndrome [60] caused by these structural variants are of high penetrance, recurrent, and affect multiple unrelated individuals of diverse haplotype backgrounds [3]. Narrowing studies to genomic hotspots caused by the recurrent NAHR rearrangements increases the efficiency of discovering both *de novo* and inherited CNVs in many neurodevelopmental-related cases, to name a few, autism, schizophrenia [9], bipolar disorder [16, 22], and attention deficit hyperactivity disorder (ADHD) [3]. The analysis of genomic backbones is also important because some CNVs can possibly arise from nonrecurrent mechanisms that encompass microhomology/microsatellite mediated break-induced repair, nonhomologous end joining and template switching [59]. Significant nonrecurrent events in *SHANK2*, *SHANK3* and *PTCHD1* genes have been previously linked to autism; deletions and duplications in *ASTN2* and *NRXN1* were linked to autism and schizophrenia [1, 3].

Many research groups have recently focused on identification of enriched gene sets and new candidate pathways leading to autism and ASD [16,22]. Alteration of dosage sensitive genes in chromosomal regions previously linked to autism and schizophrenia led to changes in distribution of cortical neuron projections, neuronal stem cells and basal progenitors as well as reduction of cortical gray matter based on detection from brain imaging and postmortem analysis [60]. Zikopoulos and Barbas (2010) discussed that a range of axonal growth projections were shorter and excessively connected to neighboring cells in autistic individuals as detected from brain imaging analysis [66]. Disconnection of axonal projections and excessive connectivity were further found to disrupt emotional pathways in prefrontal cortices [66, 72]. Such mechanism was relevant to the observable phenotype of autistic individuals that included attention deficiency, repetitive behavior and avoidance of social interactions [19, 66].

There were also connections made between enriched CNV density found in regions that potentially disrupt gene sets involved in cellular proliferation, projection, motility, vesicular transport and GTPase/Ras signaling [7, 60, 67]. The Rho GTPases are important in regulating dendrite and spine plasticity and have been previously associated with intellectual disability (ID) [61, 62]. Co-expression analysis further showed that gene expression in a fetal human brain of an autistic individual differed from the brain of a normal developing person by down regulation of neuronal pathways. In contrast to low expression of genes involved in synaptic function, there was a significant increase in expression of autism susceptibility genes that included *CADPS2*, *AH11*, *CNTNAP2* and *SLC25A12* [67]. Transcriptome analysis also showed that heightened immunological response was detected in the brain of autistic individuals, signifying possibility of excessive microglial activity due to an abnormal neural plasticity [67,68]. It was postulated that phenotypic diversity observed in autistic individuals that included varying degrees of social impairment, communication barriers and repetitive behaviors were caused by non-overlapping genes associated with different brain regions [63]. Previous research studies that involved integrating autism and brain pathology have shown that autistic individuals were characterized with a heightened brain growth in early childhood [84], curtailed neuronal development [85], unusual brain cytoarchitecture [85], altered metabolic functions from amyloid protein precursor (APP) [86] and increased cell organelles turnover as well as glial activation [87]. It has also been recently discovered that genes that may be associated with ASD and intellectual disability caused adverse effects on microtubule cytoskeleton, glycosylation and central nervous system development as well as adhesion [61, 63, 64]. Whole-genome CNV studies led to discovery of autism susceptibility genes such as *NRXN1* and *CNTN4* that encoded the neuronal cell-adhesion molecules and *UBE3A*, *PARK2* and *FBXO40* involved in the ubiquitin pathways [65]. Rare CNVs in regions associated with specific genes provide plausible indications that genes involved in neuronal cell-adhesion and

ubiquitin degradation in the central nervous system contribute to genetic susceptibility to ASD including autism [65].

Apart from that, it also appears that genomic instability has been linked to neuropsychiatric disorders that include schizophrenia and autism [70, 71]. An interesting study has pointed out that genomic instability that included chromosomal instabilities, and base pair alterations showed increased tendency of occurrence at fragile sites, which are also cancer-prone regions [70].

Technology advances, such as the methods based on high-density oligonucleotide arrays, provided an avenue to discovering structural variants caused mainly by segmental duplications. Many CNVs situated in gene-rich regions can potentially result in functional consequences, such as gene dosage alterations, disruption of genes as well as modulation of gene activity [69]. These genetic variations further affect susceptibility of an individual to diseases, drug response, and sometimes impede human genome evaluation [6, 69]. This research was conducted to focus on deciphering significant *de novo* structural variants that arise mainly from recurrent NAHR events in genomic hotspot regions with the purpose of finding autism susceptible genes and association of these genes with other forms of neuropsychiatric disorders.

Chapter 2. MATERIALS AND METHODS

Ethics statement

DNA blood samples from patients of University California Davis Medical Investigation of Neurodevelopmental Disorders (MIND) Institute for study cohort were collected for the purpose of genetic studies and clinical analysis upon approval and informed consent of the patients as well as their respective family members.

Sample collection and ascertainment

The collaborative study provided an assessment based on 554 blood DNA samples from the Childhood Autism Risks from Genetics and Environment (CHARGE) [10] initiated by the MIND Institute. The CHARGE study highlighted a wide range of chemical and biological exposures, susceptibility factors and interactions between these factors. Data collection by the CHARGE study cohort included prudent developmental assessments, medical information, questionnaire data and biological specimens provided by the University of California Davis Center for Children's Environmental Health laboratories [10]. The autistic group consisted of 274 individuals while the control group consisted of 280 individuals, both of various ethnicity backgrounds. The criteria for selecting blood DNA samples of autistic (AU) individuals required that these children conform to the ASD prototype using Autism Diagnostic Observation Schedule (ADOS) [15], Autism Diagnostic Interview Revised (ADI-R) and further clinical diagnosis judgement by the staff of CHARGE study [11]. The nonverbal IQ scores of the autistic sample cohort were required to be above 35. Children that were found to have significant hearing or vision difficulties, motor problems and birth complications as well as Autism Spectrum Disorder (ASD) related characteristics such as Fragile X syndrome [12] were excluded from the study. Individuals with relatives diagnosed with ASD or siblings with ASD symptoms were also excluded from

the analysis. Standard data collection, entry and validation methods were used for the diagnosis to assess the cognitive development and perform the phenotypic categorization to enhance the reliability of sample collection method [5].

The control cohort was selected from the general population that did not meet the DSM-IV criteria for major depression, bipolar disorder and psychosis [5]. In addition, the study included CNV data from 8329 extra cell line and blood-derived controls to evaluate the frequency of recurring pathogenic CNVs in broader domain of neurologically normal developing individuals [5]. The CNV data for these controls originated from genome-wide association studies in which individuals provided informed consent but were not diagnosed with neurological-related phenotypes [5].

Mullen Scales of Early Learning and Vineland Adaptive Behavior Scales were used to assess the children on different domains including communication, daily living skills, socialization, motor skills and language expression [25]. According to our evaluations and study design, patients with average Mullen and Vineland scores below 85 were considered having mild or severe symptoms of autism or ASD. It is important to note that the control group included some typically developing (TD) individuals that had mild symptoms of mental retardation (MR) or developmental delay (DD) in spite of their normal cognitive performance had only slightly decreased Mullen and Vineland scores.

The 274 AU individuals encompassed diverse ethnicities of 143 Caucasian, 74 Hispanics, 30 Mixed Race, 12 Asians, and six African Americans. For the control group, the 280 TD individuals consisted of 147 Caucasians, 83 Hispanics, 37 Mixed Race, seven Asians, five African American, and one Pacific Islander/ Hawaiian Native. Table 1 illustrated the ethnical background of the patients included in CHARGE study cohort for AU and TD samples used in this analysis.

| Ethnicity | Typically developing (TD) | Autistic (AU) | Total Samples |
|----------------------|--------------------------------------|--------------------------|----------------------|
| Hispanic | 83 | 74 | 157 |
| Caucasian | 147 | 143 | 290 |
| Asian | 7 | 12 | 19 |
| African American | 5 | 6 | 11 |
| Mixed Race | 37 | 30 | 67 |
| Native Hawaiian | 1 | 0 | 1 |
| Total samples | 280 | 274 | 554 |

Table 1: Summary of the CHARGE study showing distribution of samples based on diverse ethnicity backgrounds.

Design of array Comparative Genomic Hybridization (aCGH)

CNVs were detected using a customized design of the hotspot v1.0 array by Roche Nimblegen. The 12-plex array was tailored with 135,000 probes. 107 genomic hotspot regions were covered with higher probes density with average probe spacing of 2.6 kb. The rest of the genome (genomic backbone) was covered with the probes of lower density with an average spacing of 36 kb. Another hotspot v3.1 array was also designed to validate the samples that showed large CNV loads in the Roche Nimblegen 12-plex array described above. Validation was conducted using a custom-designed 4x180K Agilent chip provided by Agilent Technologies with the median probe spacing of 2 kb in genomic hotspot regions and 36 kb in whole-genome backbone regions [5].

Labelling and hybridization

The Nimblegen array CGH kit (Roche Nimblegen Inc.) was used to label the DNA samples analyzed using the hotspot v1.0 arrays and hotspot v3.1 arrays prior to hybridization. Procedures of labelling and hybridization were carried out according to manufacturer's protocols. All microarray hybridization experiments were conducted using a single unaffected male (GM15724 from Coriell) as a reference DNA sample. The AU and TD DNA samples were differentiated from the reference DNA using a different fluorescent dye color. Cy3 dye was used to label the AU and TD samples while Cy5 dye was utilized to label the reference DNA sample. Both AU and TD samples (labelled with Cy3) and the reference DNA sample (labelled with Cy5) were then mixed and hybridized to the Nimblegen hotspot v1.0 arrays using the Nimblegen 12 Bay Hybridization System (Roche Nimblegen) according to the manufacturer's instructions. For validation purposes, hybridization was also conducted by placing hybridization slides in Agilent SureHyb chamber (Agilent Technologies). Further steps were performed according to the manufacturer's protocol. The hybridized and washed slides from both hotspot v1.0 arrays and hotspot v3.1 arrays were then scanned using GenePix 400B Microarray Scanner (Molecular Devices) based on the specific scanning and feature extraction settings provided by the company.

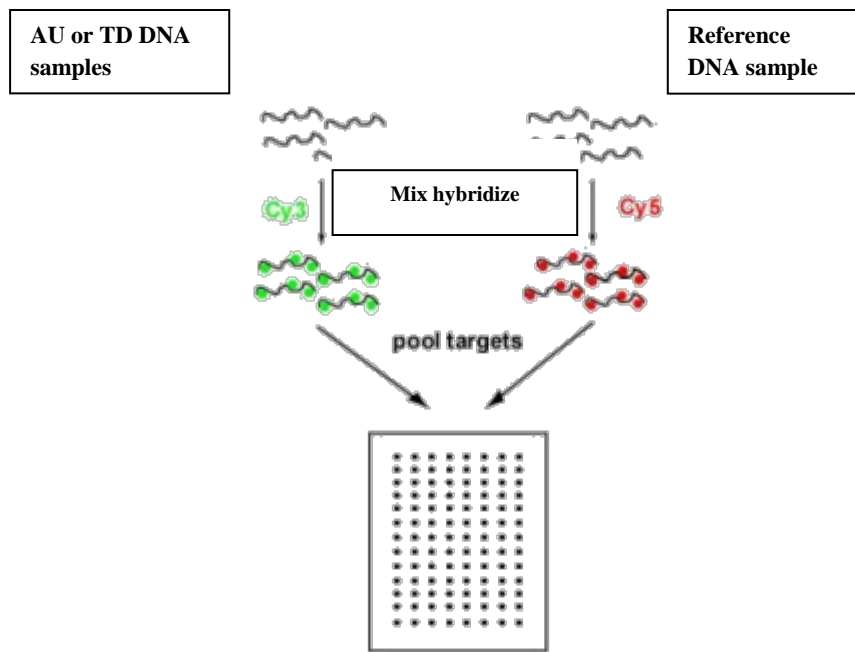


Figure 1. A Scheme of the sample labelling and hybridization used in array experiments. AU and TD samples were labeled with Cy3 and reference DNA sample (GM15724 from Coriell) was labelled with Cy5. The DNA samples of the subjects tested and the reference DNA samples were then mixed and hybridized to the Nimblegen 12 x 135K array (Roche Nimblegen) for the initial analysis, and 4 x 180K Agilent Chips (Agilent Technologies) for validation.

Analysis of the aCGH

The University of California, Santa Cruz (UCSC) genome browser was used for CNVs' detection. Analyses of all arrays were conducted by mapping the probe coordinates to the human genome assembly Build 36 (hg18 array) . Specific means, standard deviations and log intensity ratios were translated into z-scores to classify the copy number states. The HMMSeg was used to segment continuous genomic datasets based on the Hidden Markov Model (HMM) [17]. This process enabled the identification of significant CNVs in the DNA samples and rule out small events or events in segmental duplicated regions. Stringent quality control (QC) measures were also applied after selection was done using HMM.

The QC criteria included events with absolute z-score of more than 1.5 that contained more than 10 probes [5]. Manual curation which involved a manual detection of the array

output of samples from the genome browser Build36 (hg18) based on the normalized log intensity ratio, was also conducted to further identify events that were missed and not detected using the HMM application. An example of an output for the manual curation purposes obtained from UCSC genome portal is shown and described in Figure 2. The CNVs detected using Hotspot 1.0 array were shown in the last two lanes by green (duplication) or red bars (deletions). The heights of the bars represented the z value that reflected the strength of the aCGH signal. The criteria for evaluation shown in the Figure 2 included the minimum 50 kb size of a chromosomal region covered; the absence of events in segmental duplicated regions, which are shown in gray and orange bars, less than two analogous events seen among the 8329 controls. The CNVs that overlapped with the area previously known to contain signature genetic events for neurological disorders were given special consideration. Selected AU and TD samples, validated with the Agilent arrays were analyzed using feature extraction of ADM-2 setting according to the manufacturer [5].

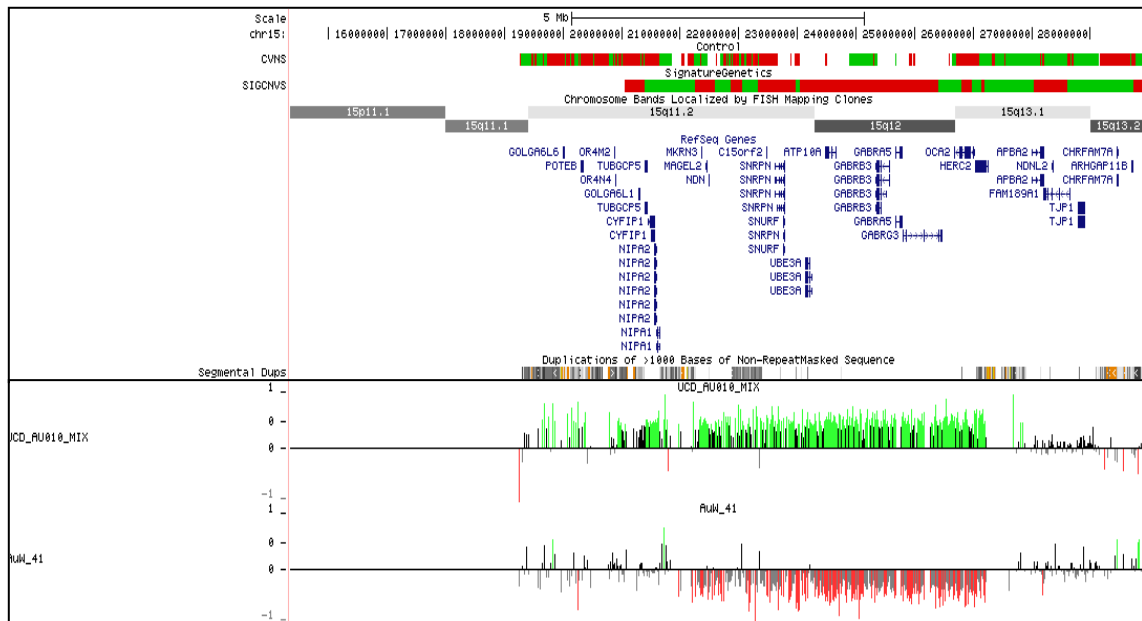


Figure 2: An example of figure output utilized for further analysis as obtained from the UCSC genome browser. The exact chromosomal location of the region and the scale are shown at the very top of the figure. Below that, the deletions or duplications in the specific loci found in the 8329 controls from the UCSC database are shown in red or green respectively. In the lane directly below that, the figure also shows the corresponding signature genetics events for the neurological disorders that were found previously. Increased number of signature genetics events and absence of control cases were preferred to define the region as being rare and significant in our intended analysis. The map below that provides information on genes located in the selected region. The next lane shows the position of the segmental duplication regions, which are indicated in grey and orange colors. The probes that were detected to be duplicated or deleted by aCGH are shown below as green and red color bars, respectively. The red or green bars are clustered together indicates a deletion or an expansion event. The criterion for the size of the region to be considered as significant was equivalent to or larger than 50kbp. Figure was obtained from the UCSC genome browser.

Chapter 3. RESULTS

Out of the 554 samples hybridized and analyzed, 499 samples (243 AU and 256TD) were confirmed to pass the HMM and QC measure with an absolute z-score of more than 1.5 and standard deviation less than 0.32 (Table 2). Analysis was carried out using hotspot 1.0 array that focused on the 107 genomic hotspot regions that contained a higher density of probes. Some of the samples were validated using the higher density hotspot 3.1 Agilent array to confirm events of duplication and deletions. Samples that contained significant CNVs and those that did not pass quality control were also analyzed again using the hotspot 1.0 Nimblegen array.

| Ethnicity | Autistic (AU) | Typically developing (TD) | Total |
|------------------|----------------------|----------------------------------|--------------|
| Hispanic | 57 | 81 | 138 |
| Caucasian | 133 | 134 | 267 |
| Mixed Race | 27 | 30 | 57 |
| Asian | 20 | 7 | 27 |
| African American | 6 | 4 | 10 |
| Total | 243 | 256 | 499 |

Table 2: The ethnical distribution of the AU and TD samples that passed QC and HMM measures.

Our study had revealed and confirmed sixteen significant events of duplications and deletions, with the majority of them found in autistic individuals. There were some significant CNV events found in AU samples that covered a region that contained candidate genes highly associated with autism. It is essential to point out that several significant, large CNV duplication and deletion events were also found in the typical developing samples. In conjunction with the results obtained, further analysis showed that these typically developing

individuals with fairly large CNV events suffered from other neurological disorders and mild autistic behavior.

A summary of all the events including the type of copy number variant, chromosomal size band, race, ethnicity and significant genes involved in the region, is presented in the Table 3. The table also shows amount of analogous events that could be detected in comparison to a expanded map developed from a set of 8329 normal individuals genotyped using Illumina microarrays [5]. It also lists some other CNV events that we have found not to be as significant as those that would be discussed in detail in subsequent parts of the result description. These significant regions were selected based on minimal length threshold of at least 50kbp. Most of the insignificant events that were detected should be confirmed by repeating the hotspot 1.0 microarray analysis and further validated with the hotspot 3.1 Agilent arrays.

| <i>Chr</i> | <i>Approximate Size</i> | <i>Chr. Band</i> | <i>CNV</i> | <i>Sample</i> | <i>Cohort</i> | <i>Race</i> | <i>Inheritance</i> | <i>Control count</i> | <i>Genes</i> |
|------------|-------------------------|------------------|-------------|---------------|----------------------|------------------|--------------------|----------------------|--|
| 1 | 1000000 | 1q21.1 | Deletion | AU_W_01 | Autism | Caucasian | Both | 0/8329 | <i>CHD1L/ALC1</i> |
| 4 | 5000000 | 4q32.3 | Deletion | AUMA_14 | Autism | Hispanic | Paternal | 0/8329 | <i>SPOCK-3, KLHL-2</i> |
| 4 | 3000000 | 4q13.1 | Duplication | AU_AS_13 | Autism | Asian | de novo | 0/8329 | <i>LPHN3</i> |
| 6 | 3000000 | 6q22.31 | Duplication | AU_W_67 | Autism | Caucasian | de novo | 0/8329 | <i>GJA1</i> |
| 6 | 1000000 | 6q23.2 | Deletion | TD_W_33 | Typically developing | Caucasian | de novo | 0/8329 | <i>MED23</i> |
| 6 | 1000000 | 6q25 | Duplication | TD_W_87 | Typically developing | Caucasian | de novo | 2/8329 | <i>PARK2</i> |
| 7 | 3000000 | 7q11.22 | Deletion | AU_W_118 | Autism | Caucasian | de novo | 0/8329 | <i>AUTS2</i> |
| 7 | 1500000 | 7q11.23 | Deletion | TD_W_104 | Typically developing | Caucasian | de novo | 0/8329 | <i>WBSCR17 and 19, CALNI, ELN, STX1A</i> |
| 15 | 5000000 | 15q11.2 | Duplication | AU_MIX_10 | Autism | Mixed Race | Maternal | 0/8329 | <i>UBE3A, NIPA2, CYFIP1, FMR1</i> |
| 15 | 5000000 | 15q11.2 | Duplication | AU_W_41 | Autism | Caucasian | Maternal | 0/8329 | <i>UBE3A, NIPA2, CYFIP1, FMR1</i> |
| 15 | 5000000 | 15q11.2 | Duplication | AU_MIX_29 | Autism | Mixed Race | Maternal | 0/8329 | <i>UBE3A, NIPA2, CYFIP1, FMR2</i> |
| 15 | 500000 | 15q13.3 | Duplication | AU_MIX_28 | Autism | Mixed Race | Either | 51/8329 | <i>CHRNA7</i> |
| 15 | 500000 | 15q13.3 | Duplication | AUMA_60 | Autism | Hispanic | Either | 51/8329 | <i>CHRNA7</i> |
| 15 | 400000 | 15q13.3 | Duplication | TDMA_71 | Typically developing | Hispanic | Either | 51/8329 | <i>CHRNA7</i> |
| 15 | 400000 | 15q13.3 | Duplication | TDMA_55 | Typically developing | Hispanic | Either | 51/8329 | <i>KLF13</i> |
| 16 | 400000 | 16p11.2 | Duplication | AU_MIX_04 | Autism | Mixed Race | Either | 2/8329 | <i>VIPR2,</i> |
| 16 | 200000 | 16p11.2 | Duplication | TD_W_29 | Typically developing | Caucasian | Either | 2/8329 | <i>SEZ6L2, SH2B1</i> |
| 16 | 200000 | 16p11.2 | Deletion | TD_AA_05 | Typically developing | African American | Either | 2/8329 | <i>SEZ6L2, SH2B2</i> |
| 17 | 50000 | 17p11.2 | Deletion | AU_W_19 | Autism | Caucasian | de novo | 0/8329 | <i>SPECC1</i> |
| 17 | 50000 | 17p11.2 | Deletion | AU_W_26 | Autism | Caucasian | de novo | 0/8329 | <i>SPECC1, SLC47A2</i> |
| 17 | 400000 | 17q12 | Duplication | AU_MIX_04 | Autism | Mixed Race | Paternal | 0/8329 | <i>ACACA</i> |
| 22 | 200000 | 22q11.21 | Duplication | AUMA_69 | Autism | Hispanic | de novo | 8/8329 | <i>P14KA</i> |
| 22 | 2000000 | 22q11.22 | Duplication | TD_W_23 | Typically developing | Caucasian | de novo | 36/8329 | <i>P14KA</i> |
| 22 | 500000 | 22q11.22 | Duplication | AU_AS_13 | Autism | Asian | de novo | 36/8329 | <i>P14KA</i> |

Table 3: Summary of events found for both AU and TD samples with their corresponding locations, type of CNV event, ethnicity, inheritance and affected genes that may be related to autism . The statistics on the control cases in this table was calculated based on the data obtained from the UCSC genome browser.

The Mullen Early Scales of Learning and Vineland Adaptive Behavior test was used to evaluate the children in our study (Table 4). The Vineland Adaptive Behavior scale provided valuable information about the adaptive functioning throughout the individuals' life span that was associated with intellectual functioning [88]. The Mullen Scales of Early Learning reflected the cognitive development of infants, toddlers and children [88]. The results of the test summarized in this table clearly demonstrate that that most autistic individuals have relatively lower Mullen and Vineland scores. However, in certain unusual circumstances, typically developing individuals may also have a relatively low Mullen and Vineland scores. These low scores were often accompanied with a presence of a large CNV event in their genome. These typically developing individuals may be affected by other forms of neurological disorders and developmental or intellectual delay problems.

| Autistic (AU) | Mullen | Vineland |
|----------------------------------|---------------|-----------------|
| AU_AS_13 | 49 | 48 |
| AU_MIX_04 | 49 | 51 |
| AU_MIX_10 | 49 | 50 |
| AU_MIX_28 | 49 | 59 |
| AU_MIX_29 | 49 | 50 |
| AU_W_01 | 111 | 77 |
| AUMA_69 | 49 | 58 |
| AU_W_19 | 52 | 59 |
| AU_W_26 | 49 | 56 |
| AU_W_41 | 49 | 57 |
| AU_W_67 | 114 | 87 |
| AUMA_14 | 49 | 55 |
| AUMA_60 | 49 | 46 |
| AU_W_118 | 53 | 71 |
| Typically developing (TD) | Mullen | Vineland |
| TD_AA_05 | 80 | 88 |
| TD_W_23 | 70 | 72 |
| TD_W_29 | 81 | 90 |
| TD_W_33 | 92 | 105 |
| TD_W_87 | 127 | 109 |
| TD_W_104 | 77 | 71 |
| TDMA_55 | 82 | 111 |
| TDMA_71 | 122 | 102 |

Table 4: Summary of Mullen Early Scales of Learning and Vineland Adaptive Behavior scores of the autistic (AU) and typically developing (TD) individuals listed earlier in Table 3. Most AU individuals have relatively low Mullen and Vineland scores (below 80) although there are a few exceptions of scores above 100. TD individuals have higher Mullen scores (above the average of 80). Nevertheless, three TD individuals have low Mullen and Vineland scores and this is in conjunction with our findings regarding large copy number variant changes in their genomic hotspot regions.

For further detailed description of the detected CNV events, we separated our data into three main categories: the first category contained novel events found in autistic individuals; the second category included events found in typically developing individuals and the third category consisted of events found in both autistic and typically developing individuals. In total, there were six significant cases of CNVs in autistic individuals detected in our analysis, three large CNVs found in the typically developing group and two CNVs that appeared significant and found in both autistic and typically developing samples. From all these eleven significant events, most occurred de novo for the exception of one categorized as paternally inherited (4q32.3), and another one (15q1.2) of maternal inheritance. Two events (16p11.2 and 15q13.3) were found in both autistic and typically developing samples and could be either inherited from parents or de novo. However, the cases listed in Table 3 accounted to a total of 24, which was more than the eleven significant events chosen to be discussed. This was due to the fact that some cases listed in different individuals were found in the same chromosomal regions such as chromosome bands of 15q11.2, 15q13.3, 16p11.2, 17p11.2 and 22q11.21-22. Moreover, deletions in AU_W_19 and AU_W_26 that spanned 50kbp were not discussed here due to a smaller CNV size; repeated analysis using the hotspot 1.0 array (Nimblegen Inc.) and validation using hotspot 3.1 array (Agilent) would be necessary to confirm this results.

Novel, rare CNV events found in autistic individuals

From the Hidden Markov Model computational analysis and manual curation, we discovered a deletion of approximately 5Mb that covered a region of the chromosome 4 (4q32.3). This region has not been previously described but could be highly associated with autism. The deletion at 4q32.3 that we detected included the *KLHL-2* actin cytoskeleton gene expressed in brain and *SPOCK-3* gene that may be involved in neurogenesis [73] (Figure 4). We also conducted a further analysis of blood DNA samples for this autistic individual that revealed that the 4q32.3 deletion was paternally inherited (Figure 4). This result had been confirmed with the hotspot 3.1 array. Comparison with Mullen score of 49 and Vineland score of 55 were consistent in showing that the individual had slow learning and social skills.

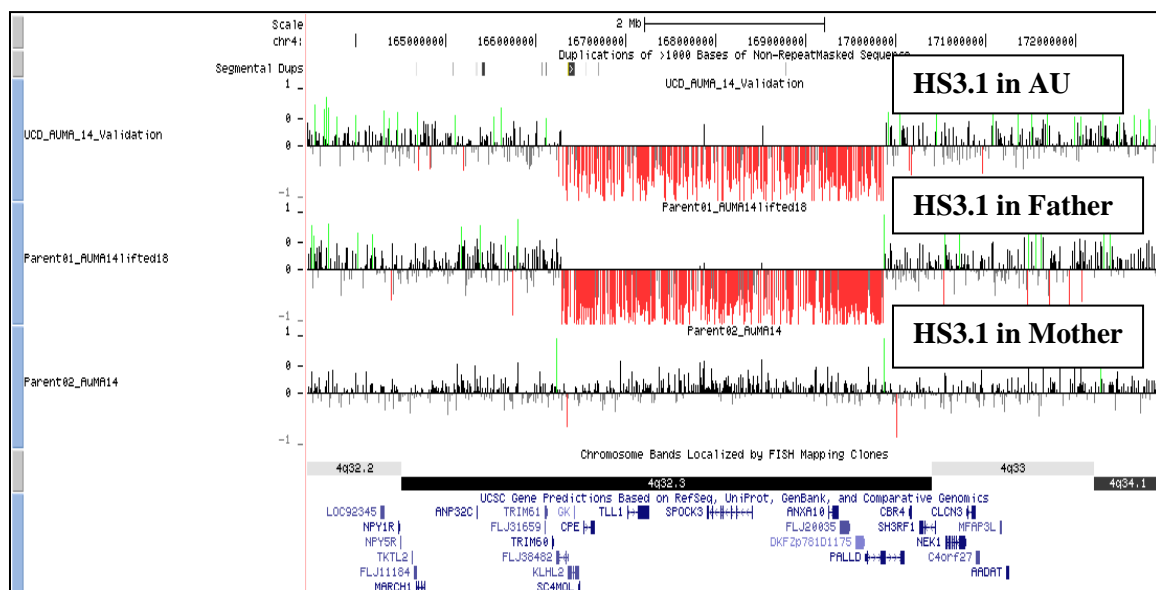


Figure 3: A large deletion of 5Mbp at 4q32.3 in AUMA_14 (Autistic individual, Hispanic) encompassing *SPOCK-3* gene and *KLHL2*-actin cytoskeleton in brain cells. This CNV was paternally inherited. The samples were validated with hotspot 3.1 Agilent arrays. The map region of the genome was retrieved from the UCSC genome browser.

Another novel 2Mbp duplication not previously linked to autism was found in chromosome 4(4q13.1) of an AU individual. As observed in Figure 4, the duplication that event occurred adjacent to the *LPHN3* and *EPHA5* gene. The *LPHN3* gene has been previously linked to the attention deficit hyperactivity disorder (ADHD) [74], but was never associated with autism. On the other hand, the *EPHA5* gene had been previously reported to be associated with autism [51]. This autistic individual had relatively low Mullen and Vineland scores of 49 and 48.

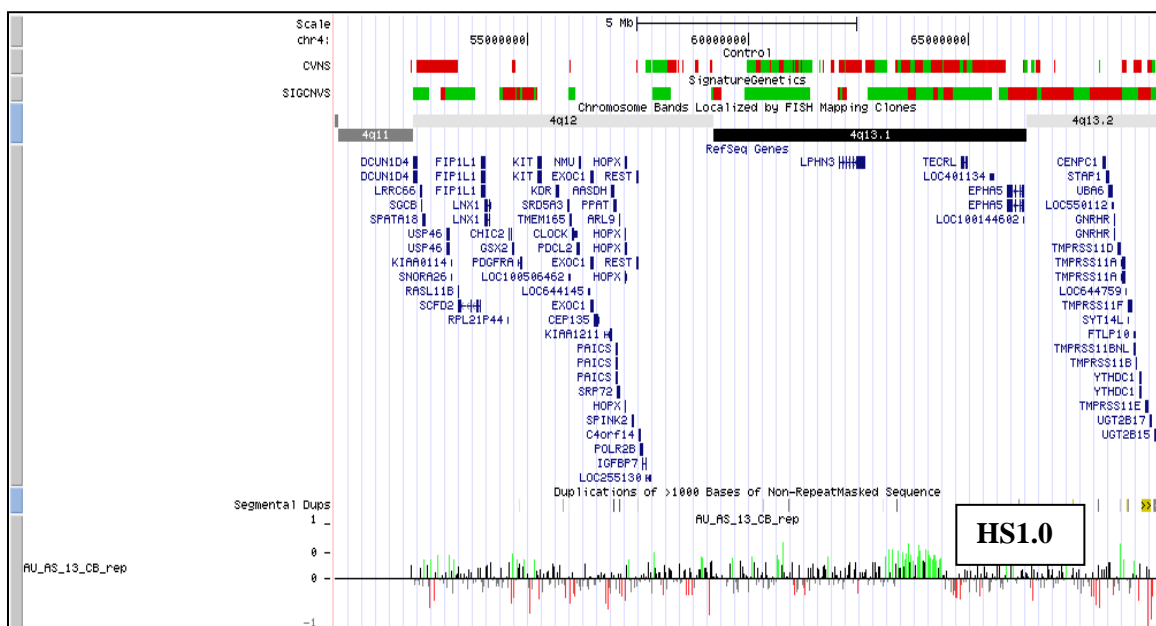


Figure 4: 2Mbp duplication at 14q13.1 that spanned the *LPHN3* gene in AU_AS_13 (Autistic individual, Asian). The figure was obtained directly from the UCSC genome browser.

We also found a rare duplication event that was not observed in control groups in an autistic individual at chromosome region of 6q22.31 that contained the *GJA1* gene (Figure 5). Despite the significant duplication observed, this autistic individual had relatively high Mullen score of 114 as well as Vineland score of 87. Sometimes, minority of autistic individuals may obtain high intelligence scores but perform poorly in interactive skills. The observation results from possible occurrence of varying phenotypic differences among autistic individuals in cognitive and social performance due to diverse genes disruptions [66].

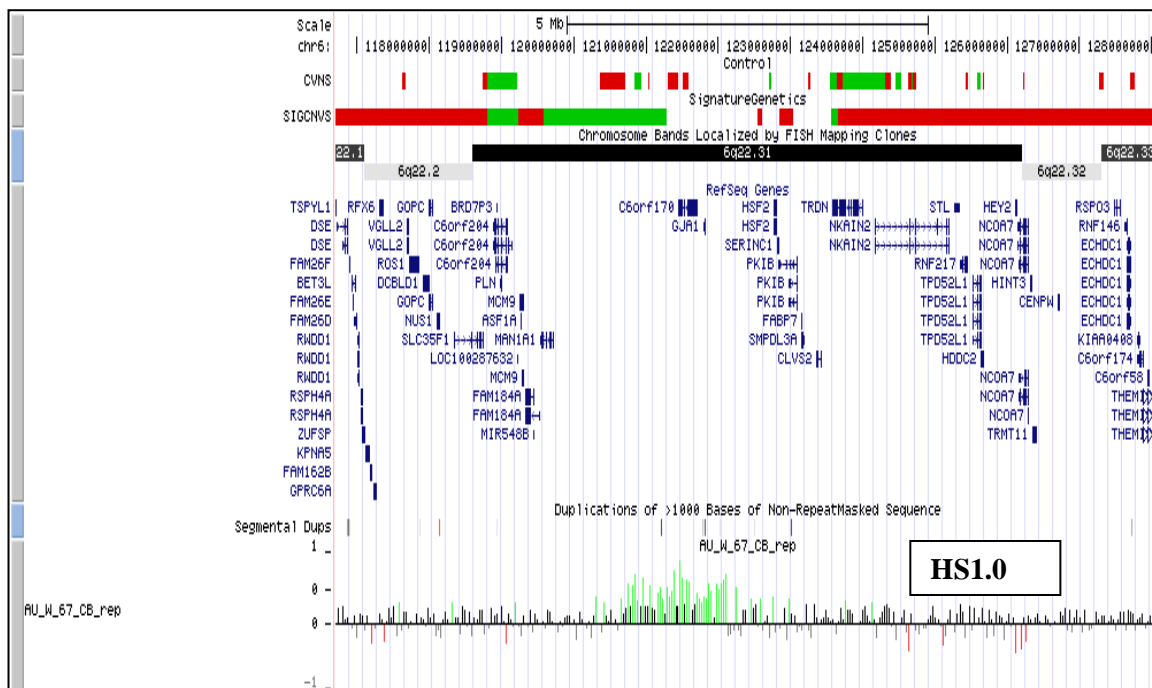


Figure 5: A 2Mbp novel duplication at 6q22.31 diagnosed in blood sample of AU_W_67 (autistic individual, Caucasian) that includes the *GJA1* gene. The map of the area was obtained from the UCSC genome browser.

In another autistic group sample, we observed a novel large deletion of about 1Mbp within chromosome 1 (1q21.1) that has not been previously described to have an association with autism (Figure 6). Although this region did not encompass genes directly linked to autism, the duplication at this location had not been observed in any of the 8329 controls. Similarly to AU_W_67 discussed in Figure 5, this individual had a high Mullen score of 111, but a low Vineland score of 77.

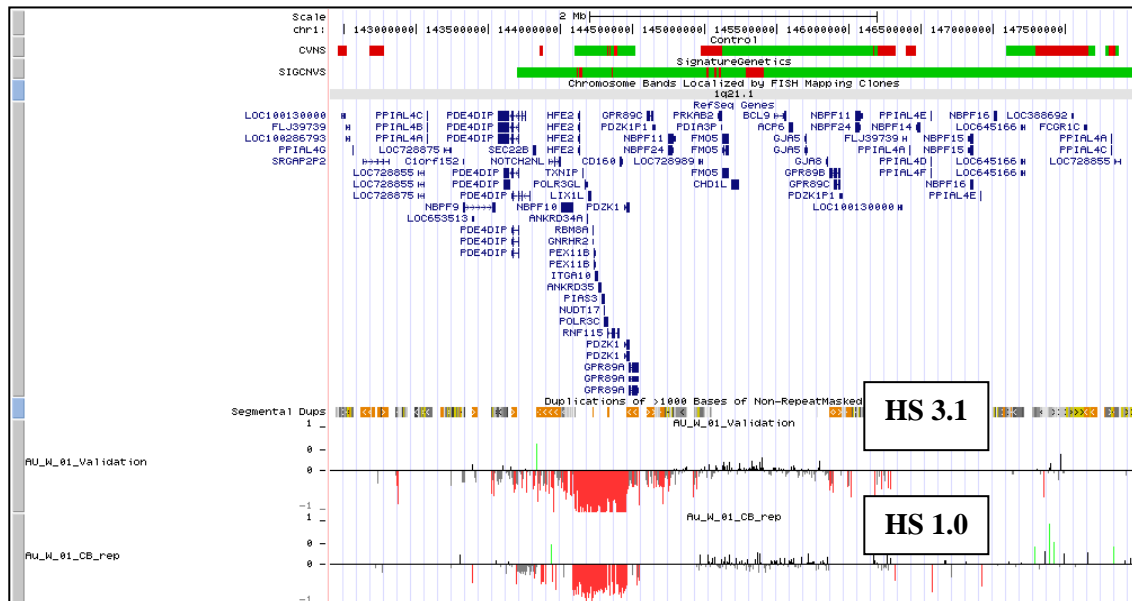


Figure 6: Deletion of 1Mbp at position 1q21.1 found in AU_W_01 (autistic individual, Caucasian). This region had not been discovered with any association with autism but was well known to be responsible for TAR syndrome as previously explained by Uhrig et al. [20]. Results from the validation with the hotspot 3.1, labeled HS3.1, were consistent with results from hotspot 1.0 Nimblegen array, labeled HS1.0. The detailed map of the area was generated with the UCSC genome portal.

We found several new deletion events that were associated with the area of chromosome 7q11.22 that encompassed the autism candidate gene 2 (*AUTS2*), shown in Figure 7 below. So far, no events of autism disorder had been linked to the region of 7q11.22 although our analysis using the UCSC genome browser showed loss of *AUTS2* gene. Deletion at this region had been reported responsible for Williams-Beuren syndrome previously [29]. Autism characteristics were reliably confirmed by low Mullen and Vineland scores of 53 and 71, respectively.

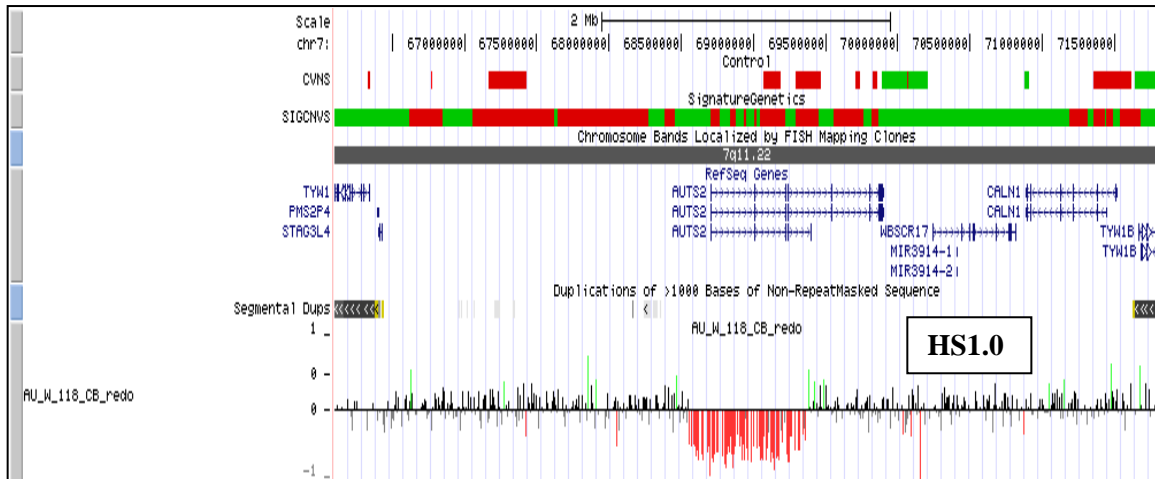


Figure 7: A 2Mb deletion of a region of 7q11.22 that contained the *AUTS2* gene was detected in AU_W_118 (autistic individual, Caucasian). The figure was secured from the UCSC genome browser.

A duplication event in a region of chromosome 15 (15q11.2) was identified in an autistic individual (shown in Figure 8). A similar large CNV event was previously described to be a maternally inherited duplication that led to autism with probability of 1-1.5% of the cases [28]. This individual had a low Mullen and Vineland scores of 49 and 50.

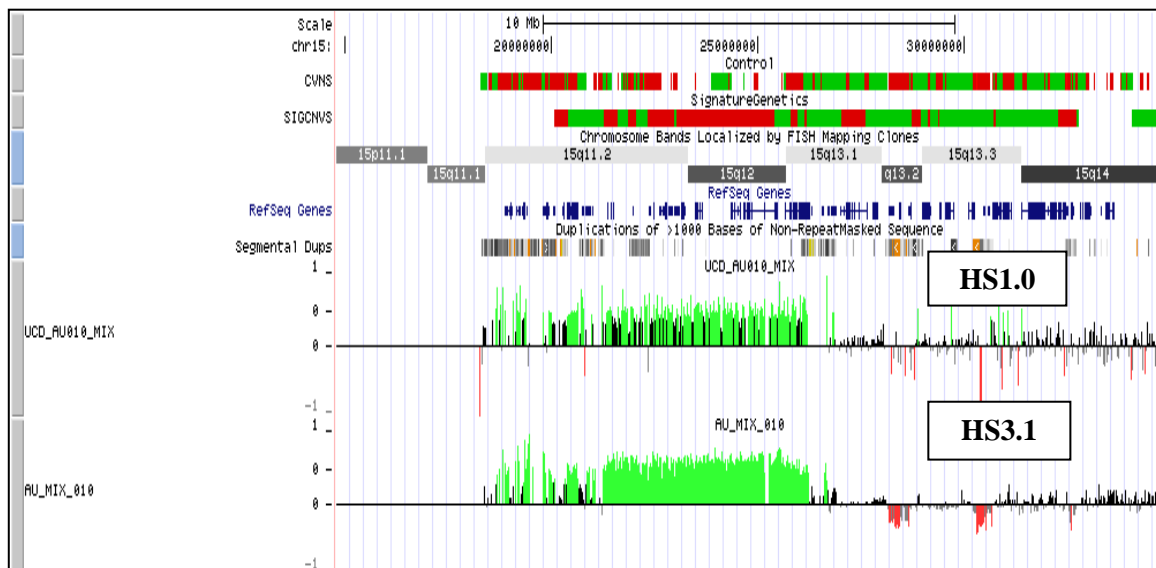


Figure 8: Large duplication of 5Mbp at chromosome 15 (15q11.2) found in an AU_MIX_10 (Autistic individual of mixed race). The first part of the figure illustrated large CNV event using Nimblegen hotspot v1.0 array. Second part of figure showed validation of sample using the Agilent hotspot v3.1 array. Both sections of the figure were retrieved from the UCSC genome website.

Novel CNV events in the typically developing group

It was expected that some large CNVs would be found in typically developing individuals because they were not thoroughly diagnosed and selected based on the absence of other neuropsychiatric disorders. An atypical deletion was found in the region of chromosome 6 (6q23.2) of a typically developing individual (Figure 9). A CNV in this region has not been found in other samples of the control group and would definitely require further validation using higher density arrays. Compared to autistic individuals, it was obvious that this typically developing individual had a higher Mullen score of 92 and Vineland score of 105. To exclude possibility of a false positive result, it would be necessary to repeat the microarray analysis of this sample.

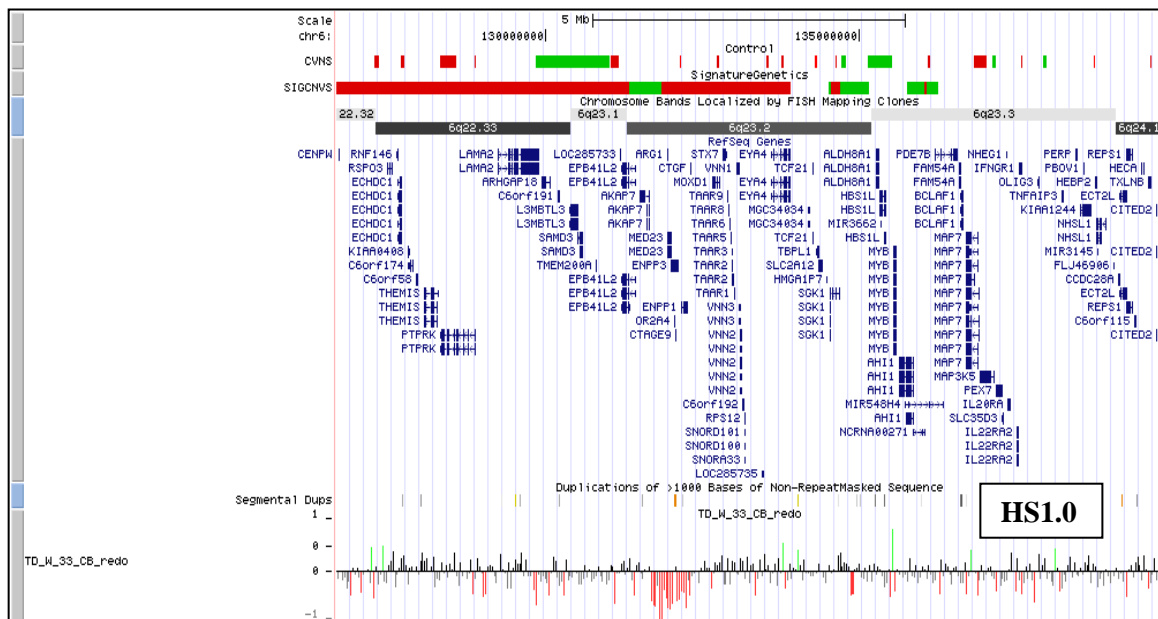


Figure 9: A rare, atypical deletion of 1Mbp at 6q23.2 in TD_W_33 (typically developing individual, Caucasian). Detailed map region of the genome above was attained from the UCSC genome browser.

An interesting 1 Mbp deletion event was discovered at the chromosome 7 (7q11.23) of a typically developing individual (Figure 10). This individual had Mullen scores of 77 and Vineland scores of 71, which were below the average normal intellectual development scores. This region has not been previously linked to autism spectrum disorder.

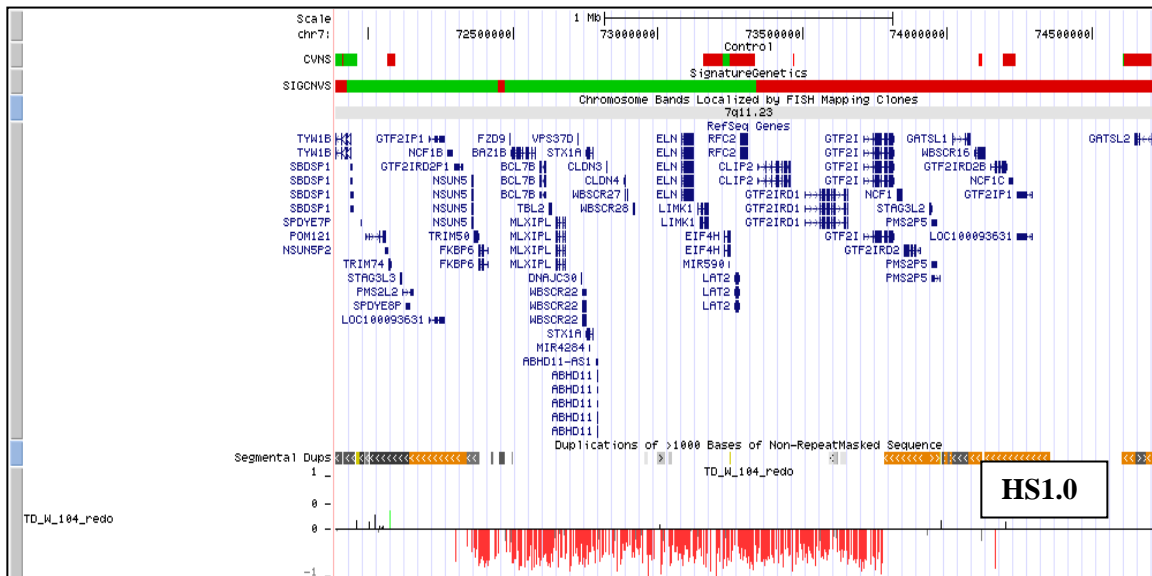


Figure 10: A large deletion of about 1Mbp at 7q11.23 in TD_W_104 (Typically developing, Caucasian). This region has never been linked to autism. Figure containing region of interest was generated using the UCSC genome browser.

Another large duplication event in an individual from the typically developing group was observed in a chromosome 22 region 22q11.22. The size of the duplication was about 1.5Mbp (Figure 11). The Mullen and Vineland scores of this individual were also below average for the typically developing group, 70 and 72. The atypical results of these individuals could be considered as an indication to contact them as well as their parents for further clinical and genetic analysis.

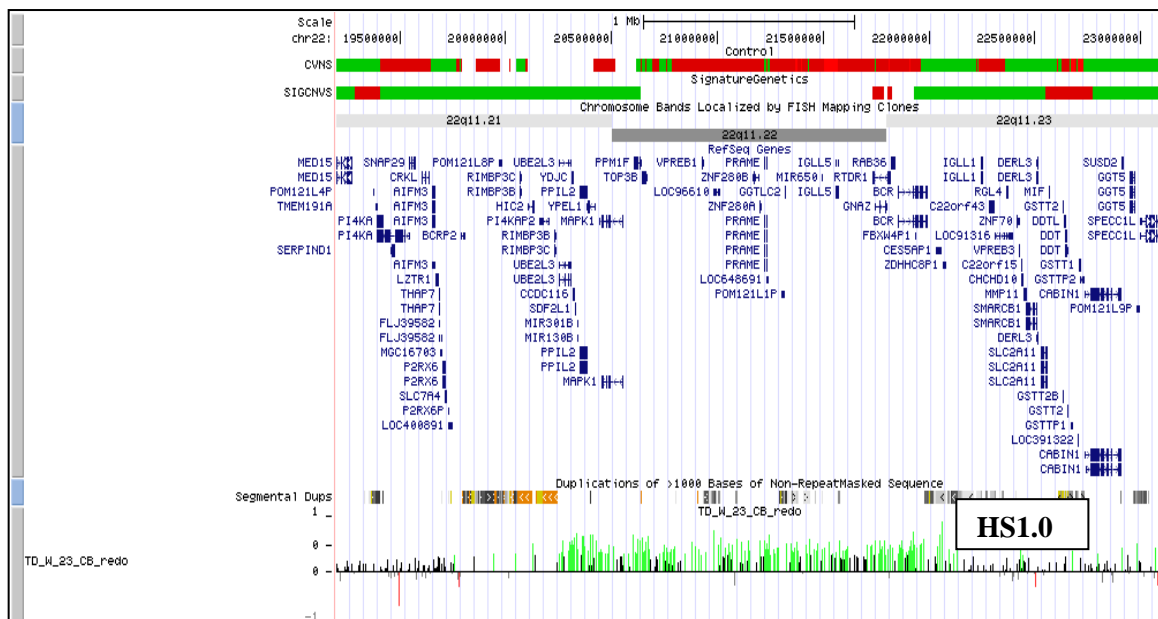


Figure 11: A Large duplication event of approximate 1.5Mbp in the region 22q11.22 of TD_W_23 (typically developing individual, Caucasian). Mapped region with significant event was acquired from the UCSC genome browser.

Significant CNV events found in both autistic and typically developing cohort

In this study we detected some significant and large CNV events that occurred in both autistic and typically developing samples. One autistic individual blood sample analysis revealed a large area of duplication that encompassed a 500 Kb region of chromosome 16 (16p11.2). As shown in Figure 12, this duplication event contained a candidate gene for seizure and autism spectrum disorder, *SEZ6L2* [75]. This autistic individual of mixed race ethnicity had a low Mullen score of 49 and Vineland score of 51. Analysis of the CHARGE cohort led to the discovery of CNVs involving the very same region in some typically developing (TD) samples. These CNVs could potentially result in a loss or gain of genes responsible for neurological disorders as well.

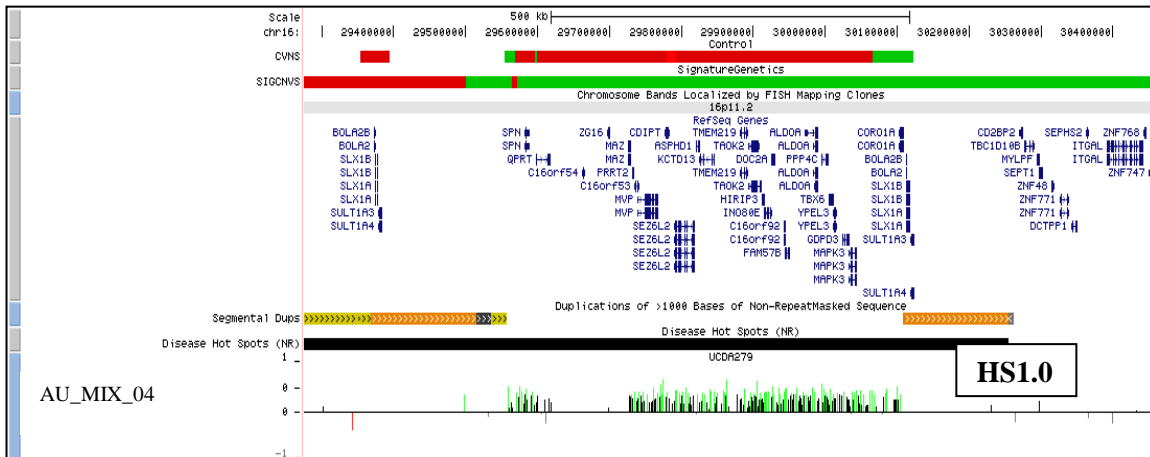


Figure 12: A 500kbp duplication in AU_MIX_04 (autistic individual, mixed race) that encompassed the *SEZ6L2* gene at region 16p11.2. The region discussed was captured from the UCSC genome browser.

Another interesting finding from our analysis was two typically developing individuals with both duplication and deletion events in this region (Figure 13). It would be interesting to address the fact that these two TD individuals had relatively low Mullen scores of 80 (TD_AA_05) and 81 (TD_W_29). Vineland scores of the individuals were in the normal range for the typically developing children, 88 for TD_AA_05 and 90 for TD_W_29.

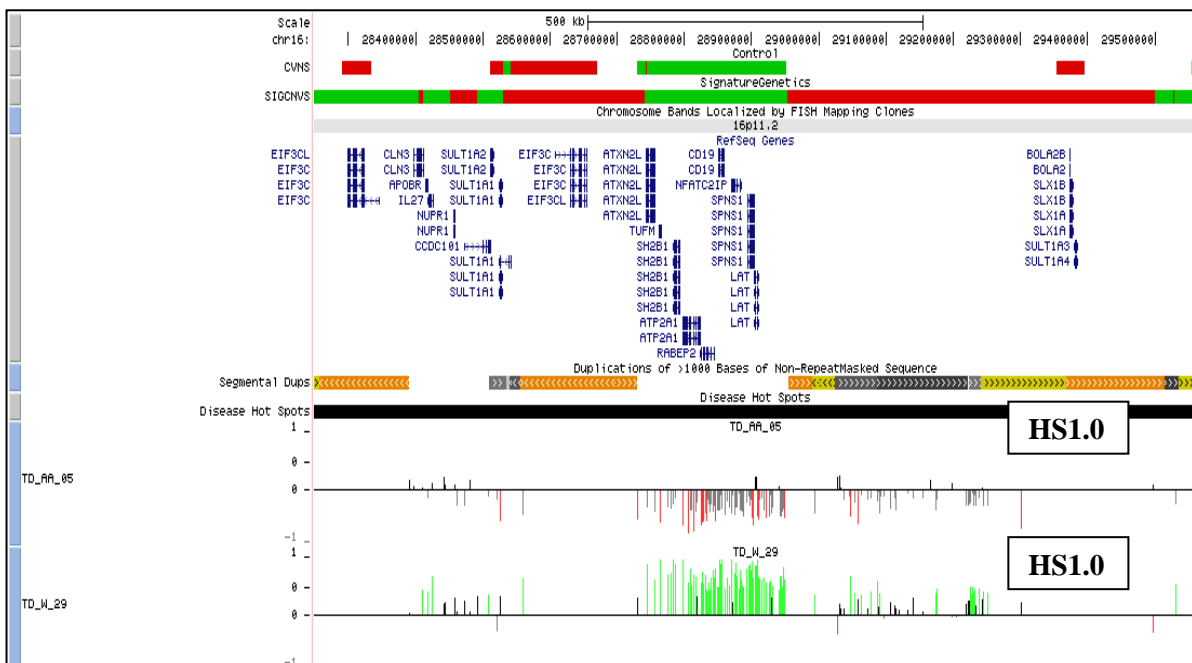


Figure 13: Deletion and duplication events found in two separate typically developing individuals, TD_AA_05 and TD_W_29 (Typically developing, African American and Caucasian). Both were assessed clinically to have low Mullen. However, Vineland scores of both individuals fell within the normal average range. Deletion was previously described to cause obesity and intellectual disability [15, 21]. Duplication had also been associated with mild symptoms of autism spectrum disorder [21,

75], attention deficit hyperactivity disorder (ADHD), bipolar disorder [15], schizophrenia [22, 75] as well as microcephaly [15]. The mapped regions were both attained from the UCSC genome browser.

Duplication and deletion events were found in some other region of chromosome 15 (15q13.3) of two autistic individuals. Both duplication and deletion events spanned a 500 kb region, and contained the *CHRNA7* gene, which had been previously shown to be affected in epilepsy, mental retardation, schizophrenia [27], and autism [26] (Figure 14). Mullen and Vineland scores of both autistic individuals confirmed the autistic phenotypes. AU_MIX_28 had a low Mullen score of 49 and Vineland of 59. AUMA_60 also showed a low Mullen score of 49 and Vineland of 46. At the same time, duplication events at different locations within the same region (15q13.3) were also detected in two typically developing individuals. One of the duplication events fell in the region of 15q13.3 containing *KLF13* gene while another covered the region containing *CHRNA7* gene (Figure 15). Despite the fact that these typically developing individuals had duplications in the area close to the duplication observed in autistic individuals, both of them had relatively high and normal range of Vineland scores, but still fairly low Mullen scores. Both individuals were of Hispanic ethnicity. Individual labeled TDMA_55 had quite low Mullen score of 82 but high Vineland score of 111. Another individual, TDMA_71 had high Mullen score of 122 and Vineland score of 102.

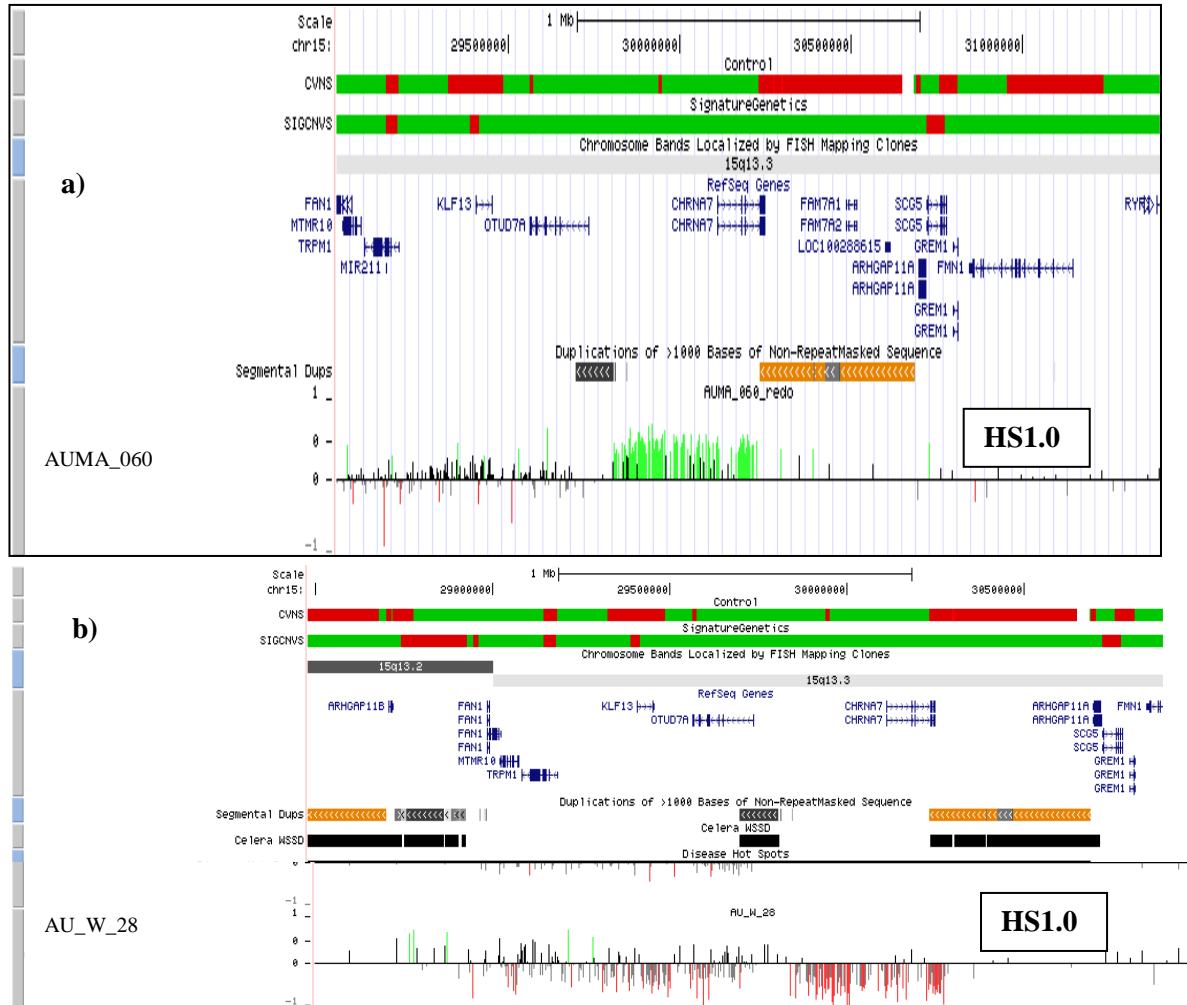


Figure 14: **a)** A 500kbp duplication in AUMA_060 (autistic individual, Hispanic). **b)** At the very same position, a 500kbp deletion is observed in AU_W_28 (autistic individual, Caucasian). CNV events in both individuals caused a disruption of the *CHRNA7* gene. Captured figure was obtained from the UCSC genome browser.

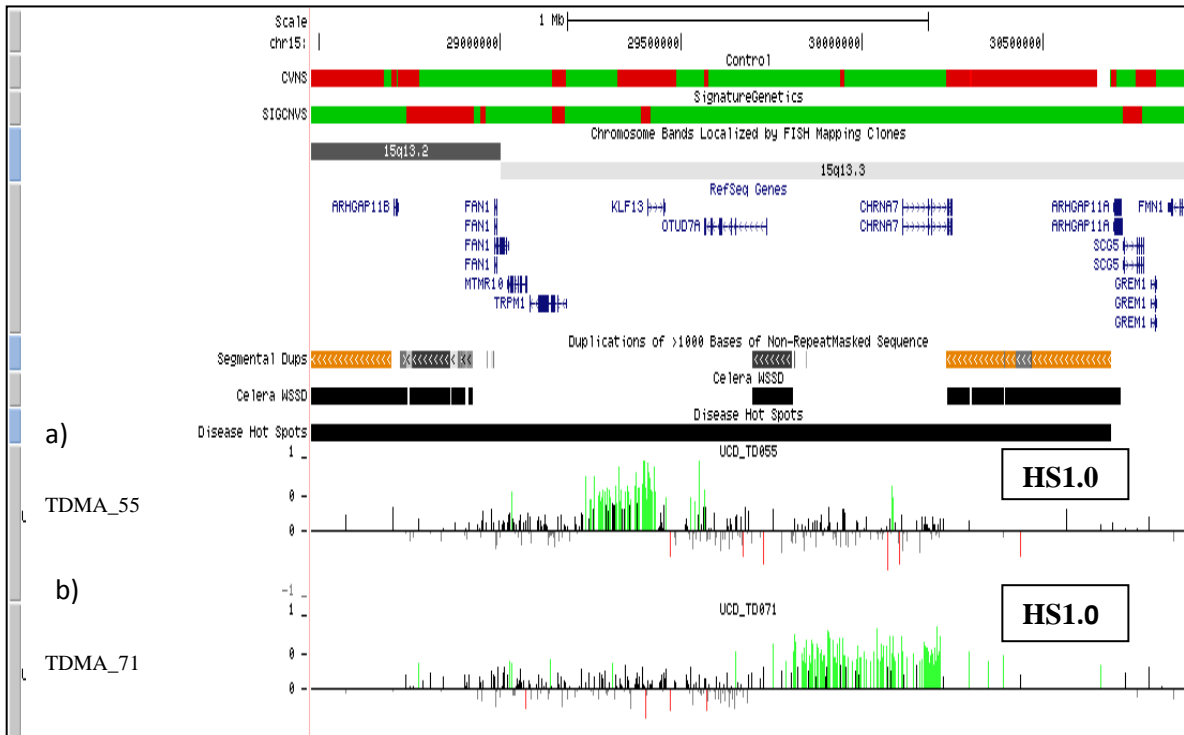


Figure 15: a) An atypical duplication that encompassed *KLF13* gene has never been observed in any controls. **b)** Duplication in another typically developing individual (TDMA_71) of Mexican heritage looked very similar to the duplication observed at region 15q13.3 of autistic sample, AUMA_060. The mapped genomic regions were generated from UCSC genome portal.

Chapter 4. DISCUSSION

Our search for *de novo* CNVs associated with autism had led to the discovery of many areas affected by significant duplications or deletions in autistic population. Some of the CNVs we found were previously described, and some of them were previously shown to be implicated in other neurodevelopmental and genetically inherited disorders such as schizophrenia, attention deficit hyperactivity disorder (ADHD), epilepsy and thrombocytopenia with absent radius (TAR) syndrome. The design of the oligonucleotide microarray used in this study was customized with an enriched probe density in the selective 107 genomic hotspot regions that could potentially cause a detection bias [5]. The genomic hotspot regions were covered with probes with approximate 25-fold enrichment compared to other commercial arrays [6]. Most of the rare pathogenic CNVs that were discovered in this study covered regions of more than 100kb, which was consistent with results from previous studies [3, 4, 5, 6]. Events of larger than 1Mb were previously found to be relatively less common in the typically developing population than in autistic individuals [6, 27, 36]. CNVs detected in autistic individuals were located in the regions that were more gene-enriched [5]. For the significant events that we found in our analysis, we discuss some previously described events that occurred within the same or surrounding region.

A novel deletion in chromosome region 4q32.3 detected in an autistic individual

For the CNV that we detected in the 4q32.3 region, it was previously shown that it can cause occipital encephalocele as well as delayed neuropsychomotor movements and learning disabilities in the patient [44]. However, it was never directly associated with autism. The terminal deletion at the neighboring region, 4q32.2, has been previously described (Figure 3). It occurred *de novo* as karyotyping and microarray analysis proved that both parents had normal chromosomes [44]. Other sources had also linked a 4q deletion to brain

abnormalities such as absence of corpus callosum [45], cerebral atromosome [46] and severe learning difficulties [47]. Individuals with a 4q32-33 deletion also exhibited other associated phenotypes such as facial and digital anomalies, congenital heart defects and postnatal growth deficiency [44, 48]. In addition to that, a deletion in another proximal region, 4q34.2 has been previously found to cause autism and malformations of limbs. Similarly to deletion event at 4q32.3, the loss of structural variant at 4q34.2 was also of paternal inheritance [19]. This *de novo* CNV at 4q32.3 that was found in our current analysis has not yet been directly associated to autism.

An atypical duplication at region of 4q13.1 in autistic individual of Asian ethnicity

An unstable region of 4q13.1 that we identified has not been previously linked to autism. We suggest that *EphA5* gene located in this region may be implicated in autism. Ephrin and Eph receptor were shown to be responsible for early synaptogenesis at the hippocampus in mice [51].

Unique region of duplication at 6p22.31 identified in autistic individual

Our analysis of the CHARGE cohort revealed a novel duplication in an autistic child at the region 6p22.31, which included the *GJA1* gene (Figure 4). According to previous neurological studies, a missense mutation in *GJA1* contributed to a rare autosomal dominant inherited disease oculodentodigital dysplasia (ODDD). Patients with ODDD suffered from facial, teeth, limbs and heart defects as well as sensory deficits [55]. Since *GJA1* gene codes for the connexin protein, its mutations can potentially influence communications between cells, which can affect the sensory or physiological responses of individuals [55]. It would be of interest to examine the autistic individual with the *GJA1*-containing duplication for the presence of other neurological disorders besides autism. Further genetic and clinical analysis

could also be conducted to explore a possibility of overlapping candidate genes and specific CNVs that could cause both oculodentodigital dysplasia and autism.

1q21.1 deletion detected in autistic individual

1q21.1 deletions similar to the one we detected have been previously described to be responsible for 10% of autism spectrum disorders and attention deficit hyperactivity disorder cases [76, 78] and for 0.2%-0.6% of patients with schizophrenia [27]. Deletion at 1q21.1 was also associated with thrombocytopenia with absent radius (TAR) syndrome) due to nonallelic recombination as the region was found to be flanked by four copies of segmental duplications in direct orientation and of high degree of sequence identity [20, 76]. Chromosomal microarray analysis and cytogenetic testing have also linked a deletion at 1q21.1 to developmental delays, mild-to-moderate intellectual disability, mild dysmorphic facial features and microcephaly [76]. The microdeletion of 1q21.1 could be inherited from parents who were mildly affected but not showing obvious phenotypical differences or from parents who appeared completely normal [76, 77, 78]. Such findings suggested that the 1q21.1 deletion had a reduced penetrance and a variable expressivity.

Event of deletions at region 7q11.22-23 in autistic and typically developing individuals

A large deletion covering a region of 7q11.22-23 has been implicated in Williams's syndrome that is associated with a mild developmental delay (DD) [29]. In this region, we detected deletions in an autistic, and in a typically developing patient (Figure 7 and 10). Previous study showed that this area contained *AUTS-2* gene that consisted of 19 exons. They detected a *de novo* reciprocal translocation that could potentially lead to the *AUTS2* gene being truncated between exons 2 and 7 in region 7q11.22 [52] in six patients with signs of mental retardation and autism spectrum disorder. Similarly to cases reported in studies of microdeletion in 1q21.1, a mutation that was associated with CNVs at 7q11.22 showed

variable expressivity as severity of symptoms varied between patients. The authors hypothesized that clinical variability could be caused by differences in breakpoint locations within the *AUTS2* gene [52].

Significant 15q11.2 duplication found in three autistic individuals

Previous study of the region 15q11.2 indicated that copy number variants at this region, especially duplications, were maternally inherited, and led to autism in 1-1.5% of the cases [28]. We found several children from both autistic and control groups that had duplications in this area. The maternal heterodisomy of microduplication and microdeletion at the region of 15q11-13 had been previously associated with Prader-Willi and Angelman syndrome [37, 38]. However, another study found an individual with genetic instability in this region that exhibited some characteristics relevant to autism and met the criteria of the DSM-IV test [28]. This region of chromosome 15 encompassed candidate genes of E3 ubiquitin ligase gene (*UBE3A*) and a cluster of gamma-aminobutyric acid (*GABA(A)*) receptor subunit genes [39] that are essential for neurotransmission in central and peripheral nervous systems. A duplication that resulted in several copies of maternally expressed *UBE3A* genes led to epigenetic changes in the postnatal neurons [41]. The CNV events present in this region can cause an imbalance of the neurotransmission activities in the brain that could lead to the autistic behavior in these children [40].

Novel deletion at region of 6q23.2 as detected in a typically developing subject

Our analysis also uncovered some interesting events in typically developing individuals. The deletion in region 6q23.2 of our sample interfered with the *MED23* gene. Hashimoto et al. associated the substitution of amino acid 617 in *MED23* protein with moderate intellectual disability. Furthermore, the authors discovered that a mutation in transcriptional machinery that encoded the *MED23* subunit could potentially alter the

immediate early genes regulation, thus interfering with learning and memory formation [79, 80].

Atypical deletion of 7q11.23 encountered in typically developing subject

A large deletion at 7q11.23 that we observed in a typically developing individual encompassed the *STX1A* gene that regulated the serotonin transporter (5-HTT) for controlling glutamate levels in neurotransmission [53]. Nakamura et al had conducted an autistic cohort study in Japan that determined the association between low levels of *STX1A* and higher risk of autism. On the other hand, *GTF2i* gene located in this region was shown to play an important role in controlling social behavior and verbal communication skills [54]. A simultaneous deletion of three genes (*STX1A*, *CYLN2* and *GTF2i*) in this region was strongly linked to Williams-Beuren Syndrome and autism spectrum disorder [35]. However, other studies had proven that patients with duplication events at 7q11.23 exhibited overt symptoms of ASD such as visual, language expression and social complications [35]. A typically developing individual with the 7q11.23 deletion detected in our study had relatively low Mullen (77) and Vineland (71) scores, which indicated that this person might have other mild intellectual disability and social adaptive skills, although not considered in the category of autism. Further genetic analysis and clinical diagnosis would be necessary to confirm the condition of this individual.

Duplication event at 22q11.22 in a typically developing individual

A large duplication of 22q11.22 that we found in a typically developing individual was novel as previous studies had identified deletions in this region associated with intellectual disability, short stature, eye anomalies, cardiac defects and schizophrenia, but no CNVs associated with autism has been yet uncovered [60, 76].

Occurrence of duplications involving 15q13.3 in autistic and typically developing individuals

Our analysis yielded unexpected results of detecting significant structural variants in both autistic and typically developing individuals. Many microduplication events were observed in the region of 15q13.3 in both the autistic and typically developing samples, most likely due to nonallelic homologous recombination (NAHR) events between the BP4 and BP5 low copy repeats [42]. Events in one autistic individual and one typically developing person duplicated the entire *CHRNA7* gene while another autistic individual was shown to have a deletion of *CHRNA7*. This gene was responsible for encoding the alpha7 subunit of neuronal nicotinic acetylcholine receptor (nAChR) that formed pentameric ligated cation channels [42]. Imbalance of the alpha7 nAChR could affect the cerebral cortex and hippocampus leading to impairments in cognitive processing and social behavior related to autism spectrum disorders [43]. Microdeletions in this area have been previously linked to epilepsy, schizophrenia and epilepsy [43].

Deletion and duplication events at 16p11.2 occurred in both autistic and typically developing individuals

We also uncovered some deletion and duplication events in the region 16p11.2 in both autistic and typically developing samples. Hence, it is unclear if this region is implicated in autism. Both deletion (in a single TD sample) and duplications (in a single TD and a single AU samples) were flanked by 99% identical low copy repeats that would make this region susceptible to non-allelic homologous recombination events. It was possible that recombination events contribute to altering dosage-sensitive genes, thus causing autism. Genetic variation in the autism-specific R386H amino acid substitution in exon 7 of seizure-related gene, *SEZ6L2* located in this region could contribute to autism in European descent [49]. Duplications at 16p11.2 were linked to different forms of psychiatric disorders such as

schizophrenia [21, 50], anxiety, depression [21] and autism [22]. Deletions at 16p11.2 were shown to result in attention deficit hyperactivity disorder (ADHD), developmental delay (DD), autism spectrum disorder (ASD) [50] as well as facial dysmorphism [21]. The genetically unstable region at 16p11.2 also encompassed the *SH2B1* gene that had been implicated in neuronal differentiation and development of sympathetic neurons causing neurodevelopmental problems as well as decreased growth hormone signaling leading to development of diabetes and obesity [32]. *SH2B1-3* genes are also known to regulate various physiological responses through signaling, gene expression and cell adhesion [32]. Thus genetic instability in this region was not only specific for autism, but was also observed in individuals that were diagnosed with other mild forms of mental, psychological and developmental abnormalities.

Chapter 5. CONCLUSION AND FUTURE DIRECTIONS

Overall, our data suggested that most of the rare pathogenic CNVs covered regions of more than 100kb, which was consistent with results from previous studies [3, 4, 5, 6]. Deletions and duplications of genomic structures also occurred more frequently in autistic group. Out of twenty-two samples discussed here, sixteen belonged to autistic individuals. This was consistent with previous studies that also pointed out that events of larger than 1Mb were relatively less common in the typically developing population than in autistic individuals [6, 27, 36]. The genomic regions affected by CNV events in autistic individuals contained some genes that were not previously associated with autism but influenced neuronal signals and physiological responses such as the *GJA1*, *UBE3A*, *MED23*, *EphA5* and *SH2B1*. We suggest that those genes should be further investigated in mice and drosophila model systems for their potential role in autism and autism spectrum disorders.

It is imperative to note that the detection of CNV in our study may be biased towards the hotspot recombination areas, since the oligonucleotide microarray used in this study had an enriched probe density in the selective 107 genomic hotspot regions [5]. The genomic hotspot regions were covered with probes with approximate 25-fold enrichment compared to other commercial arrays [6]. Other studies [81, 82] have reported that many false-negative CNV calls reside within the genomic backbone or non-hotspot regions due to insufficient probe coverage [5]. However, in our analysis we limited our consideration to CNVs that contained at least 10 probes, even in the genomic backbone area, and concentrated on the CNVs that spanned at least 50,000 bp.

The discovery of copy number variants provides an outstanding breakthrough in medical and genetic fields to question the belief that certain disorders only arise from single nucleotide alteration. It has been previously suggested that at least fifteen percent of human developmental diseases were caused by large structural variants that contributed to dosage

imbalance of multiple genes [5]. Further research is necessary to provide more detailed data from high-resolution arrays and next-generation exome sequencing of all samples detected with significant copy number variants. This would allow us to make thorough comparison of different microarray analyses data and provide support and validation to previous results. Increase in sample size that includes samples from patients with other neurological disorders should also be put into consideration for detection of low copy number repeats and genes in the affected regions that could potentially contribute to multiple disorders. In addition, compilation of clinical data available from CHARGE cohort of MINDS Institute, University California Davis should be utilized to evaluate the relative contributions of genetic and environmental factors to autism. A recent monozygotic twins study had raised concerns that environmental exposures contributed to genetic susceptibility to autism spectrum disorders [83]. In conclusion, CNV analysis is instrumental in providing preliminary information on genetic and epigenetic mechanisms contributing to genetic predisposition to diseases that can be further evaluated and extended in subsequent studies.

References

1. Marshall, C. R et al. (2008). Structural variation of chromosomes in autism spectrum disorder. *American journal of human genetics*, 82(2), 477-88. doi:10.1016/j.ajhg.2007.12.009
2. Feuk, L et al. (2006). Structural variants: changing the landscape of chromosomes and design of disease studies. *Human Molecular Genetics*, 15(1), doi:10.1093/hmg/ddl057
3. Girirajan, S et al. (2011). Human Copy Number Variation and Complex Genetic Disease. *Annual Review of Genetics*, doi:10.1146/annurev-genet-102209-163544
4. Girirajan, S., & Eichler, E. (2010). Phenotypic variability and genetic susceptibility to genomic disorders. *Human Molecular Genetics*, 19(Review Issue 2), doi:10.1093/hmg/ddq366
5. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, et al. (2011) Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. *PLoSGenet* 7(11): e1002334. doi:10.1371/journal.pgen.1002334
6. Itsara, A et al. (2009). Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease. *The American Society of Human Genetics*, 84, 148–161. doi:10.1016/j.ajhg.2008.12.014
7. Morrow, E. M et al. (2010). Genomic Copy Number Variation in Disorders of Cognitive Development. *JOURNAL OF THE AMERICAN ACADEMY OF CHILD & ADOLESCENT PSYCHIATRY*, 49(11), doi:10.1016/j.jaac.2010.08.009
8. Shinawi, M., & Cheung, S. W. (2008). The array CGH and its clinical applications. *Drug Discovery Today*, 13(17/18), doi:10.1016/j.drudis.2008.06.007
9. Williams, S. R et al. (2010). Array comparative genomic hybridisation of 52 subjects with a SmitheMagenis-like phenotype: identification of dosage sensitive loci also associated with schizophrenia, autism, and developmental delay. *J Med Genet*, 47, 223e229. doi:10.1136/jmg.2009.068072
10. Hertz-Picciotto I, Croen LA, Hansen R, Jones CR, van de Water J, Pessah IN. (2006). The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Environ Health Perspect* 114(7):1119-1125.
11. Lord C, Risi S, Lambrecht L, Cook EH, Jr., Leventhal BL, et al. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and

- communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30: 205-223.
12. Lord C, Rutter M, Le Couteur A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24: 659-685.
 13. Chen Y-Z, Matsushita M, Girirajan S, Lisowski M, Sun E, Sul Y, Bernier R, Estes A, Dawson G, Minshew N, Shellenberg GD, Eichler EE, Rieder MJ, Nickerson DA, Tsuang DW, Tsuang MT, Wijsman EM, Raskind WH, Brkanac Z. (2012). Evidence for Involvement of GNB1L in Autism. *Am J Med Genet Part B* 159B:61–71.
 14. Girirajan, S., & Eichler, E. E. (2011). De Novo CNVs in Bipolar Disorder: Recurrent Themes or New Directions. *Neuron*, (72), doi:10.1016/j.neuron.2011.12.008
 15. Rosenfeld, J. A., Coppinger, J., Bejjani, B. A., Girirajan, S., Eichler, E. E., Shaffe, L. G., & Ballif, B. C. (2009). Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications. *J Neurodevelop Disord*, 2, 26-38. doi: 10.1007/s11689-009-9037-4
 16. Malhotra, D., McCarthy, S., Michaelson, J. J., Vacic, V., Burdick, K. E., Yoon, S., ... Sebat, J. (2011). High Frequencies of De Novo CNVs in Bipolar Disorder and Schizophrenia. *Neuron*, 72, doi: 10.1016/j.neuron.2011.11.00
 17. Day, N., Hemmaplard, A., Thurman, R. E., Stamatoyannopoulos, J. A., & Noble, W. S. (2007). Unsupervised segmentation of continuous genomic data. *Bioinformatics*, 23(11), 1424-1426. doi:10.1093/bioinformatics/btm096
 18. Mefford, H.C., Cooper, G.M., Zerr, T., Smith, J.D., Baker, C., Shafer, N., Thorland, E.C., Skinner, C. & Eichler, E. E. (2009). A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res*, 19 (9). 1579-1585.
 19. Davis, L. K., Meyer, K. J., Rudd, D. S., Librant, A. L., Epping, E. A., Sheffield, V. C., & Wassink, T. H. (2009). Novel copy number variants in children with autism and additional developmental anomalies. *J Neurodevelop Disord*, 1, 292-301. doi: 10.1007/s11689-009-9013-
 20. Uhrig, S. (2007). Impact of Array Comparative Genomic Hybridization–Derived Information on Genetic Counseling Demonstrated by Prenatal Diagnosis of the TAR (Thrombocytopenia Absent–Radius) Syndrome–Associated Microdeletion 1q21. *The American Journal of Human Genetics*, 81(4), 866–868.

21. Weiss, L. A et al. (2008). Association between Microdeletion and Microduplication at 16p11.2 and Autism. *N Engl J Med*, 358(7), 667-75.
22. McCarthy, S et al. (2009). Microduplications of 16p11.2 are Associated with Schizophrenia. *Nat Genet*, 41(11), 1223-1227.
23. Fernell, E. (2011). Early intervention in 208 Swedish preschoolers with autism spectrum disorder. A prospective naturalistic study. *Research in developmental disabilities*, 32(6), 2092-101.doi:10.1016/j.ridd.2011.08.002
24. Ray-Subramanian, C. E. (2011). Brief report: adaptive behavior and cognitive skills for toddlers on the autism spectrum. *Journal of autism and developmental disorders*, 41(5), 679-84.doi:10.1007/s10803-010-1083-y
25. Akshoomoff, N. (2006). Use of the Mullen Scales of Early Learning for the assessment of young children with Autism Spectrum Disorders. *Child neuropsychology*, 12(4-5), 269-77.doi:10.1080/09297040500473714
26. Helbig, I et al. (2009). 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nature genetics*, 41(2), 160-2.doi:10.1038/ng.292
27. International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455, 237–241.
28. Kwasnicka-Crawford, D. A et al. (2007). Characterization of an autism-associated segmental maternal heterodisomy of the chromosome 15q11-13 region. *Journal of autism and developmental disorders*, 37(4), 694-702.doi:10.1007/s10803-006-0225-8
29. Blyth, M et al. (2008). A novel 2.43 Mb deletion of 7q11.22-q11.23. *American journal of medical genetics. Part A*, 146A(24), 3206-10.doi:10.1002/ajmg.a.32584
30. Ben-David, E et al. (2011). Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. *Human molecular genetics*, 20(18), 3632-41.doi:10.1093/hmg/ddr283
31. Zollino, M et al. (1995). Further contribution to the description of phenotypes associated with partial 4q duplication. *American journal of medical genetics*, 57(1), 69-73.doi:10.1002/ajmg.1320570116
32. Wang, T et al. (2011). The adaptor protein SH2B3 (Lnk) negatively regulates neurite outgrowth of PC12 cells and cortical neurons. *PloS one*, 6(10), e26433.doi:10.1371/journal.pone.0026433
33. Lavallée, G et al. (2006). The Kruppel-like transcription factor KLF13 is a novel regulator of heart development. *The EMBO journal*, 25(21), 5201-13.doi:10.1038/sj.emboj.7601379

34. Canastar, A et al. (2011). Promoter Methylation and Tissue-Specific Transcription of the $\alpha 7$ Nicotinic Receptor Gene, *CHRNA7*. *Journal of Molecular Neuroscience*.doi:10.1007/s12031-011-9663-7
35. Malenfant, P et al. (2011). Association of GTF2i in the Williams-Beuren Syndrome Critical Region with Autism Spectrum Disorders. *Journal of autism and developmental disorders*. doi:10.1007/s10803-011-1389-4
36. McCarroll, S.A., Kuruville, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A. ,et al . (2008). Integrated detection and polpulation-genetic analysis of SNPs and copy number variation. *Nat. Genet.* 40, 1166-1174.
37. Nicholls, R. D., Saitoh, S., & Horsthemke, B. (1998). Imprinting in Prader-Willi and Angelman syndromes. *Trends in Genetics*, 14, 194–200.
38. Roberts, S., Maggouta, F., Thompson, R., Price, S., & Thomas, S. (2002). A patient with a supernumerary marker chromosome (15), Angelman syndrome, and uniparental disomy resulting from paternal meiosis II non-disjunction. *Journal of Medical Genetics*, 39, E9.
39. Schroer, R. J., Phelan, M. C., Michaelis, R. C., Crawford, E. C., Skinner, S. A., Cuccaro, M., Simensen, R. J., Bishop, J., Skinner, C., Fender, D., & Stevenson, R. E. (1998). Autism and maternally derived aberrations of chromosome 15q. *American Journal of Medical Genetics*, 76, 327–336.
40. Olsen, R. W., & Tobin, A. J. (1990). Molecular biology of GABAA receptors. *Faseb Journal*, 4, 1469–1480.
41. Scoles, H. A. (2011-12-12). Increased copy number for methylated maternal 15q duplications leads to changes in gene and protein expression in human cortical samples. *Molecular autism*, 2(1), 19.doi:10.1186/2040-2392-2-19
42. Szafranski, P. (2010-07). Structures and molecular mechanisms for common 15q13.3 microduplications involving *CHRNA7*: benign or pathological?. *Human mutation*, 31(7), 840-850.doi:10.1002/humu.21284
43. Deutsch, S. I. (2011-11). Pharmacotherapeutic implications of the association between genomic instability at chromosome 15q13.3 and autism spectrum disorders. *Clinical neuropharmacology*, 34(6), 203-5.doi:10.1097/WNF.0b013e31823a1247
44. Roberto Quadrelli, Eugen M. Strehle, Alicia Vaglio, Mariela Larrandaburu, Búrix Mechoso, Andrea Quadrelli, Yao-Shan Fan, and Taosheng Huang. *Genetic Testing*. April 2007, 11(1): 4-10. doi:10.1089/gte.2006.9995.

45. Fagan K, Gill A (1989) A new interstitial deletion of 4q(q21.1;q22.1). *J Med Genet* 26:644–647.
46. Marneros AG, Olsen BR (2005) Physiological role of collagen XVIII and endostatin. *FASEB J* 19:716–728
47. Lin AE, Garver KL, Diggans G, Clemens M, Wenger SL, Steele MW, Jones MC, Israel J (1988) Interstitial and terminal deletions of the long arm of chromosome 4: further delineation of phenotypes. *Am J Med Genet* 31:533-548.
48. Strehle EM, Bantock HM (2003) The phenotype of patients with 4qsyndrome. *Genet Couns* 14:195–205.
49. Kumar, R.A., Marshall, C.R., Badner, J.A., Babatz, T.D., Mukamel, Z. , et al. (2009). Association and Mutation Analysis of 16p11.2 Autism Candidate Genes. *PLoS One*, 4(2): e4582.
50. Fernandez, B.A, et al. (2010). Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. *Journal of medical genetics*, 47(3), 195-203. doi:10.1136/jmg.2009.069369
51. Akaneya, Y, et al. (2010). Ephrin-A5 and EphA5 interaction induces synaptogenesis during early hippocampal development. *PloS one*, 5(8), e12486. doi:10.1371/journal.pone.0012486
52. Huang, X, et al. (2010). A de novo balanced translocation breakpoint truncating the autism susceptibility candidate 2 (AUTS2) gene in a patient with autism. *American journal of medical genetics. Part A*, 152A(8), 2112-4. doi:10.1002/ajmg.a.33497
53. Nakamura, K et al. (2011). Replication study of Japanese cohorts supports the role of STX1A in autism susceptibility. *Progress in neuro-psychopharmacology & biological psychiatry*, 35(2), 454-8. doi:10.1016/j.pnpbp.2010.11.033
54. Dai, L., Bellugi, U., Chen, X.N., Pulst-Korenberg, A.M., Jarvinen-Pasley, A., Tirosh-Wagner, T., et al. (2009). Is it Williams syndrome? GTF2IRD1 implicated in visual-spatial construction and GTF2I in sociability revealed by high resolution arrays. *American Journal of Medical Genetics Part A*, 149A, 302-314.
55. Furuta, N., Ikeda, M., Hirayanagi, K., Fujita, Y., Amanuma, M and Okamoto, K. (2012). A Novel GJA1 Mutation in Oculodentodigital Dysplasia with Progressive Spastic Paraplegia and Sensory Deficits. *Intern Med*, 51(1), 93-8.

56. Duan, J. (2004). Polymorphisms in the trace amine receptor 4 (TRAR4) gene on chromosome 6q23.2 are associated with susceptibility to schizophrenia. *American journal of human genetics*, 75(4), 624-38.doi:10.1086/424887
57. Konstantareas, M. M. (1999). Chromosomal abnormalities in a series of children with autistic disorder. *Journal of autism and developmental disorders*, 29(4), 275-85.
58. Anney, R et al. (2010). A genome-wide scan for common alleles affecting risk for autism. *Human molecular genetics*, 19(20), 4072-82.doi:10.1093/hmg/ddq307
59. Hastings, P. J., Lupski, J.R., Rosenberg, S.M. and Ira, G. (2009). Mechanisms of change in gene copy number. *Nature reviews. Genetics*, 10(8), 551-64.doi:10.1038/nrg2593
60. Meechan, D. W., Tucker, E.S., Maynard, T.M. and LaMantia, A. (2009). Diminished dosage of 22q11 genes disrupts neurogenesis and cortical development in a mouse model of 22q11 deletion/DiGeorge syndrome. *PNAS : Proceedings of the National Academy of Sciences*, 106(38), 16434-45.doi:10.1073/pnas.0905696106
61. Pinto, D. et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature (London)*, 466(7304), 368-72.doi:10.1038/nature09146
62. Long, H., Zhu, X., Yang, P., Gao, Q., Chen, Y. and Ma, L. (2012). Myo9b and RICS Modulate Dendritic Morphology of Cortical Neurons. *Cerebral cortex (New York, N.Y. 1991)*doi:10.1093/cercor/bhr378
63. Wegiel, J et al. (2010). The neuropathology of autism: defects of neurogenesis and neuronal migration, and dysplastic changes. *Acta neuropathologica*, 119(6), 755-70.doi:10.1007/s00401-010-0655-4
64. Palmen, S et al. (2004). Neuropathological findings in autism. *Brain*, 127, 2572-83.doi:10.1093/brain/awh287
65. Glessner, J. T et al. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature (London)*, 459(7246), 569-73.doi:10.1038/nature07953
66. Zikopoulos, B. & Barbas, H. (2010). Changes in Prefrontal Axons May Disrupt the Network in Autism. *The Journal of neuroscience*, 30(44), 14595-14609.doi:10.1523/JNEUROSCI.2257-10.2010
67. Voineagu, I. et al (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature (London)*, 474(7351), 380-384.doi:10.1038/nature10110

68. Boulanger, L. M. (2009). Immune proteins in brain development and synaptic plasticity. *Neuron* (Cambridge, Mass.), 64(1), 93-109. doi:10.1016/j.neuron.2009.09.001
69. Shastry, B.S. (2008). Copy number variation and susceptibility to human disorders. *Molecular Medicine Reports*, 2, 143-147. doi:10.3892/mmr_00000074
70. Smith, C. L. , Bolton, A. & Nguyen, G. (2010). Genomic and epigenomic instability, fragile sites, schizophrenia and autism. *Current genomics*, 11(6), 447-69. doi:10.2174/138920210793176001
71. Nguyen, G.H. et al (2003). DNA stability and schizophrenia in twins. *Am J. Med. Genet. B Neuropsychiatry*, 120B (1), 1-10
72. Loveland, K.A., Bachevellar, J., Deborah, A.P. & Lane, D.M. (2008). Fronto-Limbic functioning in children and adolescents with and without autism. *Neuropsychologia*, 46(1), 49-62.
73. Charbonnier, F et al. (1998). Genomic organization of the human SPOCK gene and its chromosomal localization to 5q31. *Genomics* (San Diego, Calif.), 48(3), 377-80. doi:10.1006/geno.1997.5199
74. Ribasés, M et al. (2011). Contribution of LPHN3 to the genetic susceptibility to ADHD in adulthood: a replication study. *Genes, brain and behavior*, 10(2), 149-57. doi:10.1111/j.1601-183X.2010.00649.x
75. Konyukh, M et al. (2011). Variations of the candidate SEZ6L2 gene on Chromosome 16p11.2 in patients with autism spectrum disorders and in human populations. *PloS one*, 6(3), e17289. doi:10.1371/journal.pone.0017289
76. Haldeman-Englert C, Jewett T. In: Pagon RA, Bird TD, Dolan CR, and Stephens K. (2011) 1q21.1 Microdeletion. *GeneReviews*.
77. Mefford HC et al. (2008). Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med*, 359, 1685–99.
78. Stefansson, H. et al (2008). Large recurrent microdeletions associated with schizophrenia. *Nature*, 455, 232–6.
79. Goh, Y.S. and Grants, J.M. (2011). Mutations in the mediator subunit MED23 link intellectual disability to immediate early gene regulation. *Clinical Genetics*. doi: 10.1111/j.1399-0004.2011.01821.x
80. Alberini CM. Transcription factors in long-term memory and synaptic plasticity. *Physiol Rev* 2009; 89: 121–145.

81. Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A. et al (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, 70 :886-897
82. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011). Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. *Neuron* 70:863–885.
83. Joachim, H. et al. (2011). Genetic Heritability and Shared Environmental Factors Among Twin Pairs with Autism. *Arch Gen Psychiatry*, 68 (11), 1095-1102
84. Redcay E, Courchesne E (2005) When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biol Psychiatry* 58: 1-9
85. Bauman ML, Kemper TL (1985) Histoanatomic observations of the brain in early infantile autism. *Neurology* 35:866–867
86. Bailey AR, Giunta BN, Obregon D et al (2008) Peripheral biomarkers in autism: secreted amyloid precursor protein-alpha as a probable key player in early diagnosis. *Int J Clin Exp Med* 1:338–344
87. Lopez-Hurtado E, Prieto JJ (2008) A microscopic study of language-related cortex in autism. *Am J Biochem Biotechnol* 4:130–145
88. Caudle, S. E. (2012). Nonverbal Cognitive Development in Children With Cochlear Implants: Relationship Between the Mullen Scales of Early Learning and Later Performance on the Leiter International Performance Scales-Revised. *Assessment* (Odessa, Fla.)doi:10.1177/1073191112437594
89. Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J. et al. (2010). A genome-wide scan for common alleles affecting risk for autism. *Human Molecular Genetics*, 19 (20): 4072-4082.

ACADEMIC VITA

Khoo Su Jen
Undergraduate Honors Scholar

The Huck Institute of Life Sciences
Pennsylvania State University
206 Life Science Building
University Park,
State College, PA 16802

Tel: 814-441-7386
E-mail: ksk5116@psu.edu

Education

Pennsylvania State University, State College, PA

- Bachelor of Science in Biotechnology, 2009-present
- Minor in Microbiology, 2009-present

Research and Work Experience

Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, PA

- Undergraduate honors researcher for Dr Maria Krasilnikova, July 2010-present
- Research project: Effects and mechanisms of microsatellite repeats on chromosomal instability

Department of Genome Sciences, University of Washington, Seattle, WA

- Summer internship with Dr Santhosh Girirajan, May-July 2011
- Research project: Study of structural variants and large-scale genomic rearrangements contributing to autism.

Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, PA

- Teaching assistant for Dr Ola Sodeinde, Jan-May 2012
- Task: Conduct and guide students in microbiology-related experiments

Honors and Awards

- Outstanding scholar of Schreyer Honor's College working on thesis research, 2010-present
- Selected for Dean's List Academic Achievement, 2009-present
- Evan Pugh Honorary Award, 2012

Scientific Activities

- Competed in PSU Undergraduate Research Exhibition, 2011
- Participated in Summer Undergraduate Research Fellowship, 2011
- Engaged in Undergraduate Discovery Grant, 2011

Research Skills and Techniques

- Experience with culturing mammalian cells
- Carry out bacteria transformation and transfection of mammalian cells
- Analyze DNA fragments using 2D gel electrophoresis and Southern blot hybridization
- Study of bacteria morphology using light microscopy and gram staining
- Study of chemical properties using NMR and UV spectroscopy
- Apply chromatography methods to purify chemical compounds
- Perform array comparative genomic hybridization method to discover and characterize deletions and duplications in the human genome

Service Engagements

- Aided in environment preservation during Fresh Start Day of Service
- Partake in New Student Orientation Programs
- Volunteered in Schreyer Honor's Day of Service
- Performed sketch in Malaysian Culture Night
- Participated in educating deaf kids as a Celcom Youth Ambassador

References

Scott B. Selleck, MD, PhD (research advisor)
Professor and Head of Biochemistry and Molecular Biology
The Hucks Institute of Life Sciences
Pennsylvania State University
206D Life Sciences Building
University Park,
State College, PA 16802
Email: sbs24@psu.edu
Tel: (814) 867-3274

Maria Krasilnikova, PhD (research and thesis advisor)
Research Assistant Professor of Biochemistry and Molecular Biology
The Hucks Institute of Life Sciences
Pennsylvania State University
206 Life Science Building

University Park,
State College, PA 16802
Email: muk19@psu.edu
Tel: (814) 863-5555

David Gilmour, PhD (honors advisor)
Professor of Molecular and Cell Biology
The Hucks Institute of Life Sciences
Pennsylvania State University
465A North Frear Laboratory
University Park, PA 16802
Email: dsg11@psu.edu
Tel: (814) 863-8905

Santhosh Girirajan, PhD (intern advisor)
Postdoctoral Fellow of Genome Sciences
Howard Hughes Medical Institute Investigator
Department of Genome Sciences
Foege S413A, 3720 15th Ave NE
University of Washington School of Medicine
Seattle, WA 98195-5065
Email: sangi@u.washington.edu
Tel: (206) 685-7336