DNA COPY NUMBER VARIATION IN PSYCHOSIS TRIO SAMPLES USING BAC ARRAY CGH AND REAL TIME QUANTITATIVE PCR

by

BRIANNA LEIGH MELNYK

B.Sc. Hon., The University of British Columbia, 2003

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS OF THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

April 2006

Abstract

Schizophrenia and bipolar disorder are debilitating mental illnesses. Due to their high genetic predisposition, efforts have focused on attempting to find candidate loci. Numerous regions and loci have been suggested and investigated for potential candidate genes, but none have been found to be necessary or sufficient for the development of either disease. DNA copy number changes are often important in genetic disease. For example, changes in DNA copy number are linked to mental retardation (Klein et al., 2004) and cancer (Albertson et al., 2000). Bacterial artificial chromosome (BAC) array comparative genomic hybridization (aCGH) and real time quantitative PCR (RTqPCR) were used to explore copy number variation in 20 first episode psychosis trios (proband, mother and father) with probands affected with psychosis. The genome scan results showed 11 copy number differences (9 amplifications and 2 deletions) at seven loci. Retesting three of these seven loci with RTqPCR showed 18 amplifications and three deletions. The retested locus showing the most variation in copy number was the *lipoprotein A* gene. Recently, protein levels of Lp(a) were shown to be significantly increased in patients with schizophrenia, bipolar disorder and major depression (Emanuele et al., 2006). Comparison of the aCGH and RTqPCR results revealed that of the six trios with aCGH-detected aberrations within RTqPCR-tested loci, three were confirmed in the same samples and in the same direction. The results from this study contribute to the understanding of copy number variation in the human genome using trio sets as samples, and provide insight into different methods for copy number analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aCGH	ARRAY COMPARATIVE GENOMIC HYBRIDIZATION
ARJP	AUTOSOMAL RECESSIVE JUVENILE PARKINSONISM
BAC	BACTERIAL ARTIFICIAL CHROMOSOME
bp	BASE PAIR
BPD	BIPOLAR DISORDER
BSA	BOVINE SERUM ALBUMIN
COMT	*CATECHOL-O-METHYLTRANSFERASE*
CSV	COMMA-SEPARATED VALUES
C_t	THRESHOLD CYCLE
Cy3 dCTP	CYANINE 3 DEOXY-CYTIDINE-TRIPHOSPHATE
Cy5 dCTP	CYANINE 5 DEOXY-CYTIDINE-TRIPHOSPHATE
DAO	*D-AMINO ACID OXIDASE*
DISC1	*DISRUPTED IN SCHIZOPHRENIA 1*
DISC2	*DISRUPTED IN SCHIZOPHRENIA 2*
DNA	DEOXYRIBONUCLEIC ACID
DTNBP1	*DYSTROBREVIN BINDING PROTEIN 1*
EtOH	ETHANOL
FAM	6-CARBOXY FLUORESCEIN
FCGR3B	*Fc FRAGMENT OF IgG*
FISH	FLUORESCENCE IN SITU HYBRIDIZATION
G6PD	*GLUCOSE-6-PHOSPHATE DEHYDROGENASE*
G72	*D-AMINO ACID OXIDASE ACTIVATOR*
GRK3	*G PROTEIN RECEPTOR KINASE 3*
ISHDSF	IRISH STUDY OF HIGH DENSITY SCHIZOPHRENIA FAMILES
kb	KILOBASE
LPA	*LIPOPROTEIN A*
Mb	MEGABASE
MIA	MICROARRAY IMAGE ANALYSIS
NaOAc	SODIUM ACETATE
NRG1	*NEUREGULIN 1*
PCR	POLYMERASE CHAIN REACTION
PD	PARKINSON DISEASE
PGT	*HUMAN PROSTAGLANDIN TRANSPORTER*
PMT	PHOTOMULTIPLIER TUBE
PRKCZ	*PROTEIN KINASE C ZETA*
PRODH	*PROLINE DEHYDROGENASE*
RGS4	*REGULATOR OF G-PROTEIN SIGNALING 4*
RTqPCR	REAL TIME QUANTITATIVE POLYMERASE CHAIN REACTION
ROMA	REPRESENTATIONAL OLIGONUCLEOTIDE MICROARRAY ANALYSIS
RSF-1	*REMODELING AND SPACING FACTOR 1*
SCZ	SCHIZOPHRENIA
SDS	SODIUM DODECYL SULFATE
SKY	SPECTRAL KARYOTYPING
SMRT	SUB-MEGABASE RESOLUTION TILING
SNP	SINGLE NUCLEOTIDE POLYMORPHISM
SSC	SODIUM CHLORIDE/SODIUM CITRATE
STS	SEQUENCE TAGGED SITE
SYT7	*SYNAPTOTAGMIN VII*
TAE	TRIS ACETATE ETHYLENEDIAMINE TETRAACETIC ACID
TE	TRIS ETHYLENEDIAMINE TETRAACETIC ACID
VCFS	VELO-CARDIO FACIAL SYNDROME
ZFP37	*ZINC FINGER PROTEIN 37 HOMOLOG*

ACKNOWLEDGEMENTS

I am grateful to my supervisor, Robert Holt (BC Cancer Agency Michael Smith Genome Sciences Centre, Dept of Psychiatry UBC), for his guidance and patience.

The Brain Research Group (BCCA Michael Smith GSC) for technical assistance in the lab and with the computer:
~ Gary Wilson
~ Perseus Missirlis
~ Vik Chopra
~ Carri-Lyn Mead

For constructive suggestions and discussion during committee meetings:
~ William Honer, Dept of Psychiatry UBC
~ Weihong Song, Dept of Psychiatry UBC
~ Marco Marra, BCCA Michael Smith GSC

Dr. Honer was also essential as a collaborator for this project. His participation in the Early Psychosis Identification and Intervention Program made it possible for me to use the trio samples. Thanks also to the dedicated team involved in the Program.

Stephane Flibotte (BCCA Michael Smith GSC) for the design of MIA.

The Sequencing Group (BCCA Michael Smith GSC) for help with sample labeling, BAC end sequencing, DNA prep and primer ordering.

The Mapping Group (BCCA Michael Smith GSC) for help with the BAC re-array and LMPCR.

Even though the main products of my work with the Sequencing and Mapping groups were not used for the final thesis project, they were instrumental in helping me during the initial attempts and answering subsequent questions.

Donna Albertson, Randy Davis and Greg Hamilton (University of California at San Francisco Cancer Centre) for printing the HumArray 2.0 and help with array trouble-shooting.

Spencer Watson and the Wan Lam lab (BC Cancer Research Centre) for completing the high density array hybridization.

And last but not least, thank you to my family and friends who listened to endless chatter about those array thingies and digesting DNA. You may not believe that a B.Gym and a half B.A.Sc could help an aspiring M.Sc, but my father and sister are appreciated for a myriad of things, including *just write anything* and *get yourself a newt.* A special thank you to my mother, who consistently repeated the gentle reminder that this experience was not the most difficult thing I have lived through. Internal repetition of that reminder kept me going when nothing else helped. She also kept wondering why I was *eating* the DNA, but that is a story for another time.

DEDICATION

This piece of my life is dedicated to my aunt,

Darrell Susan Drexel
9 January 1948 – 27 May 2002

1.0 INTRODUCTION

1.1 SCHIZOPHRENIA AND BIPOLAR DISORDER

1.1.1 Symptoms and Diagnosis

Schizophrenia (SCZ) and bipolar disorder (BPD) are severe mental illnesses each of which affects approximately 1% of the population (reviewed in Walker et al., 2004). These diseases do not have any consistent genetic or neuropathological markers, so diagnosis is based on the presence of positive and negative symptoms (SCZ) or manic, depressive and mixed episodes (BPD) as outlined in the *DSM IV-TR* (2000). SCZ is divided into five subtypes based on the predominant symptom, while a diagnosis of BPD is derived from the most recent episode. The symptoms of SCZ include hallucinations and delusions (psychosis), disorganized speech and behaviour, and negative symptoms (affective flattening, alogia, avolition). BPD is characterized by a disorder in mood and episodes of mood disorder can be manic, depressive, hypomanic or mixed. Patients with BPD may also experience symptoms of psychosis.

1.1.2 Treatment

Treatment of patients with psychosis is usually through administration of antipsychotic medications. In general, use of antipsychotic drugs reduces psychotic symptoms, but can also result in multiple side-effects. The classical antipsychotic medications show a high affinity for D_2 receptors; the drugs bind these receptors tightly and dissociate slowly (reviewed in Miyamoto et al., 2005). In contrast, the second generation (atypical) antipsychotics have multiple sites of action (including dopamine, serotonin, histamine and glutamate receptors) and show faster dissociation rates from the D_2 receptors (reviewed in Miyamoto et al., 2005). Both types of drugs are effective against positive symptoms (hallucinations and delusions), but the atypical medications are better for the treatment of negative symptoms (reviewed in Miyamoto

et al., 2005). Fewer extra-pyramidal symptoms (parkinsonism, dystonia, akathesia, tardive dyskinesia) are caused by the atypical drugs, but they can result in weight gain, diabetes or cardiovascular problems (reviewed in Baldessarini and Tarazi, 2001). For these reasons, a diagnosis of SCZ or BPD with psychosis is usually followed by treatment with atypical antipsychotic medication (risperidone, olanzapine or quetiapine). Even though some patients do not respond to drug treatment, or experience a worsening of symptoms, pharmacotherapy remains the most efficient way to treat SCZ and BPD (reviewed in Baldessarini and Tarazi, 2001).

1.1.3 Heritability

Although environmental factors do contribute to disease risk, SCZ and BPD are highly heritable. It is estimated that the heritability of SCZ is between 82-85% (Cardno et al., 1999), and the heritability of BPD is 85% (McGuffin et al., 2003). This means that 82-85% of the phenotypic variation in the disease is caused by genetic variation (King and Stansfield, 1997). However, even with this evidence for a strong genetic component, there is no single mutation that has been found to cause any case of disease. Nor has research into other potential biological causes (e.g. changes in brain structure, function or cognition) yielded definitive answers (Heinrichs, 2001).

In the literature to date, several chromosomal locations have been implicated as risk regions for SCZ and BPD. Most of the candidate chromosomal regions have been found through linkage studies with some clues provided by cytogenetic approaches. Follow-up studies attempt to identify mutations in individual genes that confer risk of disease development.

1.1.4 Linkage Analysis

Linkage analysis uses genotype data from large families with many cases of a disease of interest to find co-segregation between genetic markers (single nucleotide polymorphisms [SNPs], microsatellites, etc.) and the disease. Currently, linkage studies have identified at least 9 regions associated with SCZ (reviewed in Sklar, 2002; Owen et al., 2004) and 8 associated with BPD (reviewed in Sklar, 2002) (summarized in Table 1.1). However, reports from linkage studies are controversial because the findings have not been consistently confirmed (reviewed in Sklar, 2002). For example, a study with Celtic- and German-Canadian families found evidence of SCZ linkage to chromosome 1q21-22 (Brzustowicz et al., 2000). However, another group reported no significant linkage to chromosome 1q after genotyping 16 microsatellite markers in multiple populations (Levinson et al., 2002).

TABLE 1.1. LINKAGE REGIONS ASSOCIATED WITH BPD OR SCZ (adapted from Sklar, 2002; Owen et al., 2004).

Chromosomal Band	Disease
1q21-22, 6p22-24, 6q21-22, 8p21, 10p11-15, 13q32, 22q11-13	SCZ
4p15-16, 12q23, 13q32, 18q12, 18q22, 21q22, 22q11-12	BPD

Regions identified through linkage reports are often investigated for candidate genes. To determine potential candidate genes within a region of interest, smaller areas are studied using fine mapping and the disease risk of the genes are investigated using haplotype association. The alleles of specific markers (within or close to the gene of interest) and haplotypes constructed from the alleles, are analyzed for their distributions in patient and control samples. Even though there is high heritability associated with SCZ and BPD, no individual gene has been found to be responsible for disease. However, some potential candidates for increasing disease risk have been identified (Table 1.2).

TABLE 1.2. RISK GENES FOR BPD AND SCZ (adapted from Harrison and Weinberger, 2005).

Gene	Band	References	Disease
Neuregulin 1 (NRG1)	8p12	Stefansson et al., 2002	SCZ
Dystrobrevin binding protein 1 (DTNBP1)	6p22.3	Straub et al., 2002	SCZ
Catechol-o-methyltransferase (COMT)	22q11.21	Shifman et al., 2002	SCZ
D-amino acid oxidase (DAO)	12q24.11	Chumakov et al., 2002	SCZ
D-amino acid oxidase activator (G72)	13q33.2	Chumakov et al., 2002	SCZ
G-protein receptor kinase 3 (GRK3)	22q12.1	Barrett et al., 2003	BPD
Regulator of G-protein signaling 4 (RGS4)	1q21	Chowdari et al., 2002	SCZ
Disrupted in schizophrenia 1 (DISC1)	1q42.2	Hennah et al., 2003	SCZ

For example, interest in *neuregulin 1* (*NRG1*) as a risk gene for SCZ was generated by one of multiple linkage scans highlighting chromosome 8p (Stefansson et al., 2002). Subsequent fine-mapping in this study identified *NRG1* as a possible susceptibility gene. The authors found evidence of association between a *NRG1* haplotype and SCZ in an Icelandic population where the core haplotype overlapping *NRG1* was found at a higher frequency in SCZ patients when compared to controls. Association of this risk haplotype was evaluated in an independent population and a portion of the risk haplotype identified in the Icelandic patient group was also found at a higher frequency in Scottish patient population (Stefansson et al., 2003). Stefansson et al. (2002) also investigated heterozygous *NRG1* mutant mice and found that the behavioural phenotype was consistent with other SCZ mouse models. For example, the *NRG1* mutant mice were hyperactive and had impaired pre-pulse inhibition. In addition, clozapine treatment reduced the hyperactivity of mutant mice but did not change the activity of normal mice. The results from these papers show how linkage analysis and fine-mapping can be used to suggest risk loci for psychiatric disease.

1.1.5 Cytogenetic Analysis

Cytogenetic reports related to SCZ and BPD are generally anecdotal and based on isolated cases because patients with psychiatric disorders are not usually referred for cytogenetic testing (reviewed in MacIntyre et al., 2003). Cytogenetic aberrations implicated in mental illness include translocations (reciprocal and unbalanced), fragile sites, insertions, inversions and mosaicism. A small number of cytogenetic abnormalities identified have been supported by further evidence of association to disease. For example, researchers discovered that a balanced translocation in a large Scottish family disrupts two genes (*DISC1* and *DISC2*) on chromosome 1 and segregates with SCZ and affective disorders in this family (Millar et al., 2000; Blackwood et al., 2001). Subsequently, using SNP markers and haplotype analysis, *DISC1* was found to be associated with schizophrenia (Hennah et al., 2003).

A seminal cytogenetic finding in SCZ and BPD is related to velo-cardio-facial syndrome (VCFS). This syndrome is caused by a 3 Mb deletion of chromosome 22q11. Symptoms of VCFS include craniofacial abnormalities, cardiovascular defects, learning disabilities and behavioural disorders (reviewed in Shprintzen et al., 2005). In addition to physical abnormalities, patients with VCFS develop psychiatric disorders. The first report was published in the early 1990's (Shprintzen et al., 1992). Another group of researchers assessed a cohort of 50 adults with VCFS and found an increased rate of schizophrenia (24%) and psychosis (30%) when compared to the general population rate (Murphy et al., 1999). The phenotype of schizophrenia in patients with VCFS is not different from SCZ in patients without VCFS (Bassett et al., 2003). The high degree of risk for people with VCFS is only surpassed by the risk for people with a monozygotic twin or two parents with SCZ or BPD (reviewed in Bray and Owen, 2001).

Because of the high rate of psychosis in patients with VCFS, efforts have been aimed at identifying genes within the 3 Mb deletion that could be disease risk candidates. Within the approximate boundaries of the microdeletion (chr22:16,865,871-20,930,922), as viewed on the University of California at Santa Cruz (UCSC) Genome Browser (Karolchik et al., 2003), there are approximately 30 genes. *Catechol-o-methyltransferase* (*COMT*) and *proline dehydrogenase* (*PRODH*) have been investigated as potential risk genes for psychosis. Association between two haplotypes within *PRODH* and schizophrenia was found in a Chinese population (Li et al., 2004). In addition, missense mutations within *PRODH* influenced the functional activity of the gene product POX (Bender et al., 2005). The authors reported that some of the mutations resulting in decreased POX activity are polymorphisms associated with schizophrenia. The *COMT* gene product metabolizes catecholamines, including dopamine (Axelrod and Tomchick, 1958). A few papers have shown an association between *COMT* and SCZ using genetic marker and haplotype analysis. In a study of a population of Ashkenazi Jews, researchers reported significant association between SCZ and a particular *COMT* haplotype (Shifman et al., 2002). Another group attempted to replicate these findings in the Irish study of high-density schizophrenia families (ISHDSF) (Chen et al., 2004). The authors investigated three of the SNPs examined by Shifman et al. (2002), but found no association of any SNP with SCZ in the ISHDSF. The linkage disequilibrium values for the Irish families showed a similar trend to that of the Ashkenazi Jews, but were lower. In addition, Chen et al. (2004) did not find strong association between the haplotype identified by Shifman et al. (2002) and SCZ. The differences between the studies suggest that population history may influence the genes that are risk factors. In addition to genetic marker analysis, researchers have examined the effects of *COMT* genotype on cognitive function. Several studies using cognitive testing revealed better performance of

psychosis patients (Nolan et al., 2004; Rosa et al., 2004) and 22q11 deletion patients (Bearden et al., 2004) with a methionine at amino acid 158 of the *COMT* gene product.

The increase in risk for the development of psychosis in patients with VCFS suggests that changes in gene copy number are important. Evidence exists that *COMT* and *PRODH* are important for the development of SCZ and psychosis, but it remains to be seen whether other genes that show copy number changes in patients are risk factors for disease.

1.2 DNA COPY NUMBER AND GENETIC DISEASE

Linkage and cytogenetic analyses are powerful techniques to identify disease genes in which copy number aberrations segregate with disease. Examples of disease genes found through these techniques are *a-synuclein* and *parkin*. The *a-synuclein* gene was first pinpointed after a linkage study found a connection between Parkinson disease (PD) and chromosome 4q21-23 (Polymeropoulos et al., 1996). A copy number increase of the gene was found in independent families with PD (Singleton et al., 2003; Farrer et al., 2004). For the *parkin* gene, linkage studies with autosomal recessive juvenile parkinsonism (ARJP) highlighted chromosome 6q25.2-25.7 (Matsumine et al., 1997). Fine mapping refined the linkage region to the *parkin* gene (Saito et al., 1998) and this locus was also found to be deleted in ARJP patients (Kitada et al., 1998).

Another example of how copy number influences disease susceptibility is the *Fc fragment of IgG* (*FCGR3B*) gene and lupus nephritis in humans. Findings in rats indicated that reduced copy number of *Fcgr3* resulted in macrophage overactivity and glomerulonephritis susceptibility. This finding stimulated the authors to investigate families showing mendelian errors for *FCGR3B* polymorphisms (Aitman et al., 2006). The authors found that reduced copy number of *FCGR3B* was a risk factor for lupus nephritis. This study and the PD studies above, show that copy number of a single gene can influence a complex disease.

1.2.1 Array Comparative Genomic Hybridization

Array comparative genomic hybridization (aCGH) is a cytogenetic method used to detect DNA copy number changes (deletions and duplications); this technique is based on chromosomal CGH (Kallioniemi et al., 1992). Previously, chromosome banding and fluorescence *in situ* hybridization (FISH) were used to determine copy number changes in the genome, but these microscope techniques are limited by low resolution (5-20 Mb) and restricted application. For example, whole chromosomes are needed for both procedures and the target for FISH must be known prior to investigation. Instead of using metaphase whole chromosome spreads, aCGH employs the use of DNA clones printed on slides. An example of this technology is the HumArray 2.0. This array contains 2460 bacterial artificial chromosome (BAC) clones printed in triplicate on a glass slide and represents approximately 15% of the genome (Snijders et al., 2001). Each spot on the HumArray 2.0 is a polymerase chain reaction (PCR) representation of a BAC clone. Instead of using full-length BAC clones (~175kb), which would be difficult to print onto the array because of the viscous nature of the spotting solution, ligation-mediated PCR (LMPCR) is used. This technique cuts each BAC clone into fragments and amplifies each piece. A mixture of the amplified fragments from each BAC clone is printed onto a slide. The chromosomal location of each BAC is known through sequence tagged site (STS) markers (Snijders et al., 2001).

To determine copy number changes in samples, genomic DNA from two sources (patient or unaffected, and reference) is labeled with fluorescent nucleotides. Usually, DNA from a reference sample is labeled with Cyanine-5 deoxycytidine triphosphate (Cy5-dCTP; red) and the patient or unaffected control sample is labeled with Cyanine-3 deoxycytidine triphosphate (Cy3-dCTP; green). Equal amounts of labeled DNA are mixed together and incubated with human Cot-1 DNA to block repetitive DNA sequences (Figure 1.1). The

incubated mixture is hybridized to the microarray and the ratio of green to red fluorescence of each spot is measured. Since equal amounts of DNA are used, the expected ratio in the absence of a copy number change is 1:1. A significant deviation from this ratio represents a DNA copy number aberration. As seen in Figure 1.2, the red and green spots outlined in white represent sex chromosome copy number differences within family 163. both scans show a copy number decrease of a Y chromosome clone in the mother.

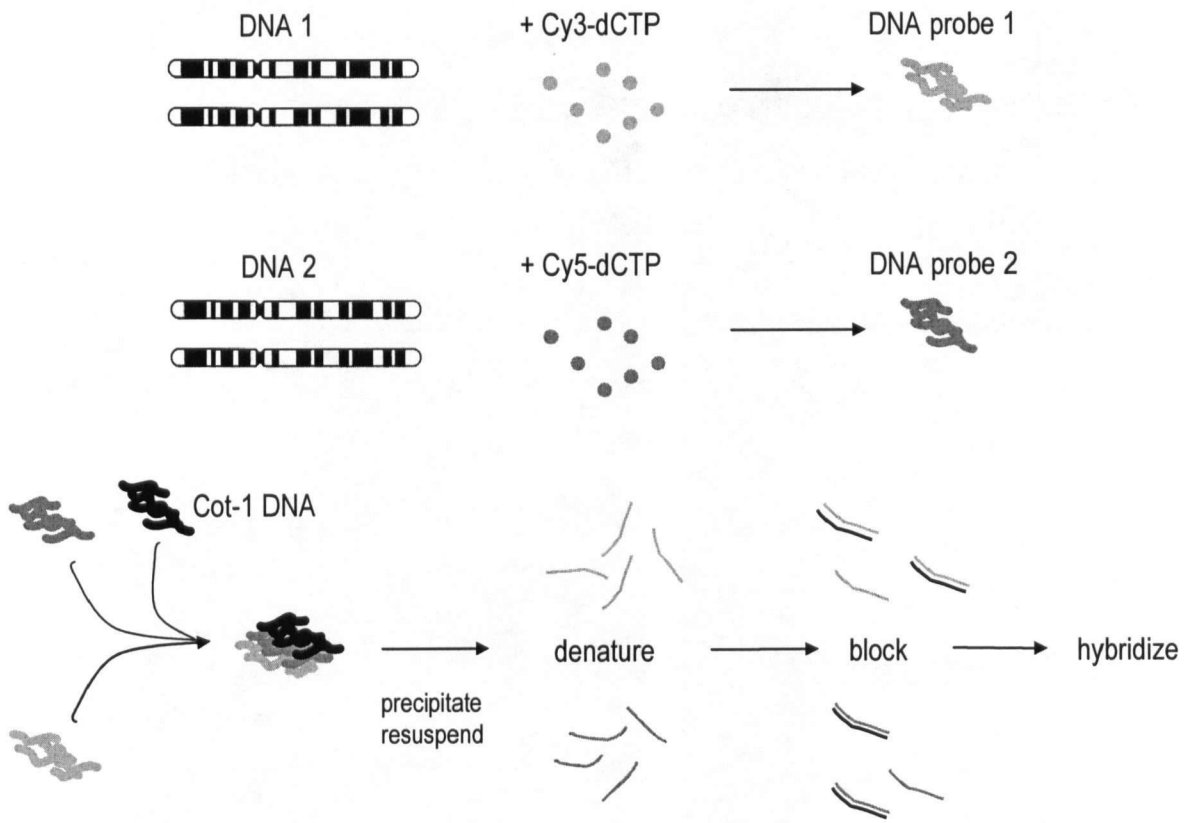

FIGURE 1.1. THE ARRAY COMPARATIVE GENOMIC HYBRIDIZATION PROCEDURE. DNA from two sources is labeled with fluorescent nucleotides (Cy3-dCTP and Cy5-dCTP). The labeled DNA is mixed with Cot-1, denatured and blocked. The mixture is hybridized to the array.

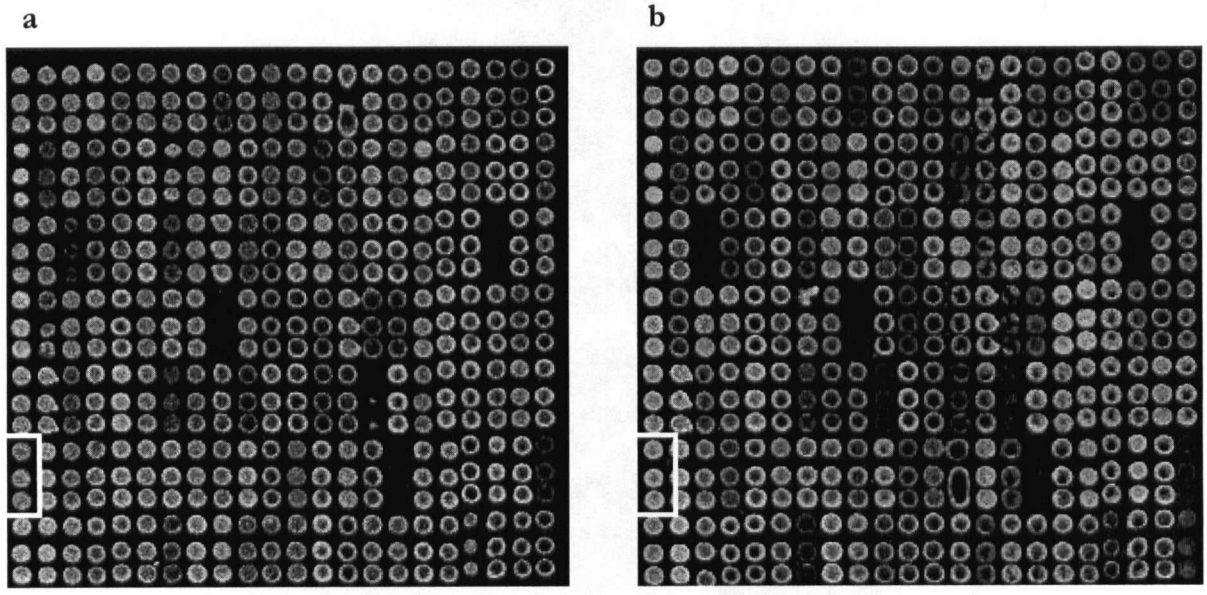

FIGURE 1.2. SCANS OF A HUMARRAY 2.0 SUBARRAY. (a) The family 163 father/mother hybridization. Red spots show a copy number decrease of a chromosome Y clone in the mother (**Cy3**) as compared to the father (**Cy5**). (b) The family 163 proband/mother hybridization. Green spots show a Y chromosome clone copy number decrease in the mother (**Cy5**) as compared to the proband (**Cy3**). The black spaces on the array are regions without spotted DNA because the plates used to print the spots were not completely full.

1.2.2 Human Copy Number Polymorphisms

In recent years, DNA copy number analysis has become a hot topic. Interest ranges from the differences between humans and other primates to normal CN polymorphisms in humans. Using the HumArray 2.0 (Snijders et al., 2001), researchers found 63 sites in the human genome that showed CN differences when compared to other primates (gorilla, pygmy chimpanzee, chimpanzee and orangutan) (Locke et al., 2003). Most variant sites were not common to all primate groups. For example, 30 variant sites were specific to orangutans. Of the 63 sites found, fourteen were investigated further and 7/7 duplications and 2/7 deletions were confirmed copy number changes. In addition, six of the 9 confirmed sites are proximal to human genes. A subsequent study of gorilla, chimpanzee and human samples found 63

chromosomal segments with DNA copy number differences using full genome coverage high density aCGH (Wilson et al., 2006b). Two of the CGH-detected duplicated genes were tested and both were validated as copy number amplifications in human DNA using real time quantitative PCR (RTqPCR).

Recently, two studies have been published that describe DNA copy number variation in the human population. The first reported a total of 76 unique copy number polymorphism sites in the genomes of 20 individuals (Sebat et al., 2004). Using representational oligonucleotide microarray analysis (ROMA, a technique similar to aCGH) the authors observed a total of 221 differences at the 76 sites. Fluorescence in situ hybridization (FISH) was used to validate the findings; nine of 12 sites tested were confirmed (two additional sites were confirmed using an alternate method). Seventy genes showing copy number variation were found within the 76 sites. The second study used aCGH with 55 individuals and found 255 clones with copy number differences (Iafrate et al., 2004). Of the identified clones, 14 were near human disease loci. The results from these two studies showed that human genome variation is more common than originally believed. However, the concordance between the two reports was low. Only 11 loci were found in common and there are some limitations of these methods (Carter, 2004). This could be explained because in these two studies neither ROMA nor aCGH have a high resolution compared to arrays that have almost complete coverage of the human genome (e.g. SMRT array (Ishkanian et al., 2004) or higher density ROMA array). The array used for CGH had one clone every 1 Mb (Iafrate et al., 2004), while the ROMA array had one probe every 35 kb (Sebat et al., 2004).

These results formed the groundwork of a database for normal copy number variation in humans (http://projects.tcag.ca/variation). The database is designed to show copy number variants within or overlapping a user-defined genomic region. For example, a query of a locus

that I have found to have copy number variation, the *lipoprotein A* (*LPA*) gene (NM_005577), shows an overlapping segmental duplication cluster of 14,270 bp (DC2038).

1.2.3 Tissue Specific Differences

Tumours are excellent examples of cell populations that show changes in DNA copy number relative to normal somatic tissues. A recent publication on ovarian cancer showed that borderline tumours and serous adenocarcinomas had more aneuploid cells than surface epithelium or inclusion cysts (Korner et al., 2005). The results also showed an elevated level of whole chromosome changes in tumours, but did not report on single gene amplifications or deletions. Another publication showed that amplification of *remodeling and spacing factor 1* (*Rsf-1*) in ovarian tumours led to shorter survival times in patients (Shih Ie et al., 2005). Given the expansive evidence of DNA copy number variation present in tumours, there is the possibility that variation between non-cancerous tissues exists. However, this phenomenon has not been investigated in normal tissues.

1.3 OBJECTIVES AND PURPOSE

My objectives for this project included: (1) complete array comparative genomic hybridization with 20 trios in which the proband suffers from psychosis; (2) identify putative inherited and spontaneous copy number aberrations within the trio; (3) design real time quantitative PCR primers and probes for putative aberrations; (4) validate the differences using RTqPCR. The purpose of this study was to identify candidate genes for psychosis that could be used as therapeutic targets or for the design of diagnostic tools.

2.0 METHODS AND MATERIALS

2.1 SUBJECT SET

The subjects for this project consisted of sets of trios. A trio is comprised of a youth (proband) aged 14-35 who has experienced a first psychotic episode (with less than 12 weeks exposure to antipsychotic medication) and his/her biological parents. Probands from the South Fraser region (Surrey, White Rock, Delta, Langley) were ascertained through referrals to the Early Psychosis Identification and Intervention (EPII) Program at Peace Arch Hospital (White Rock, British Columbia, Canada). The majority of diagnoses of probands included schizophrenia, schizoaffective disorder and affective disorders (bipolar and major depression) with psychosis. A detailed family history of psychiatric illness had not been completed for these trios. Refer to Appendix A (Table A1) for details about the proband used in this project.

2.2 SUBJECT SAMPLES

Blood samples from all three family members were sent to the BC Cancer Agency Genome Sciences Centre (BCCA GSC). Genomic DNA was extracted from lymphocytes for subsequent experiments (G. Wilson). In addition, some DNA samples arrived pre-extracted from the Kennedy lab in Toronto, Ontario, Canada. These samples were also originally from the EPII program at Peace Arch Hospital. When there was sufficient extracted DNA from subjects, it was digested with *MseI* enzyme prior to aCGH. This initial digestion improves hybridization and reduces the scatter of values around the 1:1 ratio by creating uniform DNA fragment sizes.

2.2.1 MseI Digestion, Precipitation and Resuspension

Genomic DNA from subjects (1 μg), 1 μL *MseI* enzyme (10U/μL, New England Biolabs), 1.2 μL bovine serum albumin (BSA, 100X, NEB) and 1.2 μL 10X NEB Buffer 2 were combined in

1.5 mL eppendorf tubes with a total reaction volume of 40 μL, or 96-well plates (Greiner) with a total reaction volume of 12 μL, and incubated at 37°C overnight. After 10 minutes at 70°C, 1 μL glycogen (20 mg/mL), 4 μL sodium acetate (NaOAc, 3M, pH 7, Sigma; 1.2 μL for 96-well plate) and 100 μL 95% ethanol (EtOH; 30 μL for a 96-well plate) were added. Eppendorf tubes were stored at -20°C for 10 minutes followed by centrifugation (15 minutes, 16100 x g). Plates were stored at -20°C for 30 minutes followed by centrifugation (45 minutes, 4°C, 2840 x g). The pellets were decanted, washed with 900 μL 80% EtOH (200 μL for 96-well plate) and allowed to air dry. Pellets in eppendorf tubes were resuspended in 12 μL of 10:1 TE (10 mM Tris-HCl, 1mM ethylenediamine tetraacetic acid, pH 8) and stored at 4°C overnight. For resuspension of the pellets in plates, 15 μL of TE was added to each well and the plate was incubated for 20 minutes at 37°C. After incubation, the plate was placed on the IKA-Vibrax plate agitator (Janke & Kunkel) for 5 minutes at 750 rpm. A quick spin (1 minute, 4°C, 805 x g) was followed by storage at 4°C until the DNA was qualified and quantified.

2.2.2 Quantification of Digested DNA

Dilutions (in TE) were made for each digested sample (1:100 and 1:1000) and combined with an equal amount of 1:200 pico green (10 μL) in a 384-well black skirted MJR plate. Dilutions (in TE) of calf thymus DNA were used to make the standard curve (0.7 ng/μL, 0.4, 0.2, 0.1, 0). Fluorescence was measured using the VICTOR^{3}V 1420 Multilabel Counter (Perkin-Elmer), and analysis was completed using Microsoft Excel.

2.2.3 Qualification of Digested DNA

A 0.7% agarose gel in 1X TAE (40 mM Tris acetate, 2 mM ethylenediamine tetraacetic acid) was loaded with 1 μL of sample (with 0.5 μL loading buffer and 3.5 μL TE). The samples were run

for 1.5 hours at 100 V. The gel was stained with SYBR green for 30 minutes (25 μL in 250mL 1X TAE) and then scanned using the FluorImager 595 (Molecular Dynamics).

2.3 MICROARRAY COMPARATIVE GENOMIC HYBRIDIZATION

2.3.1 Labeling

Labeled DNA fragments were prepared using the Invitrogen Bioprime labeling system. Two individual sample DNAs (400 ng each) were each combined with 10 μL of 2.5X random hexamer primers and dH_2O in separate tubes (for a total reaction volume of 20.5 μL). The DNA was denatured at 100°C for 10 minutes, then 2.5 μL of dNTP mix (2 mM dATP, 2 mM dGTP, 2 mM dTTP, 1.2 mM dCTP) was added. Cy3-dCTP (1 nmol) was added to one sample of DNA and Cy5-dCTP (1 nmol) was added to the other sample of DNA. Klenow fragment (1 μL of 40 U/μL) was added to each reaction and they were placed in a water bath in an air incubator overnight (~18 hours, 37°C). The reactions were pooled and the unincorporated nucleotides were removed using Amersham ProbeQuant Sephadex G-50 columns. To ensure blockage of repetitive DNA sequences before hybridization to the array, 140 μL of 500 ng/μL Cot-1 DNA (Invitrogen) was added to the eluate from the columns; 65 μL of NaOAc (3M, pH 5.2) and 921 μL of 95% EtOH were added and the reactions were placed on ice (10 minutes) to facilitate precipitation, and then centrifuged (30 minutes, 4°C, 18000 x g). The supernatants were decanted, pellets air-dried and the walls of the tube were dried with tissue paper. The pellets were resuspended in a mixture of 19.2 μL DIGEasy hybridization solution (Roche), 2.4 μL sheared salmon sperm DNA (10 mg/mL, Invitrogen) and 2.4 μL 100X BSA (NEB). The DNA was denatured in a dark water bath (10 minutes, 85°C), and repetitive sequences were blocked by the Cot-1 DNA present in the reaction during re-annealing in the incubator water bath (60-90 minutes, 45°C).

For the trios, an example labeling scheme is shown in Figure 2.1. This method of labeling ensures that each trio member is labeled with both dyes, but reduces the number of hybridizations needed for dye flip validation. Three arrays were completed for each trio, and each sample was labeled with each dye. For example, one array may have been the mother labeled with Cy3 (black arrow) and the father labeled with Cy5 (grey arrow). The second array would be a hybridization between the father-Cy3 and proband-Cy5. The last array is between the proband-Cy3 and the mother-Cy5. However, I did the hybridization experiments blind to sample identification. Therefore, the labeling scheme may have been different for each trio.

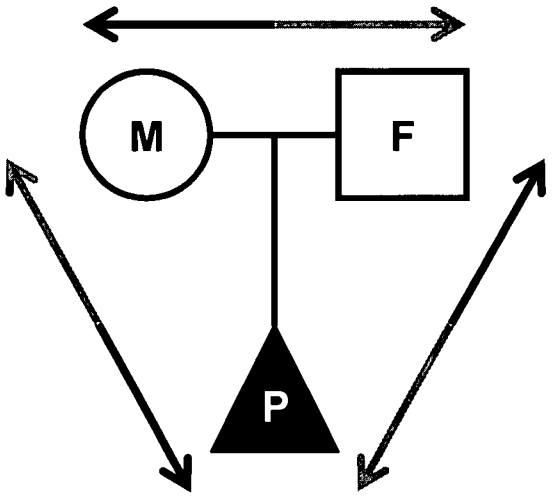

FIGURE 2.1. TRIO LABELING SCHEME. Each trio member is hybridized against the other two, instead of against a reference DNA. ○ = mother, □ = father, △ = proband (unknown sex). △ = unaffected, ▲ = affected.

2.3.2 Hybridization

To ensure a consistent temperature during application of the hybridization sample to the array, a working surface and the hybridization cassette were warmed in the incubator to 45°C. Each slide was exposed to 260,000 μJ of ultraviolet radiation prior to hybridization to link the DNA to the slide. Lifter slips (Erie Scientific) were cleaned with dH_2O and 80% EtOH. The

hybridization samples were removed from the incubator and placed in a dry block at 45°C. The working surface and hybridization cassette were removed from the incubator; and 22 μL of 2X sodium chloride/sodium citrate (SSC) was added to each reservoir of the hybridization cassette to ensure a humid environment for the slide. The slide was placed in the cassette and the lifter slips were placed over the arrays. The hybridization sample was applied to a side of the lifter slip without lifter rails (one reaction pair per array). A small amount of extra 2X SSC was added to the reservoirs and the hybridization cassette was assembled. The entire cassette was immersed in the water bath in the air incubator (48-72 hours, 45°C).

2.3.3 Washing

The hybridization cassette was removed from the water bath, dried and disassembled quickly. The lifter slips were washed off with a solution pre-warmed to 45°C (80% DIGEasy hybridization solution, 2X SSC, pH 7.0). The slide was then immersed in a Copelin jar with the same solution for 15 minutes at 45°C. The slide was transferred to a new Copelin jar (0.1X SSC, 0.1% sodium dodecyl sulfate [SDS], pH 7.0, room temperature). Fresh solution was added each time to complete three separate 5 minute washes. The Copelin jar was covered in foil during the washes to minimize exposure of the arrays to light. After the third wash, the slide was transferred to a new Copelin jar (0.1X SSC, room temperature). Fresh solution was added each time to complete four separate 30 second washes. The final wash was 5 seconds in a Copelin jar with 18 MΩ H_2O. The slide was then placed in a 50 mL conical tube and spun for 7 minutes at 350 **x** g.

2.3.4 Scanning

Each slide has two complete arrays, and each array is composed of sixteen 22 spot X 21 spot subarrays. Using a microarray scanner (ScanArray Express, Perkin-Elmer), each array was scanned multiple times at 10 μm and 5 μm resolution. Typically, initial settings for the Cy3 and Cy5 photo-multiplier tube (PMT) gains were 68% and 62%, respectively. These settings were adjusted depending on the brightness and colour of the scan images. A tan or yellow composite image (combination of Cy3 and Cy5 images) indicates a balance between the two fluorophores. If the composite image was red, the difference between the PMT values was increased; if the composite image was green, the difference between the PMT values was decreased.

2.3.5 Analysis

The Cy3 and Cy5 array images were analyzed using Microarray Image Analysis (MIA), a program designed at the BCCA GSC. For an in depth description of MIA, see Wilson et al. (2006b). Briefly, the program was designed to find the 16 subarrays of each array and then the individual spots within each subarray (addressing). This was accomplished by analyzing Fourier transformations of spot spectra and using granulometry. Tiles are placed on the array to match the spot size and the inter-spot distance was calculated by the addressing analysis. The Seeded Region Growing algorithm was used to identify the pixels that contribute to each spot (segmentation). Each pixel within the tile was labeled as a spot, background or artifact pixel. Then, each pixel intensity was calculated. An average intensity for the spot was calculated from all "spot" pixels. The $\log_2$ratio for each spot was calculated as:

$$\log_2 \text{ratio} = \log_2 \frac{\text{Cy5 intensity}}{\text{Cy3 intensity}}$$

The data were combined in a comma-separated values (CSV) output file and used for the scan selection step. MIA chooses the CSV file from the best scan of each array (determined empirically based on the standard deviation of each array) and uses this for breakpoint analysis. In breakpoint analysis, each CSV output file is analyzed for regions of the genome with multiple consecutive clones that are significantly above or below the 1:1 DNA copy number ratio. The CSV output file from the best scan was also chosen for population of the data table.

2.3.6 Data Handling

The MIA CSV output files were stored in a table (Appendix B, Table B1) within a relational database (MySQL). Queries of the data were designed to identify clones with a $\log_2$ratio that is 3.0 or more standard deviations away from the mean $\log_2$ratio of all clones on the array. Examining the output, I identified putative (a) spontaneous (i.e. found in the proband, but not the parents), and (b) inherited (i.e. found in the proband and one parent) aberrations. In addition, the data were inspected by eye for aberrations involving single clones or multiple consecutive clones.

2.3.7 SMRT Array CGH

Spencer Watson (Array CGH Laboratory, BC Cancer Research Centre) completed the sub-megabase resolution tiling (SMRT) array hybridization using DNA from the trio 456 proband and a male reference (Novagen). A full description of the array (version 1) and hybridization methods is in Ishkanian et al. (2004). Briefly, the single slide array (version 2) contains more than 26,000 overlapping BAC clones spotted in duplicate. With this number of clones, the coverage of the human genome is 1.44 fold and the theoretical resolution is 80 kb (Ishkanian et al., 2004; Krzywinski et al., 2004). Array resolution is determined by the size of the query genome divided by the number of single copy clones on the array. Selection and validation of

the spotted BAC clones was accomplished through *Hin*dIII fingerprint mapping (Krzywinski et al., 2004).

2.4 REAL TIME QUANTITATIVE PCR

2.4.1 Probe and Primer Design

Based on a high ranking (see Results section for a description of ranking procedure) and gene content, several putative aberrations were chosen for validation by real time quantitative PCR. The clone name was queried in the UCSC Genome Browser (Kent et al., 2002) with the repeat masker track on. The repeat masker track identifies regions of DNA containing short interspersed nuclear elements, long interspersed nuclear elements, micro-satellites, transfer-RNA, and other repeat family sequences using the Repbase collection of repetitive sequences from the Genetic Information Research Institute (Jurka et al., 2005). If the Browser found only fragments of the clone (200-500 bp), a region of ~200,000 bp around the fragment was searched for genes. Clone fragments are located on the reference genome based on end sequence matches of the clone. These short matches are from only a single end read, and because I did not know which end of the clone the sequence was from, I searched the 200 kb genomic region around the hit. Short stretches of DNA within the gene, but without repetitive DNA were assessed for primer/probe design. Regions of approximately 500 bp without repetitive DNA were (1) assessed for per cent identity to other genomic regions using the BLAST-like alignment tool (Blat) on the UCSC Genome Browser (Kent, 2002), (2) assessed visually to avoid long stretches of mono-nucleotides, and (3) divided into three 100 bp sub-regions for final primer/probe selection (according to instructions from Applied Biosystems). Each sub-region was assessed using Primer3 (Rozen and Skaletsky, 2000). Figure C1 is an example of a region queried for RTqPCR suitability. Once a 500 bp region was chosen and analyzed, the sequence

was sent to Applied Biosystems Assays-by-Design service for RTqPCR primer/probe design and synthesis. ABI chooses a 100 bp sub-region for the design of each primer/probe set; the solution sent back from ABI is a mixture of the primers and 6-carboxy fluoroscein (FAM) labeled probe.

2.4.2 PGT and G6PD Control Probes

Primer/probe sets for *PGT* and *G6PD* were used as controls. The *prostaglandin transporter* gene (*PGT* or *SLCO2A1*) is located on chromosome 3, and is present as two copies in each person (Lu and Schuster, 1998). However, the *glucose-6-phosphate dehydrogenase* gene (*G6PD*) is a chromosome X gene (Mason et al., 1990) and therefore present in only one copy in males. Using these two genes as controls for RTqPCR, I compared the values obtained with my designed probes to these loci of known copy number.

2.4.3 Sample Dilutions

Each of the 60 sample DNAs, as well as control male and female DNA, were diluted to 20ng/μL stocks (in TE). These dilutions (7.5 μL) were then used to make the 1ng/μL stock (in 142.5 μL TE) for RTqPCR. The control male and female samples used in this experiment are mixtures of DNA from multiple individuals of the same gender and they are commercially available from Novagen.

2.4.4 Master Mix

The master mix for each primer/probe set was prepared fresh for each experiment. The mix contained 240 μL TaqMan Universal PCR Master Mix (2X, Applied Biosystems), 19.2 μL primer/probe (to a final concentration of 0.7 μM for the primers and 0.2 μM for the probe), and 100.8 μL dH_2O. In addition, mixes for the PGT and G6PD loci were prepared. Each of these mixes contained 240 μL TaqMan Universal PCR Master Mix (2X), 4.8 μL of 20 μM FAM probe,

6.72 μL each of 50 μM forward and reverse primers, and 101.76 μL dH_2O. The Universal Master Mix contains: AmpliTaq Gold DNA Polymerase, AmpErase UNG (prevents re-amplification of PCR products containing deoxyuridine), dNTPs with dUTP, and optimized buffer components. Table C1 lists the probe and primer sequences for regions tested.

2.4.5 Plate Preparation

Fifteen μL of master mix was added to the wells of a 384-well optical PCR plate (ABI). Then, 5 μL of sample or control DNA (1ng/μL), or H_2O (no template control [NTC]) was added (2 wells per sample, 3 wells per control and 3 wells per NTC for each probe). The plate was sealed with a transparent film and centrifuged for 2 minutes at 805 x g.

2.4.6 Assay Run

Using the Applied Biosystems Prism 7900HT Sequence Detection System (SDS), the thermal profile (shown in Table 2.1) was used for the RTqPCR assays. Each well threshold cycle (C_t) was calculated using the ABI software SDS 2.2, and exported into Microsoft Excel for analysis. The C_t is equivalent to the PCR cycle at which the sample first begins to show fluorescence from the reporter dye above background levels.

TABLE 2.1. THERMAL PROFILE FOR RTQPCR ASSAYS. Primer/probe sets were designed based on the Stage 3 annealing temperature of 60°C.

	Stage 1	Stage 2	Stage 3 (40 repeats)	
Temperature (°C)	50	95	95	60
Time (min)	2:00	10:00	0:15	1:00

2.4.7 Analysis

Analysis of the RTqPCR data was completed using Microsoft Excel. The reference male C_t values were averaged for each probe. The values calculated for the 4 unique probes and the

G6PD probe were subtracted from the *PGT* probe value for each sample. Then, each of these values was subtracted from the male reference value for each probe and converted to copy number (Table C2). This is a standard assay design for gene dosage analysis using RTqPCR, and a standard method of handling RTqPCR data (McCarroll et al., 2006; Wilson et al., 2006a; Wilson et al., 2006b). These subtractions set the C_t value for each sample relative to the C_t value in a male reference sample. For example, the ratio for *G6PD* will be twice as high in females as in males because females have two copies. However, other loci with 2 copies in males and females will show a ratio of one relative to the male reference DNA. The final copy number values were averaged for each sample.

2.5 STATISTICAL ANALYSIS

2.5.1 Reference Hybridizations

The coefficient of variation (V) comparison was calculated for the reference hybridizations using the Z test (Table D1) (Zar, 1999).

3.0 RESULTS

3.1 HUMARRAY 2.0

3.1.1 Qualification of Digested DNA

Samples with at least 1 μg of DNA were digested with *MseI* enzyme. Figure 3.1 is a picture of a representative agarose gel. The digested samples show the same smear pattern as the digested male reference DNA.

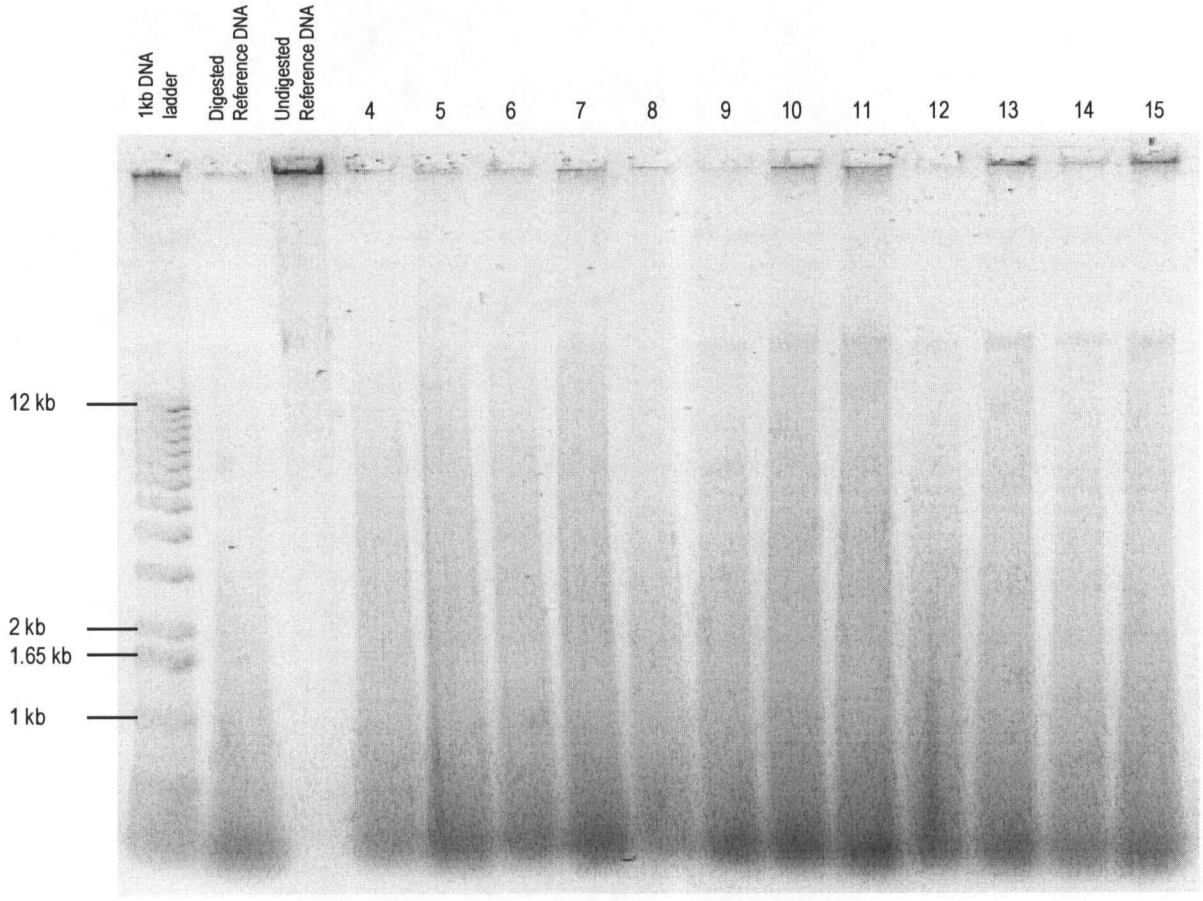

FIGURE 3.1. AGAROSE GEL OF DIGESTED DNA. The digested DNA was from three trios (lanes 4-11) and reference male DNA (lanes 13-15).

3.1.2 Sample Hybridizations

A total of 20 trios (proband, mother and father; 60 arrays) were assessed for DNA copy number changes using low density aCGH and the labeling procedure described in Figure 2.1.

3.1.3 Aberrant Clones

For each of the 63 arrays completed, Query B1 (Appendix B) was used to find individual aberrant loci from the MySQL data table (Table B1). Only loci that were aberrant in 2 of 3 arrays per trio were considered candidate aberrations. This is because a true aberration should be detected twice when an individual is compared to each of the other two trio members. Using these criteria I found 7 aberrant loci in 8 trios, and 2 aberrant loci from chromosome 13 in one trio (Table 3.1). These aberrations were found based on the following cutoffs in Query B1: the locus $\log_2$ratio was 3 or more standard deviations away from the mean $\log_2$ratio of the array (away > 3), the standard deviation of the locus triplicate was 0.2 or less (stdev < 0.2), and the intensity of the triplicate was greater than 10 units on a 16 bit scale (int > 10). I found 11 copy number changes in the seven loci, including 2 amplifications and 9 deletions. The two loci from chromosome 13 were part of a multi-clone aberration cluster. This will be examined further in the next section. For each locus, the inter-array variability was calculated (Query B2). The standard deviation of the 63 independent $\log_2$ratio values (20 trios and 3 reference hybridizations) for each aberrant locus was used to determine which loci had the least noisy between-array data. The intra-array variability was also calculated (Query B3) to determine which arrays had the least noisy data. A summary of the $\log_2$ratios for each aberrant locus, as well as the inter-array and intra-array variabilities are shown in TABLE 3.2.

TABLE 3.1. ABERRANT LOCI IDENTIFIED USING TRIO ARRAY DATA AND QUERY B1. Loci were identified as aberrant if they were present in 2 out of 3 arrays per trio.

Locus	Band[a]	Genes[b]	Trio	Putative Aberration
RP11-122N11	8p23.1	*SPAG11*	201	Father amplification
RP11-82D16	1p36.33	*PRKCZ, SKI*	231	Proband amplification
RP11-1P3	21q22.2	*PCP4, DSCAM*	257	Mother deletion
RP11-88C10	19p11.32	*CLUL1, CETN1, ENOSF1, YES1, TYMS*	257	Mother deletion
RP11-43B19	6q25.3-26	*LPA*	257	Father deletion
RP11-43B19	6q25.3-26	*LPA*	297	Proband deletion
RP11-43B19	6q25.3-26	*LPA*	309	Mother deletion
RP11-43B19	6q25.3-26	*LPA*	341	Mother deletion
RP11-95J4	9q32	*ZFP37, SLC21A2*	261	Mother deletion
RP11-88L18	5p15.1	none	261	Proband deletion
RP11-234M13	13q21.2	none	456	Proband
RP11-205J24	13q21.2	none	456	Proband

[a]From UCSC Genome Browser
[b]From RefSeq track on UCSC Genome Browser

TABLE 3.2. LOG2RATIOS FOR EACH OF THE 16 ABERRANT LOCI DETECTED BY ACGH.

Locus	Trio	Log_2_1[a]	Intra_$log_2$1[b]	Log_2_2[a]	Intra_$log_2$2[b]	Log_2_3[a]	Intra_$log_2$3[b]	Inter_var[c]
RP11-122N11	201	-0.4223	0.137	-0.0428	0.197	0.2955	0.144	0.2
RP11-82D16	231	-0.0165	0.233	0.4661	0.181	-0.1982	0.099	0.206
RP11-43B19	257	0.5391	0.144	0.1816	0.109	-0.5127	0.151	0.263
RP11-1P3	257	-0.3613	0.144	0.389	0.109	0.1553	0.151	0.175
RP11-88C10	257	-0.3693	0.144	0.3951	0.109	0.1101	0.151	0.191
RP11-95J4	261	-0.2596	0.14	0.4249	0.153	-0.0716	0.192	0.152
RP11-88L18	261	0.2456	0.14	0.1744	0.153	-0.4656	0.192	0.251
RP11-43B19	297	0.2732	0.127	0.1406	0.201	-0.4919	0.164	0.263
RP11-43B19	309	-0.586	0.221	0.3255	0.125	0.1172	0.112	0.263
RP11-43B19	341	-0.4839	0.148	0.9738	0.148	-0.3144	0.159	0.263
RP11-234M13	456	-0.2936	0.134	-0.2889	0.181	-0.2792	0.111	0.136
RP11-205J24	456	-0.2836	0.134	-0.2074	0.181	-0.3326	0.111	0.222

[a]log_2_ratio from hybridization, [b]intra-array variability for hybridization, [c]inter-array variability for locus

3.2 CHROMOSOME 13 ABERRATION

3.2.1 HumArray 2.0

The chromosome 13 multi-clone aberration was visually obvious in MIA graphical output and identified using Query B1. Trio 456 showed a proband-specific (spontaneous) aberration of chromosome 13 (Figure 3.2). The aberration was between and including clones RP11-205J24 and RP11-234M13 in a region of approximately 20 Mb. Using MIA breakpoint analysis, all of chromosome 13 was identified as aberrant.

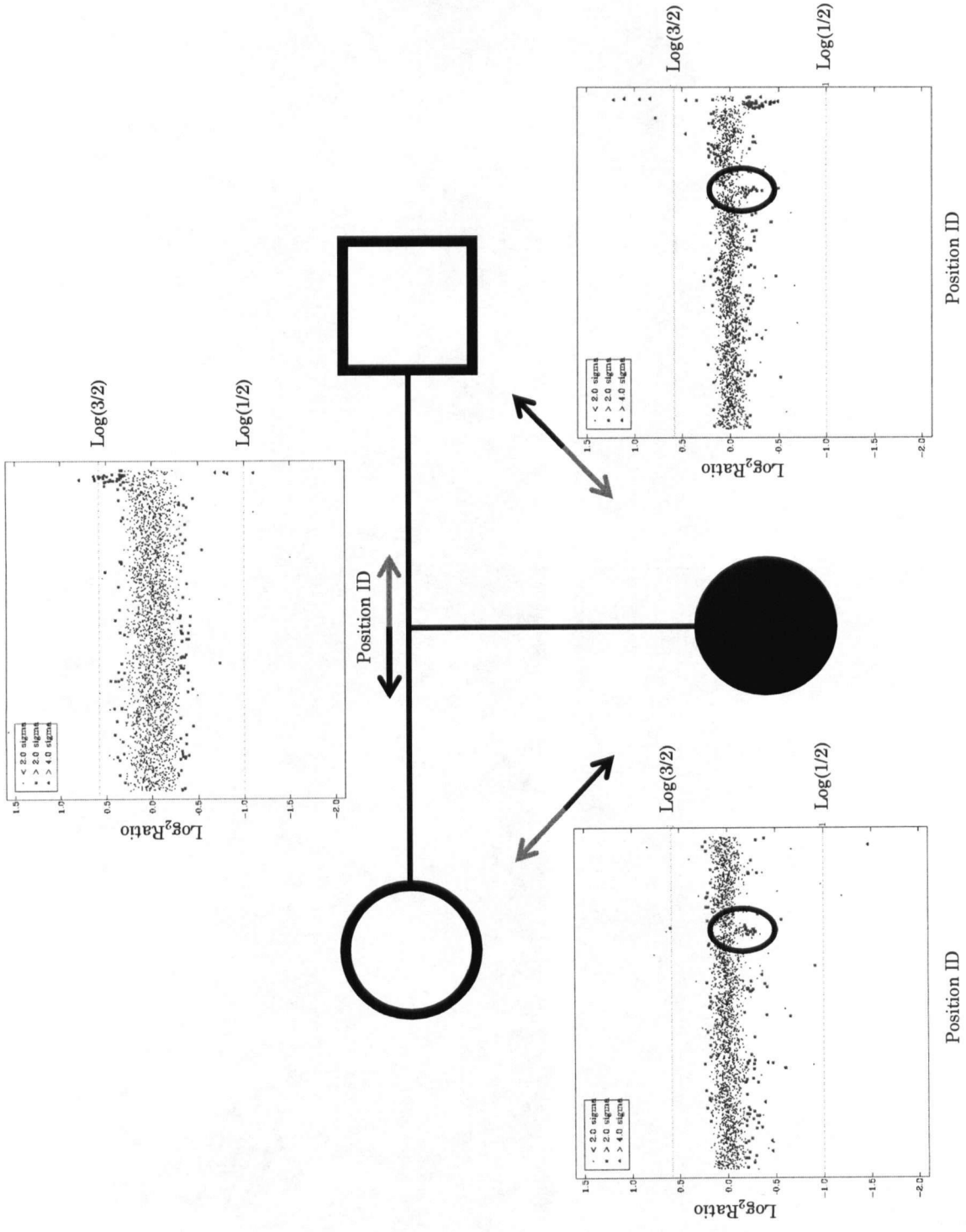

FIGURE 3.2. PEDIGREE FOR TRIO 456. The aberration on chromosome 13 is circled in black on two graphs. Females are represented by circles, males by squares. Affected individuals are shaded. Cy-3 labeled DNA is indicated by a grey arrow and Cy-5 labeled DNA indicated by a black arrow. Log_2ratio is log_2(Cy5 intensity/Cy3 intensity).

3.2.2 SMRT Array Validation

Validation of the chromosome 13 amplification was attempted by hybridizing the proband 456 sample against male reference DNA (Novagen) on the high density sub-megabase resolution tiling (SMRT) array. The array results did not validate a chromosome 13 aberration in the proband suggesting the presence of an artifact on the HumArray 2.0 (Figure 3.3) which will be addressed in the next section. However, the results do show putative aberrations on other chromosomes that I missed using the HumArray 2.0 potentially due to the lack of clone coverage for these regions. Using Query B4 with Table B2, I identified 3 potential multi-clone aberrations in the proband of trio 456. The first aberration is located on chromosome 8, between 86.2 Mb to 86.5 Mb where seven clones are implicated in this aberration. The second aberration is on chromosome 11 between 61.0 Mb and 61.3 Mb. Within the chromosome 8 aberration, there are six genes. The final aberration is on chromosome 16 between 32.5 Mb and 35.0 Mb where six clones are aberrant. None of the genes in the chromosome 8 region are involved in brain development, but the *sodium channel associated protein* (*LRRCC1*) may be involved in brain function (Table 3.3). The chromosome 11 aberration includes *synaptotagmin VII* (*SYT7*), which is involved in synaptic exocytosis and neurotransmitter release (Rao et al., 2004). The chromosome 16 aberration contained only predicted and provisional genes when viewed with the RefSeq track on the UCSC Genome Browser (Karolchik et al., 2003). I also identified a single clone on chromosome 11 with a $\log_2$ratio of -2.9. The genes (*SLC1A2*, *CD44*) located within this deletion of chromosome 11 are listed in Table 3.3.

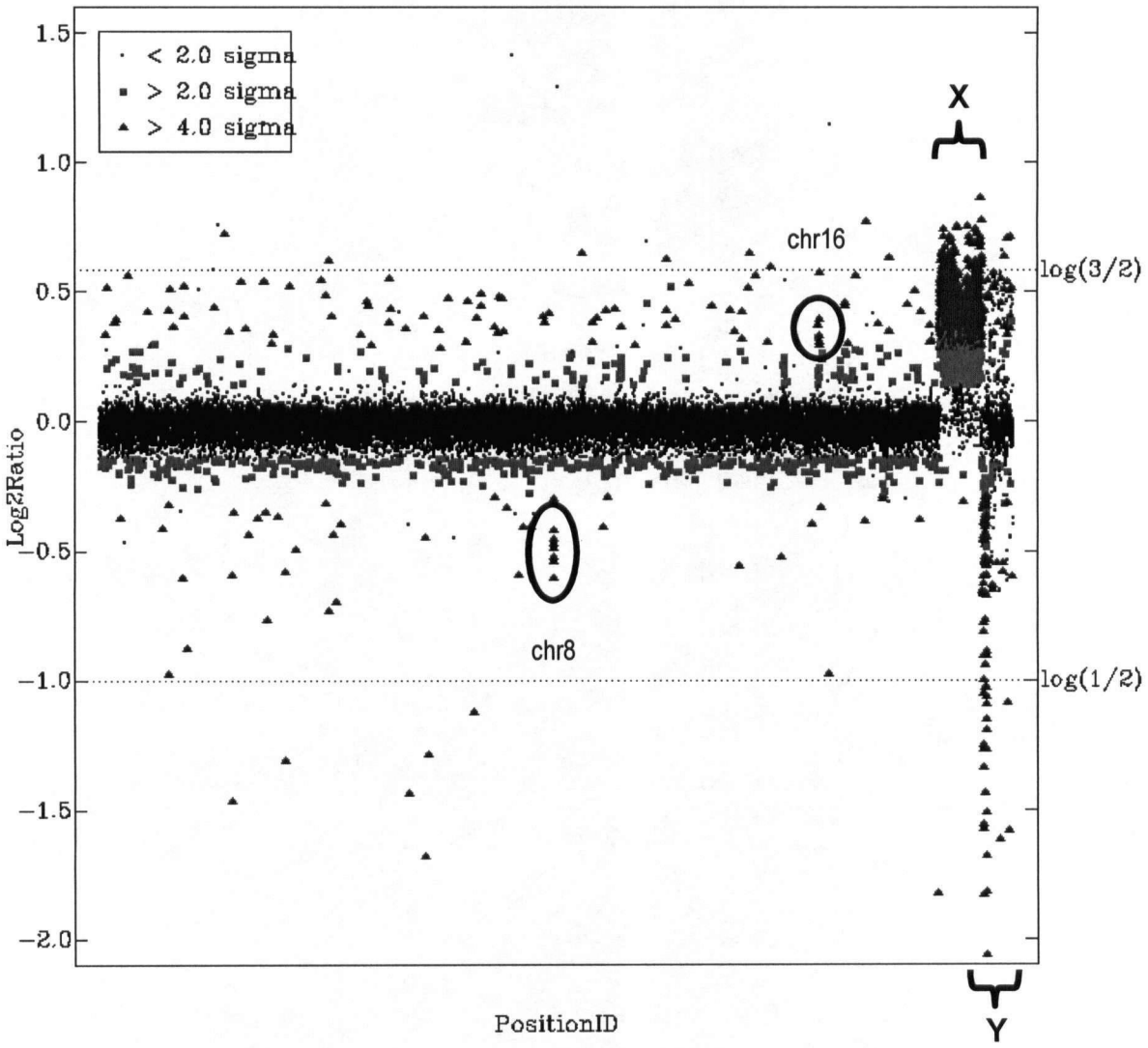

FIGURE 3.3. MIA GRAPH OF SMRT ARRAY DATA. Most clones (24336/24550 = 99.2% of autosomal clones) show no copy number difference between the trio 456 proband and male reference DNA. The **X** and **Y** chromosome clones are on the far right of the graph and verify that known copy number differences are detectable on this array platform.

TABLE 3.3. GENES LOCATED WITHIN SMRT ARRAY ABERRATIONS FOR PROBAND FROM TRIO 456.

Gene	Full Name	Chromosome	Region	RefSeq ID
LRRCC1	*sodium channel associated protein 2*	8	86206716-86245072	NM_033402.2
E2F5	*E2F transcription factor 5*	8	86276877-86314002	NM_001951.2
CA13	*carbonic anhydrase XIII*	8	86345259-86383554	NM_198584.1
CA1	*carbonic anhydrase I*	8	86427709-86477594	NM_001738.1
CA3	*carbonic anhydrase III*	8	86537710-86548526	NM_005181.2
CA2	*carbonic anhydrase II*	8	86563498-86580973	NM_000067.1
SLC1A2	*solute carrier family 1 member 2*	11	35229329-35397372	NM_004171.2
CD44	*CD44 antigen isoform 5 precursor*	11	35116993-35210524	NM_001001392.1
SYT7	*synaptotagmin VII*	11	61039362-61104874	NM_004200.2
FEN1	*flap structure-specific endonuclease 1*	11	61316726-61321284	NM_004111.1
FADS1	*fatty acid desaturase 1*	11	61323675-61340886	NM_013402.3
FADS2	*fatty acid desaturase 2*	11	61352289-61391401	NM_004265.2

3.2.3 Reference Hybridization

Because the chromosome 13 aberration was not detected by the higher resolution SMRT array, it is a putative artifact. To further explore this probability, three self hybridizations (male reference DNA versus itself) were performed on the HumArray 2.0. One of the three reference hybridizations again showed a multi-clone chromosome 13 aberration (Figure 3.4). The aberration spans clones RP11-234M13 through RP11-94F7, which matches the trio 456 proband aberration found with the HumArray 2.0 and confirms the presence of a chromosome 13 artifact. The coefficient of variation (V; also called relative variability) was calculated for each of the reference hybridizations (Table D1). A statistical comparison between the three coefficients of variation showed that V for the reference hybridization with the chromosome 13 aberration is significantly different from the other two. This suggests that the reproducibility of the array results is affected by the artifact. The reason why chromosome 13 hybridization results are

unreliable is not clear. Due to the sporadic nature of this artifact, loci on chromosome 13 have been excluded from subsequent analysis.

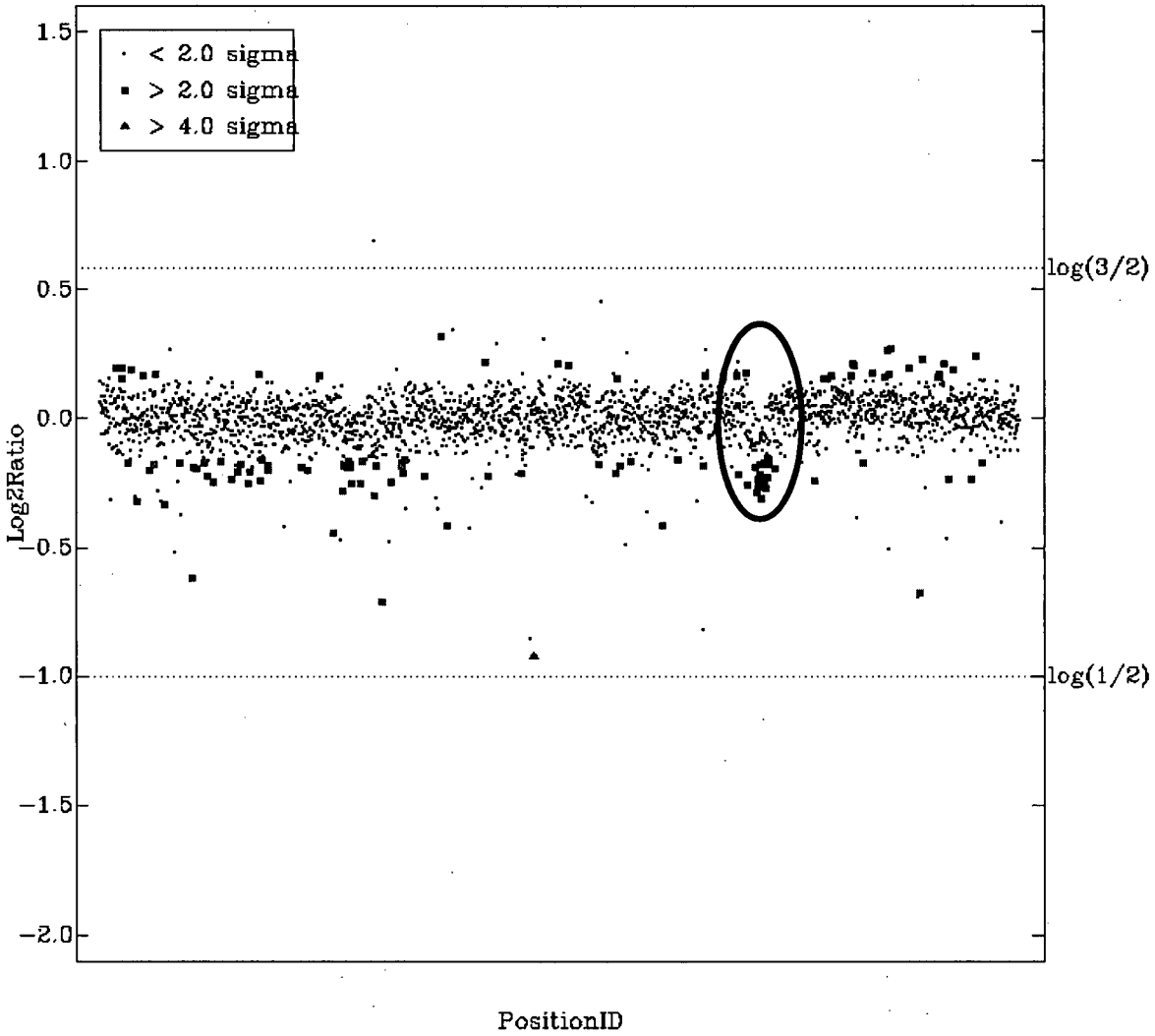

FIGURE 3.4. MIA GRAPH OF REFERENCE HYBRIDIZATION. This graph is the output of a male vs. male reference hybridization. The aberrant clones circled in black are from chromosome 13.

3.3 REAL TIME QUANTITATIVE PCR

3.3.1 Clones Chosen for Probe Design

Genomic regions for RTqPCR analysis were chosen from the aCGH-detected aberrations (TABLE 3.2). Specifically, the absolute values for the three $\log_2$ratios of an aberrant locus (one from each of three arrays per trio) were used to determine an overall $\log_2$ratio sum. For example, in Table 3.4 three loci with the absolute values for the three corresponding $\log_2$ratios have been shown (abs_$\log_2$X). The difference between each of the two highest ratios and the lowest ratio (diff_1 and diff_2) were summed and used to determine the $\log_2$ratio "ranking" of the locus. A high ranking takes into account that (1) there should be a large difference between the highest and lowest ratios, and/or (2) the lowest ratio is close to zero. Three loci were chosen for analysis based on the rank and gene content (Table 3.5). Each locus was investigated for gene content and one gene per locus was chosen for primer/probe design. The choice of gene was made based on potential function in the brain. The initial ~500bp region for *protein kinase C zeta* (*PRKCZ*) incorporated the second to last intron (chr1:2,144,215-2,144,770). Research with rats indicated that the *PRKCZ* gene product interacts with a potassium channel in brain via another protein, ZIP1 (Gong et al., 1999). The *lipoprotein A* (*LPA*) region encompassed all of the second exon and part of the second intron (chr6:161,037,320-161,037,872). A recent study found an increased level of Lp(a) protein in patients with SCZ, major depression and BPD when compared to healthy controls, and suggests that this protein may contribute to cardiovascular risk in psychiatric patients (Emanuele et al., 2006). Finally, the region for *zinc finger protein 37 homolog* (*ZFP37*) was from the last intron (chr9:112,886,269-112,886,819). Investigation of *ZFP37* in mice suggests that this gene encodes for a protein that participates in the structural integrity of neuronal nuclei (Payen et al., 1998).

TABLE 3.4. LOG2RATIO RANKING FOR LOCI.

Locus	abs_$log_2$1	abs_ $log_2$2	abs_ $log_2$3	diff_1	diff_2	Rank
RP11-43B19	0.4839	0.9738	0.3144	0.1695	0.6594	0.8289
RP11-82D16	0.0165	0.4661	0.1982	0.4496	0.1817	0.6313
RP11-95J4	0.2596	0.4249	0.0716	0.188	0.3533	0.5413

abs = absolute value
Rank = sum of diff_1 and diff_2

TABLE 3.5. REGIONS CHOSEN FOR REAL TIME QUANTITATIVE PCR VALIDATION.

Locus	Genes[a]	Full Name[b]	Potential Brain Function
RP11-43B19	***LPA***	*lipoprotein A*	Cardiovascular risk in psychiatric patients
RP11-82D16	***PRKCZ***, *SKI*	*protein kinase C zeta*	Targets potassium channels in brain
RP11-95J4	***ZFP37***, *SLC31A2*	*zinc finger protein 37 homolog*	Structural protein in rat neuronal nuclei

[a]DNA from the genes in **boldface** text was used for RTqPCR primer/probe design.
[b]Full name of chosen gene.

3.3.2 Determination of Copy Number Change

I considered two methods for determining copy number changes in the RTqPCR experiments: the normal range method (NRM), and the standard deviation method (SDM). Using the NRM the range of ratios acquired for fathers using the X chromosome *G6PD* probe (0.73-1.33) was used as the normal range of variation in my assays for a gene with one copy in the genome. In the SDM, the standard deviation of all the copy number ratios for each gene was used to determine whether an individual sample for a particular gene is aberrant. For example, the average copy number ratio (1.093) and the standard deviation (0.167) for the *PRKCZ* locus would be used to determine whether an individual sample copy number ratio was aberrant. The copy number of the individual sample would need to be the mean ± 2SD to be identified as aberrant. However, the genes tested by RTqPCR showed a high degree of variation around the normal copy number ratio of 1. These could be real aberrations, but the SDM would not identify them because the real differences had inflated the SD. For this reason I used the normal range method

Based on the NRM, ratios for the other genes were categorized as amplifications, deletions or normal if they fell within the ranges of >1.33, <0.73 or 0.73-1.33, respectively. Figure 3.5 is a graph of the *G6PD* ratios for each father showing the ranges. Male proband ratios for *G6PD* were all contained within the range of 0.73-1.33. Using this range, a number of aberrations were identified in the trio samples (Table 3.6).

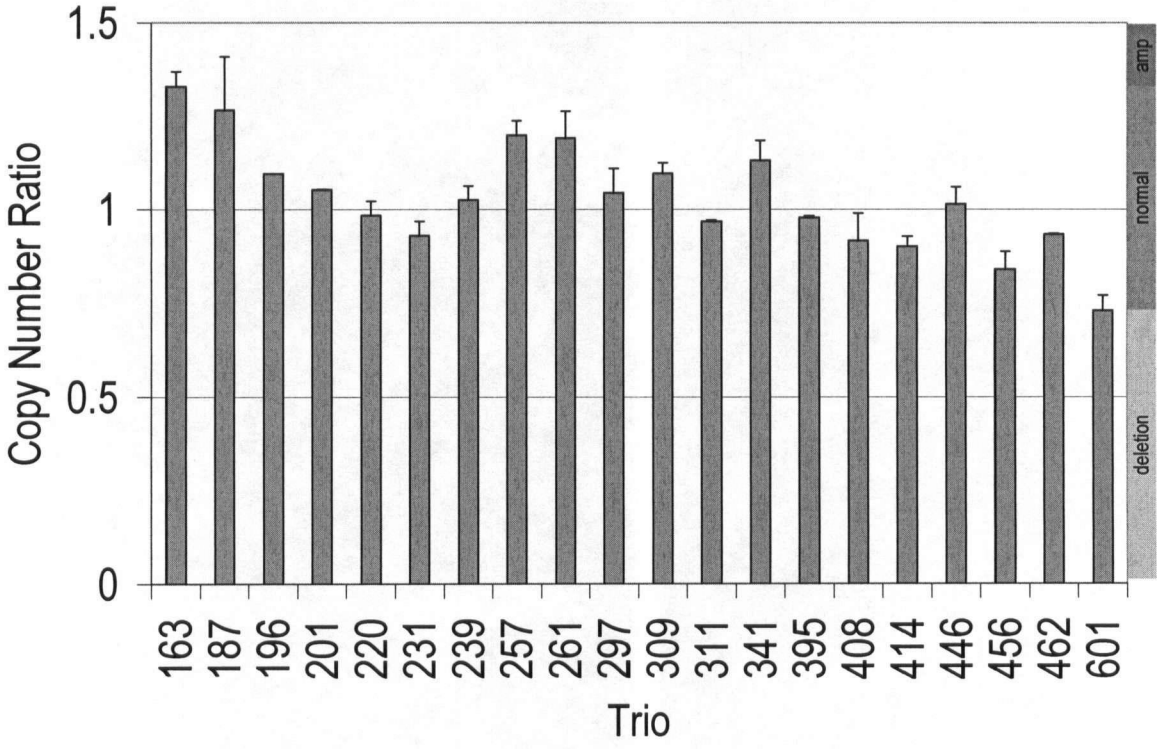

FIGURE 3.5. *G6PD* RATIOS FOR FATHER SAMPLES. Each bar represents the *G6PD* copy number for a father sample. Error bars represent the standard deviation of the two replicates done for each sample.

TABLE 3.6. TYPES OF ABERRATIONS IDENTIFIED IN EACH GENE BY RTQPCR. Numbers include all samples tested.

Aberration	*PRKCZ*	*LPA*	*ZFP37*
Amplifications	4	9	5
Deletions	0	3	0

3.3.3 Aberrant Copy Number

All samples were typed correctly for gender with the *G6PD* probe. These results are consistent with the know gender of the subject (Figure 3.6). One sample, typed as female on the sample key, presented as male (proband trio 261, ⬇). This discrepancy was identified using both aCGH and RTqPCR and is likely due to a clerical error.

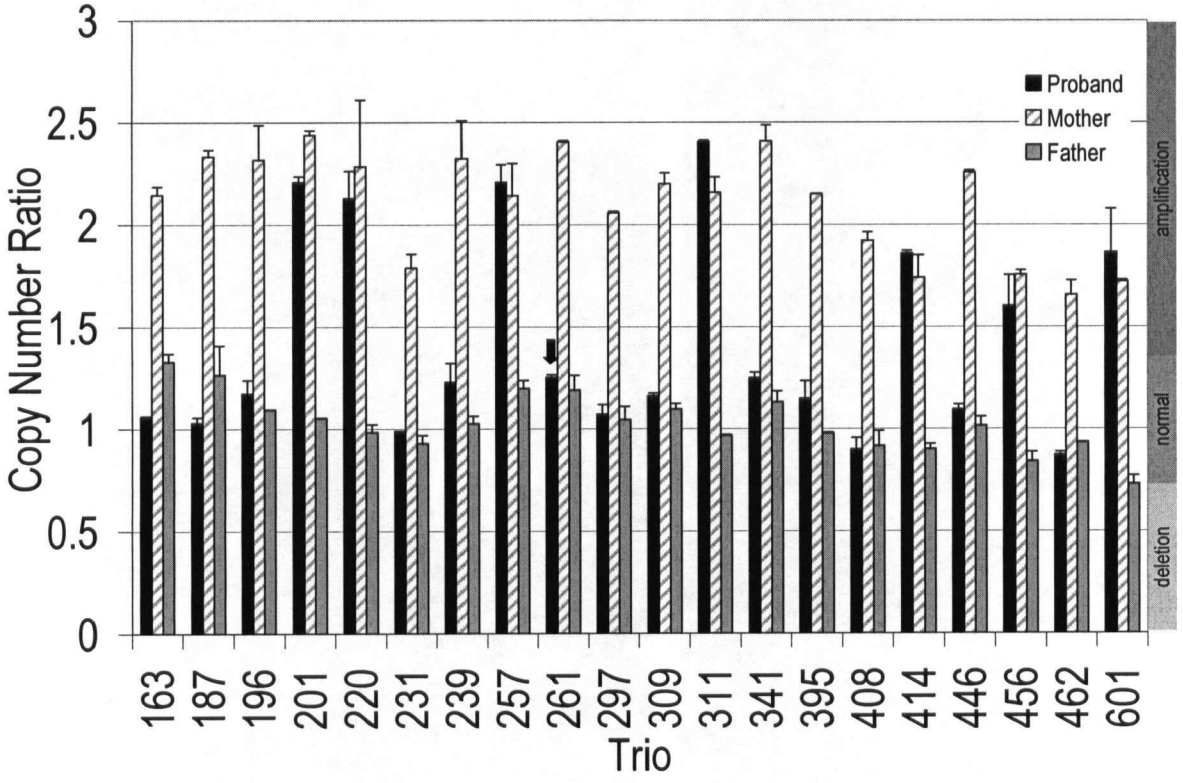

FIGURE 3.6. *G6PD* COPY NUMBER RATIOS. Error bars represent the standard deviation of the two replicates done for each sample.

By RTqPCR, three of 20 trios show a copy number increase at the *PRKCZ* locus (Figure 3.7). Trio 163 has an affected mother, while trio 257 has an affected father. Trio 239 has an affected proband and mother. The amplification "allele" from the mother was inherited by the proband. The trio (231) that showed the aCGH aberration did not show an RTqPCR-detected aberration.

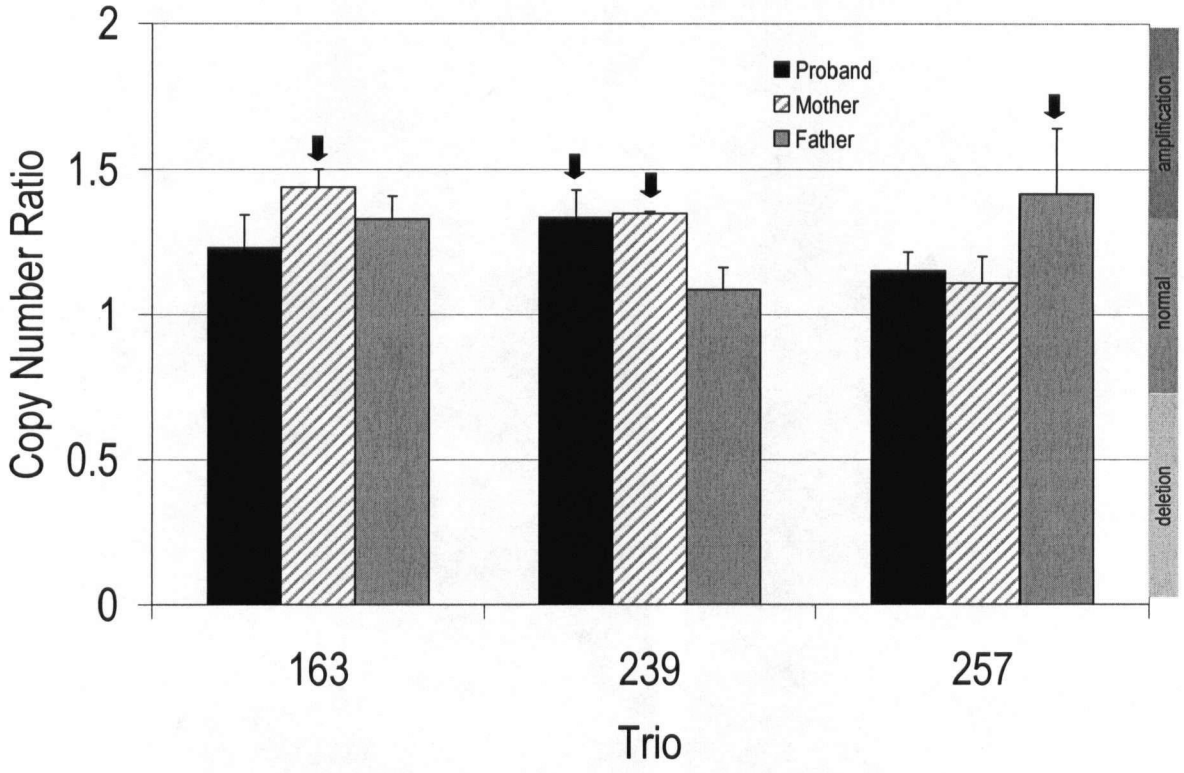

FIGURE 3.7. *PRKCZ* COPY NUMBER RATIOS. Samples with copy number aberrations for *PRKCZ* are labeled with ⬇. Error bars represent the standard deviation of the two replicates done for each sample.

Nine of 20 trios were found by RTqPCR to have copy number changes in the *LPA* locus (Figure 3.8). Of the affected probands, 2 had copy number decreases (297, 414) and 3 showed amplifications (257, 309, 395). The 2 mother samples (257, 395) showed amplifications. One father sample (601) showed a deletion while four father samples (201, 261, 309, 341) had amplifications. From these data, it was evident that two probands (257, 395) inherited amplifications from their mother, one proband (309) inherited an amplification from his father and two probands (297, 414) may have spontaneous deletions.

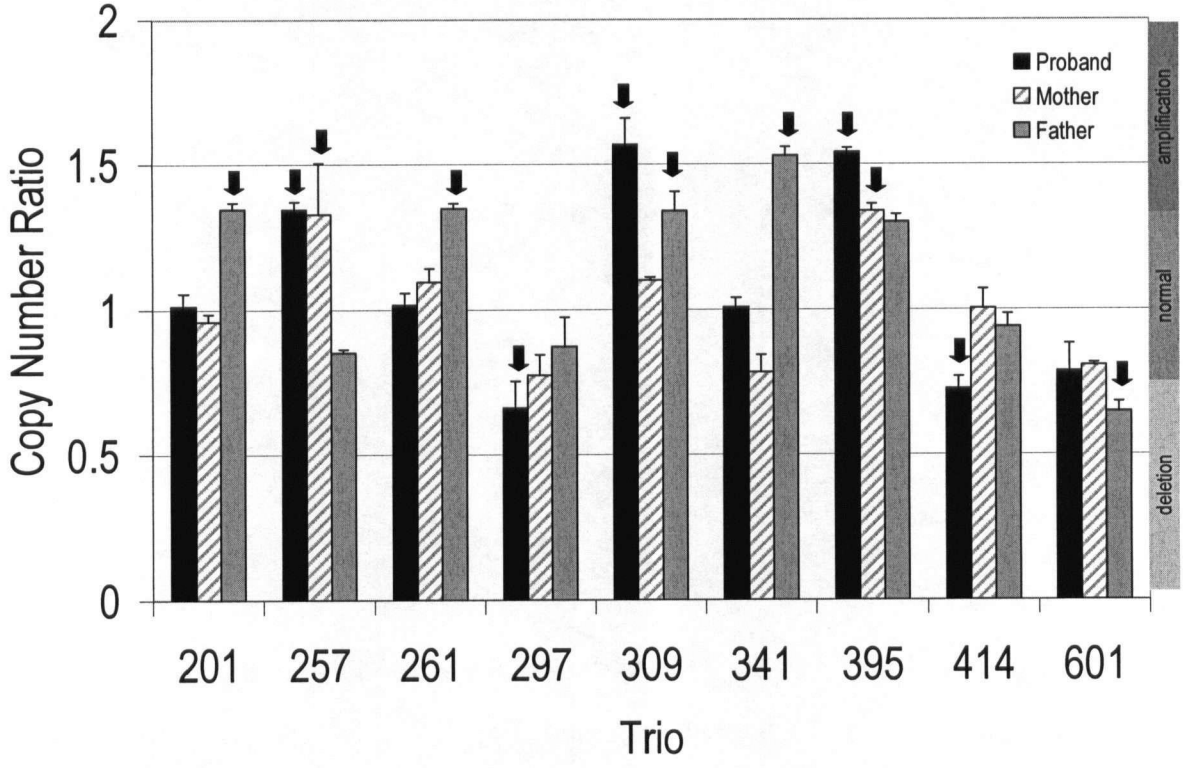

FIGURE 3.8. *LPA* COPY NUMBER. Samples with copy number aberrations are labeled with ⬇. Error bars represent the standard deviation of the two or three (trios 257, 297, 309 and 341) replicates done for each sample. The copy number ratios for trios 257, 297, 309 and 341 are average values from the *LPA* replication experiments (see FIGURE 3.11).

Five of 20 trios showed copy number changes in *ZFP37* (Figure 3.9). All aberrations were amplifications. One proband (311), one mother (201) and three fathers (257, 261, 341) have an amplification. According to these data, the proband amplification was a spontaneous change.

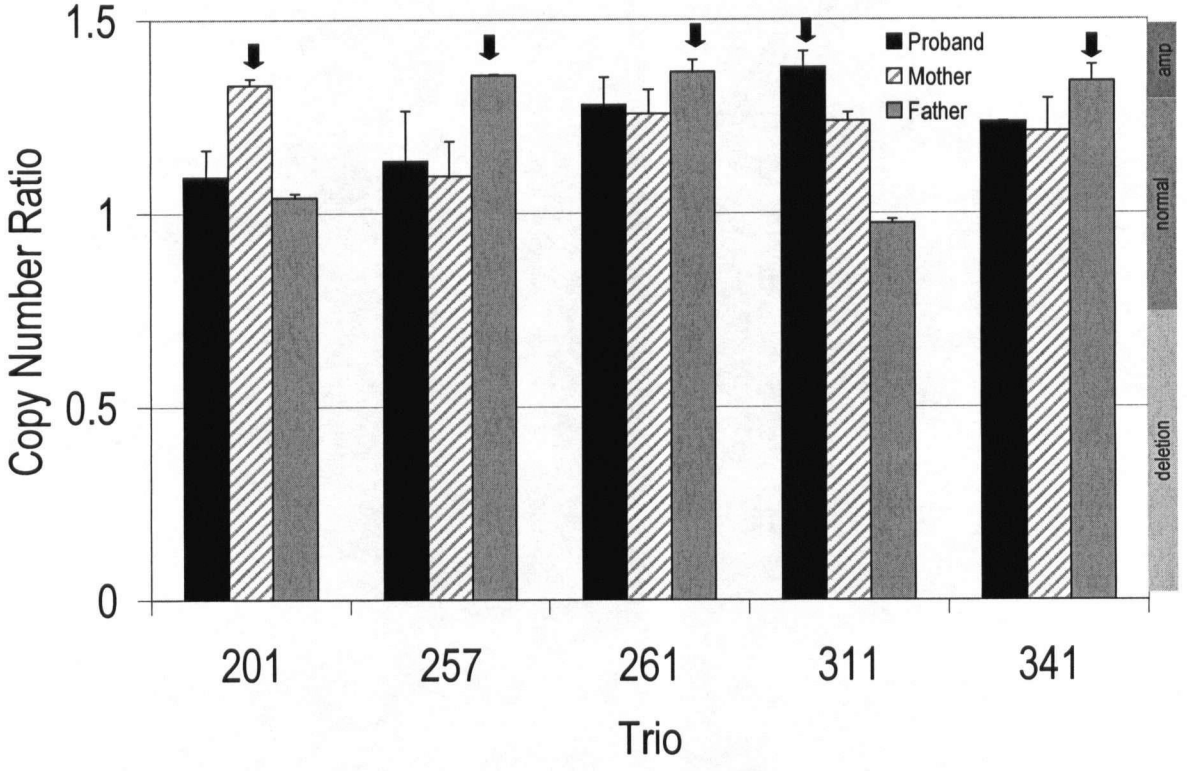

FIGURE 3.9. *ZFP37* COPY NUMBER RATIOS. Samples with copy number aberrations for *ZFP37* are labeled with ⬇. Error bars represent the standard deviation of the two replicates done for each sample.

A probe for the *synaptotagmin VII* (*SYT7*) gene (designed by G. Wilson as described in Methods), was used with trio 456 (primer and probe sequences are in Table C1). This trio showed a possible amplification on chromosome 11 in the proband sample using high density aCGH. However, the *SYT7* RTqPCR data showed no aberration in any of the trio 456 samples (Figure 3.10).

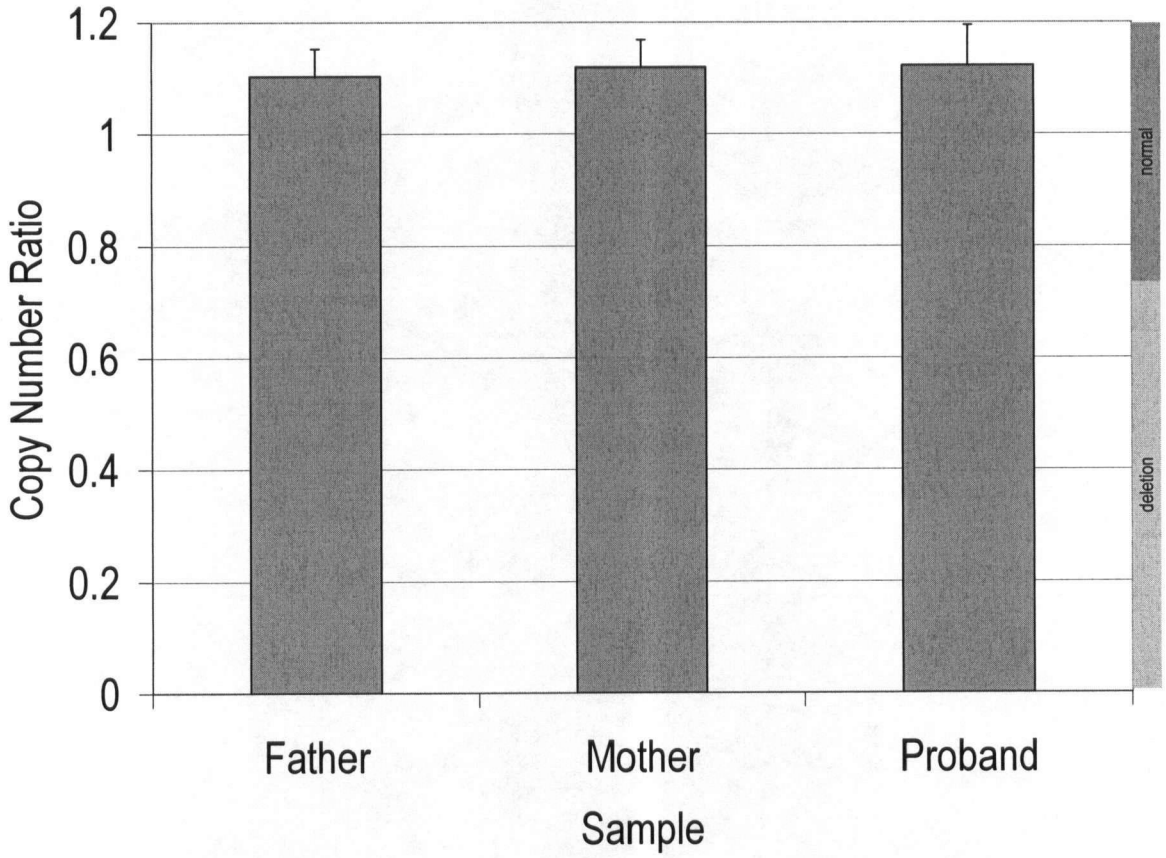

FIGURE 3.10. *SYT7* COPY NUMBER RATIO IN TRIO 456. Error bars represent the standard deviation of the three replicates done for each sample.

3.3.4 Replication

The four trios that showed an aCGH aberration in clone RP11-43B19, the *LPA* locus, were tested a second time by RTqPCR for replication purposes (i.e. to confirm that the RTqPCR finding validated the aCGH results). The replication results show strong concordance with the initial RTqPCR results for *LPA*. Figure 3.11 shows a comparison between the first and second experiment for the four trios with the *LPA* probe. Only four samples did not show the same copy number between both experiments (257 mother, 257 proband, 297 mother and 309 father). The average copy number ratio between the two experiments was used for analysis of *LPA*.

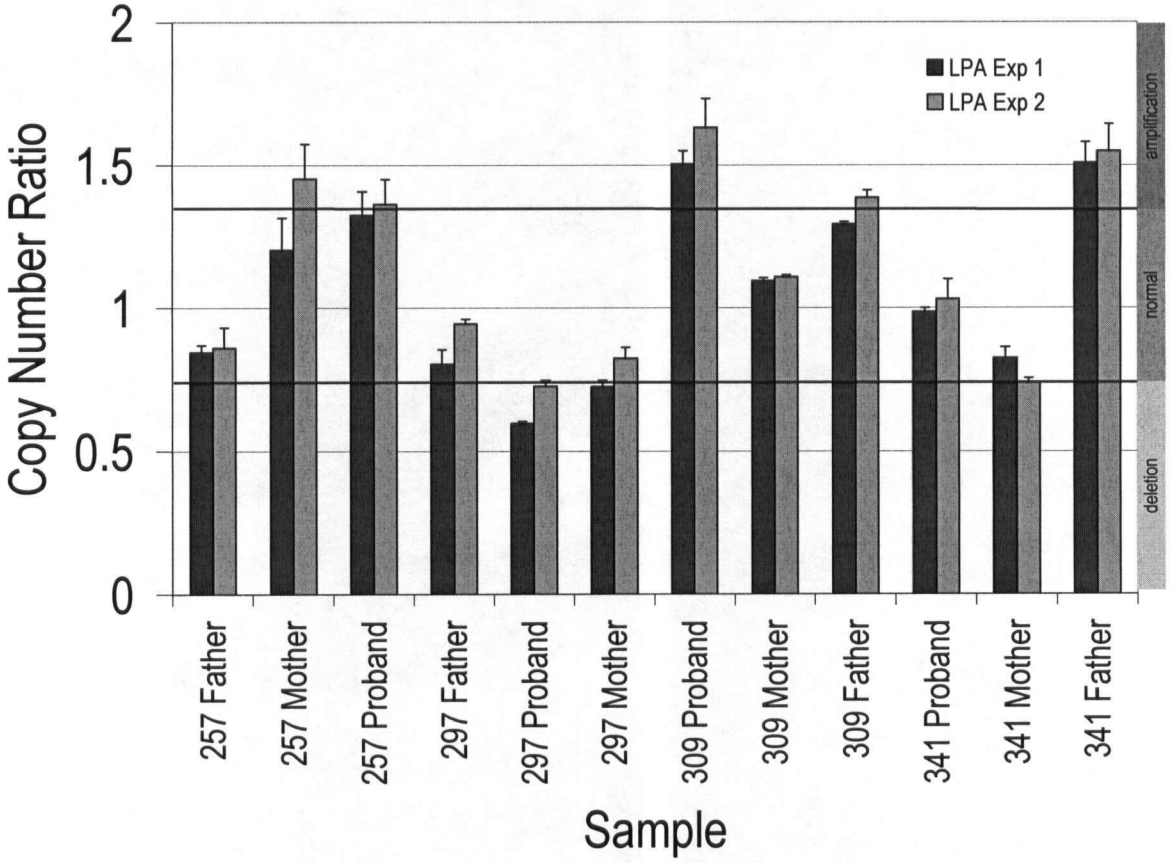

FIGURE 3.11. REPLICATION OF *LPA* COPY NUMBER RATIO IN FOUR TRIOS. Error bars represent the standard deviation of the two (LPA Exp 1) or three (LPA Exp 2) replicates done for each sample. The horizontal black lines indicate the thresholds for aberrant copy number. Ratios between the lines are normal copy number.

3.4 COMPARISON OF ACGH AND RTQPCR

The evaluation of aCGH and RTqPCR data must take into account the way in which each method measures copy number. The ratio from aCGH represents a competition between two differentially labeled DNAs such that in trios, the aCGH-defined copy number of each individual is relative to the other people in the family and the actual or absolute copy number may be ambiguous. In the absence of other information, copy number aberrations can be inferred by parsimony. In contrast to aCGH, the values produced from RTqPCR analysis are a quantitative measurement of the copy number in each sample relative to a reference DNA. The RTqPCR-defined copy number is a ratio when compared to a reference DNA, but the values within a trio are absolute with respect to each other. In some cases, results from RTqPCR may confirm the most parsimonious interpretation of an aCGH result, but in other cases the RTqPCR may confirm the less parsimonious interpretation. The following is an interpretation of aCGH copy number at each locus tested by RTqPCR.

PRKCZ (clone RP11-82D16)

No aberrations were found by RTqPCR in trio 231, so this may be a false positive aCGH finding. Other considerations will be presented in the Discussion section.

LPA (clone RP11-43B19)

Trio 257 shows the father/proband hybridization with reduced signal and the mother/proband hybridization with increased signal (Figure 3.12). The putative aberrations(s) could be a deletion in the father or amplifications in the mother and proband. The most parsimonious explanation based on the aCGH data alone is a deletion in the father. However, the RTqPCR data indicates

that there are two separate amplifications, one in the mother and one in the proband, rather than a deletion in the father (Figure 3.12).

For trio 297, aCGH showed a decreased ratio in the proband/father hybridization and an increased ratio in the proband/mother hybridization. This result could indicate a proband deletion or amplifications in the mother and father samples. However, the RTqPCR-defined copy number was consistent with a single copy of this locus in the proband. Thus, taken together, the aCGH and RTqPCR data are consistent with a proband deletion (Figure 3.13).

In trio 309, aCGH detected a decrease in the mother compared to the proband, and a decrease in the mother when compared to the father. However, this could also be interpreted as amplifications in the father and proband. Subsequently, the RTqPCR results show amplifications in the proband and father samples (Figure 3.14).

The aCGH results for trio 341 showed a decrease in the mother compared to the proband and a decrease in the mother as compared to the father. As in trio 309, these results could be interpreted as amplifications in the proband and father. The RTqPCR results showed an amplification in the father sample. Given that the proband sample does not contain an RTqPCR-detected amplification, the results for this trio are inconclusive (Figure 3.15).

ZFP37 (clone RP11-95J4)

Using aCGH, the trio 261 mother showed a decrease in copy number when compared to the father, and a decrease when compared to the proband. This result could be interpreted as amplifications in the proband and father. However, the RTqPCR results showed amplification in the father sample. Taken together, the aCGH and RTqPCR results are inconclusive (Figure 3.16).

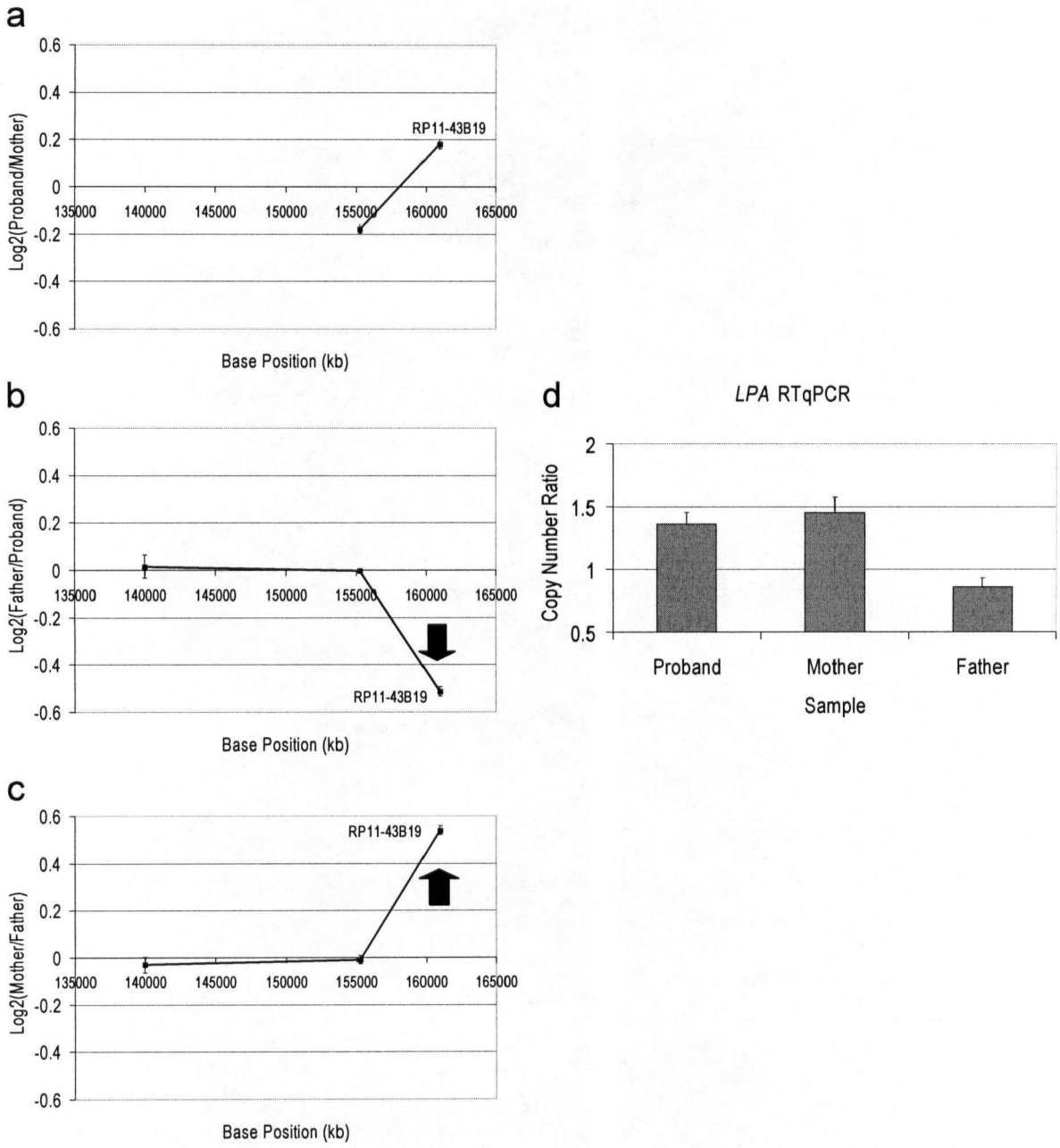

FIGURE 3.12. TRIO 257. The locus (RP11-43B19) which contains *LPA* is labeled on the aCGH graphs. Error bars represent the standard deviation for the triplicate spot (aCGH) or duplicate sample (RTqPCR). The $\log_2$ratios for RP11-82D16 that were identified as aberrant are labeled with arrows. (a) aCGH results from the proband-mother hybridization (there is no data from this array for the left-most clone). (b) aCGH results from the father-proband hybridization. (c) aCGH results from the mother-father hybridization. (d) RTqPCR results for trio 257 with the probe for *LPA*. The aCGH and RTqPCR results are consistent with amplifications in the proband and mother samples.

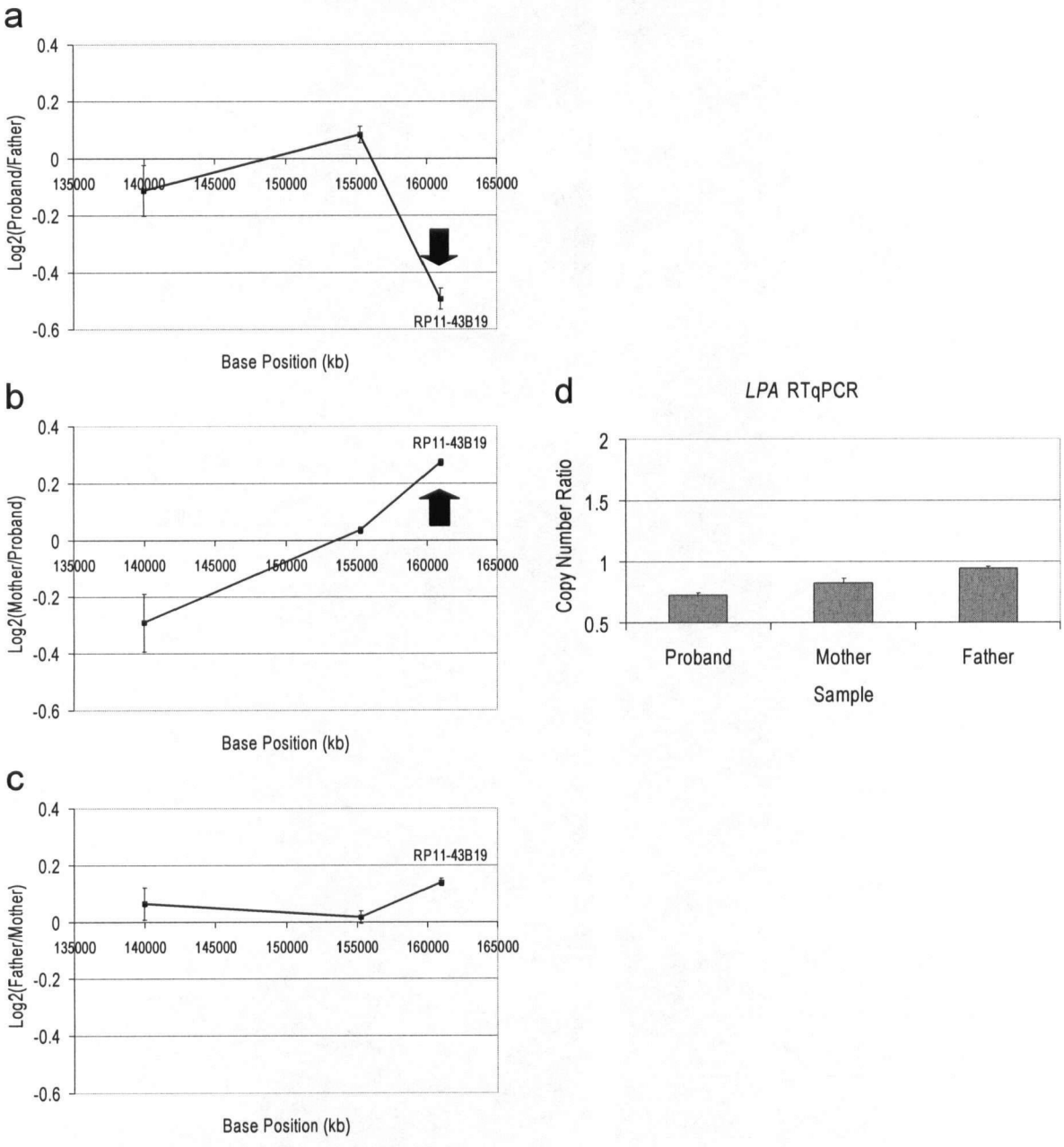

FIGURE 3.13. TRIO 297. The locus (RP11-43B19) which contains *LPA* is labeled on the aCGH graphs. Error bars represent the standard deviation for the triplicate spot (aCGH) or duplicate sample (RTqPCR). The $\log_2$ratios for RP11-82D16 that were identified as aberrant are labeled with arrows. (a) aCGH results from the proband-father hybridization. (b) aCGH results from the mother-proband hybridization. (c) aCGH results from the father-mother hybridization. (d) RTqPCR results for trio 297 with the probe for *LPA*. The aCGH and RTqPCR results are consistent with a deletion in the proband.

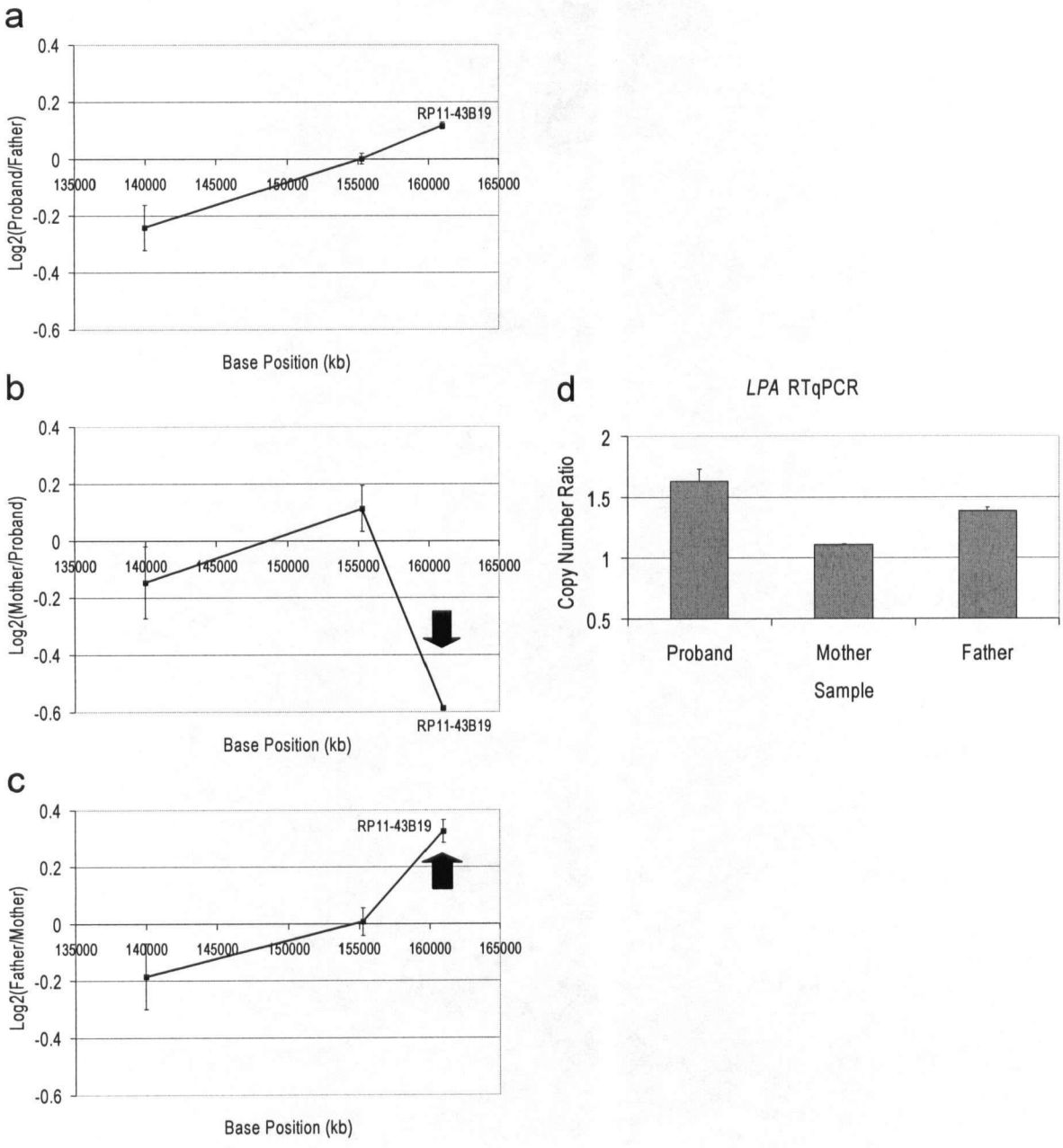

FIGURE 3.14. TRIO 309. The locus (RP11-43B19) which contains *LPA* is labeled on the aCGH graphs. Error bars represent the standard deviation for the triplicate spot (aCGH) or duplicate sample (RTqPCR). The $\log_2$ratios for RP11-82D16 that were identified as aberrant are labeled with arrows. (a) aCGH results from the proband-father hybridization. (b) aCGH results from the mother-proband hybridization. (c) aCGH results from the father-mother hybridization. (d) RTqPCR results for trio 309 with the probe for *LPA*. The aCGH and RTqPCR results are consistent with amplifications in the father and proband samples.

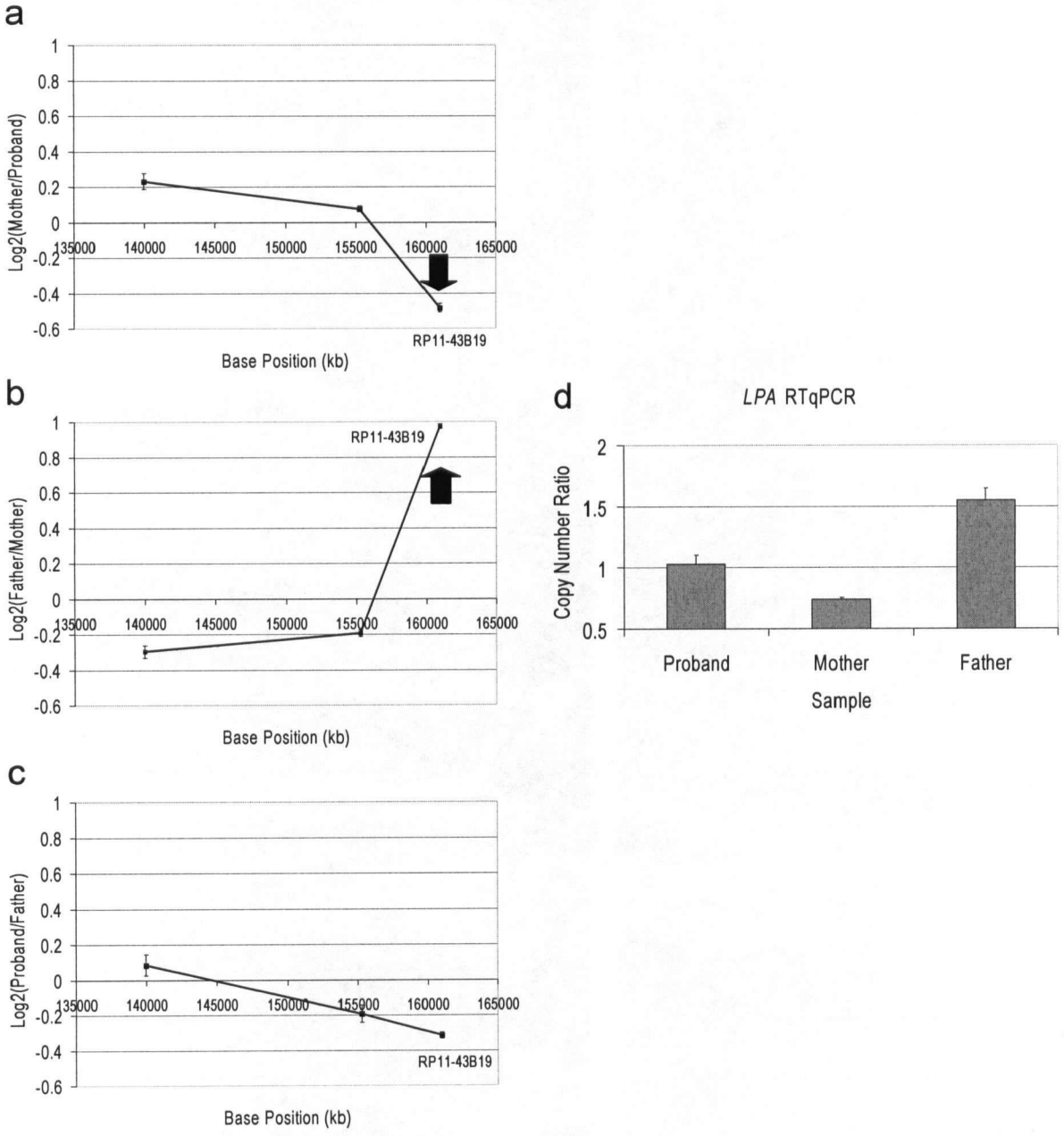

FIGURE 3.15. TRIO 341. The locus (RP11-43B19) which contains *LPA* is labeled on the aCGH graphs. Error bars represent the standard deviation for the triplicate spot (aCGH) or duplicate sample (RTqPCR). The $\log_2$ratios for RP11-82D16 that were identified as aberrant are labeled with arrows. (a) aCGH results from the mother-proband hybridization. (b) aCGH results from the father-mother hybridization. (c) aCGH results from the proband-father hybridization. (d) RTqPCR results for trio 341 with the probe for *LPA*. Taken together, the aCGH and RTqPCR results are inconclusive.

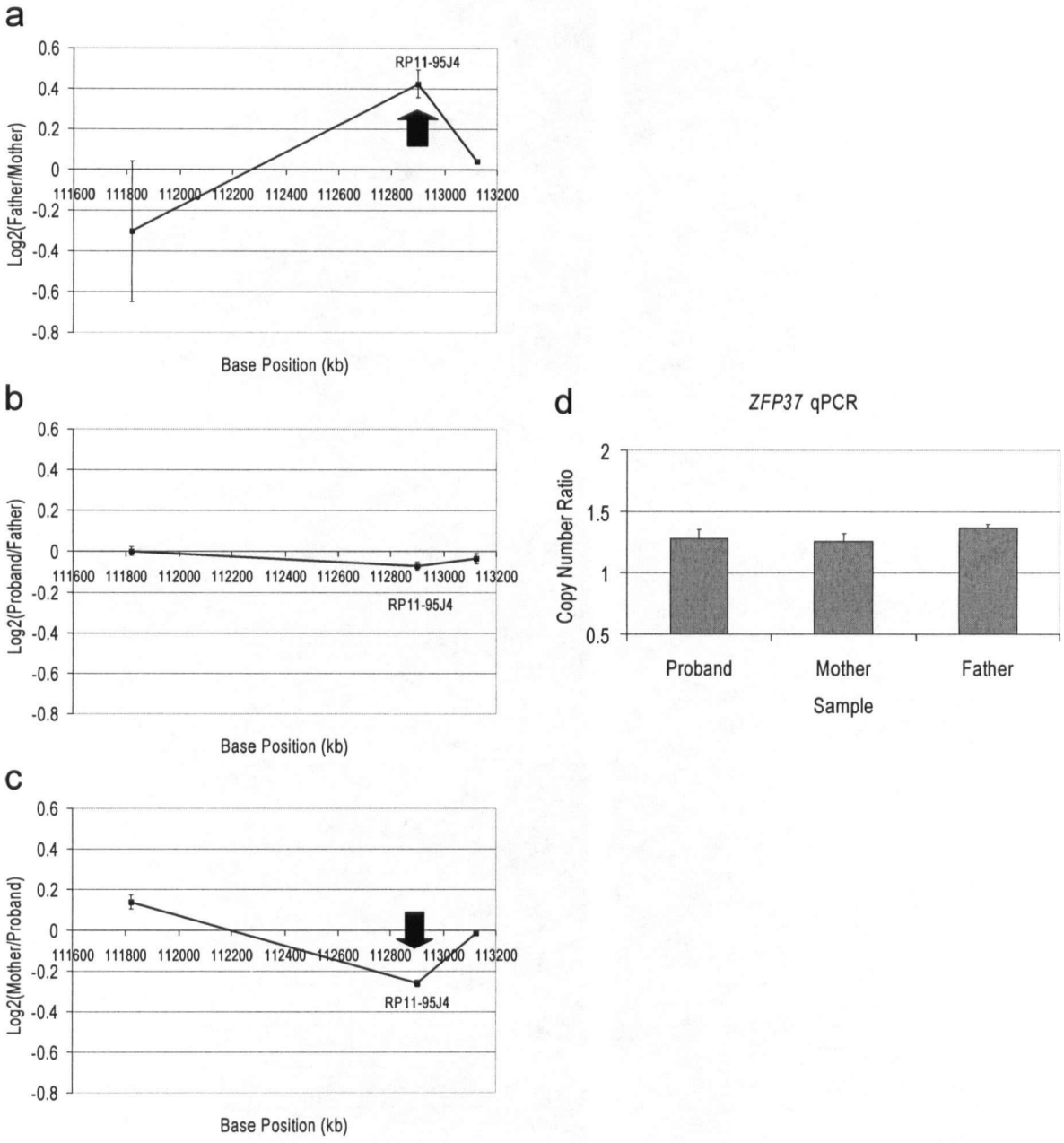

FIGURE 3.16. TRIO 261. The locus (RP11-95J4) which contains *ZFP37* is labeled on the aCGH graphs. Error bars represent the standard deviation for the triplicate spot (aCGH) or duplicate sample (RTqPCR). The $\log_2$ratios for RP11-95J4 that were identified as aberrant are labeled with arrows. (a) aCGH results from the mother-proband hybridization. (b) aCGH results from the father-mother hybridization. (c) aCGH results from the proband-father hybridization. (d) RTqPCR results for trio 261 with the probe for *ZFP37*. Taken together, the aCGH and RTqPCR results are inconclusive.

3.5 INHERITANCE AND DNA COPY NUMBER ABERRATIONS

Based on the three conclusive results between the aCGH and RTqPCR data, each person in a trio was typed for the "alleles" they carried. I used this to determine whether a proband aberration was inherited from the parents or the result of a spontaneous event. Trio 257 showed amplifications in the proband and mother samples by aCGH and RTqPCR (Figure 3.12). The alleles in the father would be two normals (N), while the alleles carried by the proband and mother would be one N and one amplified (A). These results indicated an amplification in the proband inherited from the mother. The trio 297 aCGH and RTqPCR results showed a deletion in the proband sample (Figure 3.13). Both parents carry two N alleles, so the deleted (D) allele in the proband is from a spontaneous event. The aCGH and RTqPCR results from trio 309 (Figure 3.14) showed the proband and father carrying one N and one A allele each, while the mother has two N alleles. This indicates an amplification in the proband inherited from the father.

The transmission/disequilibrium test (TDT) (reviewed in Schaid, 1998) was considered. This test determines whether there is unequal transmission of alleles from heterozygous parents to an affected offspring (reviewed in Schaid, 1998). There were only 2 trios (257 and 309) that had heterozygous parents and showed consistent detection of the aberration using aCGH and RTqPCR. Therefore, the TDT would not have the power needed to identify association between an aberration and psychosis. In addition, heterozygosity of the parent can be ambiguous is there is more than one extra copy of the gene. For example, if there were two extra copies of the gene it is unclear whether those two extra copies are on the same chromosome or not (i.e. the same allele). The TDT is based on χ^2 statistics (reviewed in Schaid,

1998) and the sample size necessary for this type of testing (at $\alpha = 0.05$, df = 2 and a medium effect size) is N = 107 (Cohen, 1992).

4.0 DISCUSSION

The purpose of this project was to find candidate genes for psychosis that could be used as therapeutic targets or in the development of diagnostic tools. I used array comparative genomic hybridization for the initial genome scan and found 11 copy number differences at seven loci. I tested three of the seven aberrant loci using real time quantitative PCR. RTqPCR was used to verify the presence of an aberration and resolve the ambiguity of aCGH results. All 20 trios were tested with the three RTqPCR probes. Eighteen amplifications (including all loci tested) and three deletions (only in *LPA*) were found using RTqPCR. Of the six trios that showed copy number differences by aCGH (within the RTqPCR-tested loci), three showed copy number changes consistent with the results of the RTqPCR assays and the other three were inconclusive.

The most intriguing result from this study is the variation seen in the *LPA* gene. Recently, the Lp(a) protein has been found at significantly increased plasma levels in patients with SCZ, major depression and BPD when compared to healthy control individuals (Emanuele et al., 2006). Although Emanuele et al. (2006) suggest a link between cardiovascular disease and SCZ, the association could be due to any number of factors. While the present study does not show a significant association between the *LPA* locus and psychosis, it is interesting that this locus, already implicated in psychiatric disease, is the most variable region of the genome in this trio set. This type of study represents the potential for development of a diagnostic test for psychiatric disorders. A larger case-control sample size to replicate the plasma protein findings is imperative. To investigate the association of copy number variation in *LPA* with psychiatric

disease, a larger number of trios and the transmission/disequilibrium test (TDT) could be used. To detect association (with 80% power) of a disease with an aberrant copy number, the study would have to include 107 families (for a gene of medium effect and $\alpha = 0.05$, df = 2) (Cohen, 1992). A case-control design could also be used. This type of experiment would have more power because all samples could be included (i.e. homozygous subjects would not be excluded). A disadvantage of this study would be population stratification. Matching the ethnicity of cases and controls would ameliorate this problem. In addition to assessing individuals for *LPA* copy number, investigating the levels of mRNA and protein could uncover a possible correlation with aberrant copy number.

The results from the family 456 proband SMRT array hybridization (Figure 3.3) and the HumArray 2.0 reference hybridization (Figure 3.4) confirm that the potential chromosome 13 aberration detected in family 456 by the HumArray 2.0 is an array artifact (Figure 3.2). However, this artifact does not routinely appear on the HumArray 2.0. Of the 63 arrays completed, only 3 arrays showed the chromosome 13 artifact. The artifact encompasses all of chromosome 13 when viewed by MIA graphical output, and MIA breakpoint analysis identified the whole chromosome as aberrant. Results from these arrays were excluded from further analysis. The artifact could be caused by (1) properties of the slide, or (2) properties of the spotted DNA. The spotted DNA is the most likely cause of the false positive signal because the only regions affected on the slide contain DNA from chromosome 13, and the clones are not segregated on the array by genomic position. To verify this it would be possible to re-generate the spotting solution for all chromosome 13 clones and a number of control loci from other chromosomes. All of this DNA, including the original spotting DNA from the HumArray could be printed on the slides and tested for artifacts. If the new chromosome 13 DNA did not show the same artifact pattern, then it would be clear that the old DNA was a problem.

However, this experiment is probably unnecessary because the chromosome 13 artifact happens only rarely and is recognizable when it does occur.

The results from the SMRT array suggest using a higher density array as a way of exploring copy number variation in the trios. Any regions of genomic DNA not represented by the HumArray 2.0 could be assessed for copy number changes using the SMRT array. In addition, using a larger sample set of trios could identify more aberrations.

One aspect to consider when investigating inherited and spontaneous aberrations is the type of tissue used. If spontaneous aberrations occur later in development, it is possible they will be restricted to specific tissues. SCZ and BPD are brain-specific disorders, so there is a possibility that important disease-specific changes in DNA copy number may be restricted to brain tissue. Post-mortem brain DNA used with aCGH (Wilson et al., 2006a) allows identification of chromosomal aberrations in the brain. However, without DNA from other tissues or the parents, we cannot discriminate between inherited and spontaneous aberrations. The same holds true when using lymphocyte DNA only, as in the present study. These samples allow discrimination between inherited (present in the proband and a parent) and spontaneous (present in proband sample only) chromosomal aberrations in lymphocytes, but do not allow identification of brain-specific aberrations. A comparison between lymphocyte (or other tissues) and brain DNA will identify the differences in copy number between these tissues. If the copy number between tissues is significantly different for more loci perhaps the reason is based on gene function. In certain tissues, particular genes may be under increased selective pressure to remain at a normal copy number. However, a relaxation of this pressure may lead to an increase in copy number changes. This could also lead to formation of multiple cell populations, within a tissue, that differ in copy number.

Six trios were analyzed for consistency between aberrations detected by aCGH and tested by RTqPCR. Three trios showed consistency when the copy number changes were compared. For example, trio 257 showed an aCGH-detected amplification in the mother and proband samples (Figure 3.12). The RTqPCR results showed that the mother and proband samples have increased copy number of the *LPA* gene (Figure 3.8 and Figure 3.12). Taken together, the aCGH and RTqPCR results are consistent with a copy number increase in the mother and proband samples. Overall, the results suggest that aCGH and RTqPCR have a decent concordance rate. This suggests that the copy number changes are not artifacts. A possible explanation for the aCGH-detected copy number changes not consistent with RTqPCR could be related to the differences in detection between the two methods. Array comparative genomic hybridization and real time quantitative PCR are two very different methods for assessing DNA copy number changes. The main difference is the size of the genomic region each method assesses; aCGH investigates the genome using, on average, 150 kb regions, while RTqPCR analyzes 100 bp sequences. Therefore, RTqPCR has an increased likelihood of finding aberrations at a single small locus than aCGH at a single large locus. For example, the signal from aCGH for a single clone is averaged over 150,000 bp. A small section of that region with aberrant copy number will never be deemed significant. On the other hand, RTqPCR could identify the difference in that small region.

Both aCGH and RTqPCR as performed here, are unable to resolve mixed cell populations. The fact that DNA from mixtures of cells has been analyzed limit the interpretation of these results. Many of the ratios observed by RTqPCR do not represent whole copy number changes. For example, the ratio may be between 2 and 3 copies. It is unlikely that inaccurate measurements are the cause, given the reproducibility of the RTqPCR results (Figure 3.11). Instead, this could be explained by the DNA. Since the results from the *LPA* replication

experiment do not suggest inaccurate measurements, perhaps, the samples with a ratio between 2 and 3 copies have lymphocytes that are comprised of cells with 2 copies of the gene of interest and cells with 3 copies of the gene. The division in unlikely to be this clean, but it demonstrates the possibility of multiple cell populations. Another example of populations of cells that differ in copy number is cancer. A study of ovarian cancer found that serous tumours contain a higher proportion of aneuploid cells than inclusion cysts and surface epithelium (Korner et al., 2005). The authors propose that aneuploid cells in the cysts and epithelium may be the precursors of ovarian tumours. These results underscore the importance of understanding the DNA copy number composition of normal tissue, and how it can lead to disease.

Using spectral karyotyping (SKY), fluorescence in situ hybridization (FISH) and flow cytometry for DNA content, it is possible to estimate the number of different populations present. A potential problem with these techniques is that they can only differentiate between cells with differing amounts of DNA or an aberrant number of large loci. Cells differing by changes in single or multiple genes may not be identified. In addition, the proportions estimated for lymphocyte populations may not be an accurate estimation for brain cell populations. However, the experiments using SKY, FISH and flow cytometry are still valid to understand the types of change within a single cell type. A study investigating whole chromosome copy number changes in mice found a significantly higher rate of aneuploidy in embryonic neuroblasts when compared to adult lymphocytes (Rehen et al., 2001). The researchers found 33% aneuploid neuroblasts and only 3% aneuploid adult lymphocytes. In addition, they found that adult mouse neurons showed lower rates of sex chromosome aneuploidy, suggesting that aneuploid cells in the brain are selected against. Given that we do not have human embryonic neuroblasts to examine, we can only speculate as to the rate of aneuploidy in human brain. However, if the findings in mice can be extrapolated to humans, this may be important in the context of human

psychiatric disease. Perhaps certain aneuploid cells in the human embryonic brain are not lost during development, but contribute to disease development or risk. Using lymphocyte DNA, we cannot identify these cells. Even using post-mortem brain tissue may not uncover the aneuploid populations. Comparing the copy number profiles of different parts of the brain, especially those important in SCZ and BPD, may prove useful for finding pathogenic changes.

The design and development of future studies may include other methods of DNA copy number detection. A higher density array for CGH would enable better detection of aberrations. With the HumArray 2.0, any aberrations in the genomic DNA not represented by clones on the array would be missed, and finding multi-clone copy number changes is unlikely. In addition, the approximate breakpoints of aberrations could be mapped by using a higher density array. Other methods of copy number detection include the GeneChip Mapping Arrays from Affymetrix, which are designed for genome-wide linkage and association (Klein et al., 2005), and copy number studies (Zhou et al., 2004). Representational oligonucleotide microarray analysis (ROMA) is another oligonucleotide probe-based method (Lucito et al., 2003). The Affymetrix GeneChip differs from aCGH and ROMA in that individual arrays must be hybridized for each sample and reference DNA. This means that the GeneChip does not represent a direct evaluation of competition between two labeled samples of DNA. Both ROMA and the GeneChip contain repeat-free oligonucleotides as probes, and therefore do not require the use of Cot-1 DNA for blocking repetitive sequences. Cot-1 DNA is expensive and can limit the number of arrays done with CGH. ROMA and the GeneChip have higher resolutions than the HumArray. ROMA has an upper limit to resolution because of the limited number of *Bgl*II fragments in the human genome, and higher resolution arrays for CGH are available (Ishkanian et al., 2004). However, BAC arrays for CGH will not achieve the resolution of oligonucleotide arrays because of the large probe size.

6.0 CONCLUSIONS

In the present study, I investigated DNA copy number changes in trio sets (proband, mother, father) in which the probands have been diagnosed with psychosis. This was an exploratory study using array CGH with 1.4 Mb resolution, I completed low density hybridizations for 20 trio sets. To confirm changes identified by the arrays, I used RTqPCR. The results show a good concordance between methods.

Although the two methods were not in 100% agreement, the practice of using them in combination to find aberrations remains valid. aCGH (especially in a high density format) is a satisfactory method for pinpointing regions of the genome that vary between members of a family, or between control and references samples. RTqPCR, if used after aCGH, can further refine the characteristics of a locus. Taken together, the two methods work well to identify and validate copy number changes. An intriguing result was found by these methods. The most abundant aberration in these trios was in the *LPA* gene, previously shown to be associated with psychiatric disorders. Further evaluation of DNA copy number at this locus could, if findings are positive, lead to a diagnostic test.

Finally, this study was important for determining the correlation between two techniques used to determine copy number and provides more insight into human copy number variation.

7.0 REFERENCES

(2000) Diagnostic and Statistical Manual of Mental Disorders: DSM IV-TR. Washington: American Psychiatric Association.

Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhangal G, Patel SG, Sheehan-Rooney K, Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyle JJ, Pusey CD, Cook HT (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 439:851-855.

Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, Kowbel D, Kuo WL, Gray JW, Pinkel D (2000) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat Genet 25:144-146.

Axelrod J, Tomchick R (1958) Enzymatic O-methylation of epinephrine and other catechols. J Biol Chem 233:702-705.

Baldessarini RJ, Tarazi FI (2001) Drugs and the Treatment of Psychiatric Disorders: Psychosis and Mania. In: Goodman and Gilman's The Pharmacological Basis of Therapeutics (Hardman JG, Limbird LE, Gilman AG, eds), pp 485-514. New York: McGraw-Hill.

Barrett TB, Hauger RL, Kennedy JL, Sadovnick AD, Remick RA, Keck PE, McElroy SL, Alexander M, Shaw SH, Kelsoe JR (2003) Evidence that a single nucleotide polymorphism in the promoter of the G protein receptor kinase 3 gene is associated with bipolar disorder. Mol Psychiatry 8:546-557.

Bassett AS, Chow EW, AbdelMalik P, Gheorghiu M, Husted J, Weksberg R (2003) The schizophrenia phenotype in 22q11 deletion syndrome. Am J Psychiatry 160:1580-1586.

Bearden CE, Jawad AF, Lynch DR, Sokol S, Kanes SJ, McDonald-McGinn DM, Saitta SC, Harris SE, Moss E, Wang PP, Zackai E, Emanuel BS, Simon TJ (2004) Effects of a functional COMT polymorphism on prefrontal cognitive function in patients with 22q11.2 deletion syndrome. Am J Psychiatry 161:1700-1702.

Bender HU, Almashanu S, Steel G, Hu CA, Lin WW, Willis A, Pulver A, Valle D (2005) Functional consequences of PRODH missense mutations. Am J Hum Genet 76:409-420.

Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ (2001) Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. Am J Hum Genet 69:428-433.

Bray NJ, Owen MJ (2001) Searching for schizophrenia genes. Trends Mol Med 7:169-174.

Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS (2000) Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. Science 288:678-682.

Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, Venturi P, Jones LA, Lewis SW, Sham PC, Gottesman, II, Farmer AE, McGuffin P, Reveley AM, Murray RM (1999) Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. Arch Gen Psychiatry 56:162-168.

Carter NP (2004) As normal as normal can be? Nat Genet 36:931-932.

Chen X, Wang X, O'Neill AF, Walsh D, Kendler KS (2004) Variants in the catechol-o-methyltransferase (COMT) gene are associated with schizophrenia in Irish high-density families. Mol Psychiatry 9:962-967.

Chowdari KV, Mirnics K, Semwal P, Wood J, Lawrence E, Bhatia T, Deshpande SN, B KT, Ferrell RE, Middleton FA, Devlin B, Levitt P, Lewis DA, Nimgaonkar VL (2002) Association and linkage analyses of RGS4 polymorphisms in schizophrenia. Hum Mol Genet 11:1373-1380.

Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, Barry C, Tanaka H, La Rosa P, Puech A, Tahri N, Cohen-Akenine A, Delabrosse S, Lissarrague S, Picard FP, Maurice K, Essioux L, Millasseau P, Grel P, Debailleul V, Simon AM, Caterina D, Dufaure I, Malekzadeh K, Belova M, Luan JJ, Bouillot M, Sambucy JL, Primas G, Saumier M, Boubkiri N, Martin-Saumier S, Nasroune M, Peixoto H, Delaye A, Pinchot V, Bastucci M, Guillou S, Chevillon M, Sainz-Fuertes R, Meguenni S, Aurich-Costa J, Cherif D, Gimalac A, Van Duijn C, Gauvreau D, Ouellette G, Fortier I, Raelson J, Sherbatich T, Riazanskaia N, Rogaev E, Raeymaekers P, Aerssens J, Konings F, Luyten W, Macciardi F, Sham PC, Straub RE, Weinberger DR, Cohen N, Cohen D (2002) Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proc Natl Acad Sci U S A 99:13675-13680.

Cohen J (1992) A power primer. Psychological Bulletin 112:155-159.

Emanuele E, Carlin MV, D'Angelo A, Peros E, Barale F, Geroldi D, Politi P (2006) Elevated plasma levels of lipoprotein(a) in psychiatric patients: a possible contribution to increased vascular risk. Eur Psychiatry 21:129-133.

Farrer M, Kachergus J, Forno L, Lincoln S, Wang DS, Hulihan M, Maraganore D, Gwinn-Hardy K, Wszolek Z, Dickson D, Langston JW (2004) Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications. Ann Neurol 55:174-179.

Gong J, Xu J, Bezanilla M, van Huizen R, Derin R, Li M (1999) Differential stimulation of PKC phosphorylation of potassium channels by ZIP1 and ZIP2. Science 285:1565-1569.

Harrison PJ, Weinberger DR (2005) Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Mol Psychiatry 10:420.

Heinrichs RW (2001) In Search of Madness. New York: Oxford University Press.

Hennah W, Varilo T, Kestila M, Paunio T, Arajarvi R, Haukka J, Parker A, Martin R, Levitzky S, Partonen T, Meyer J, Lonnqvist J, Peltonen L, Ekelund J (2003) Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. Hum Mol Genet 12:3151-3159.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949-951.

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL (2004) A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 36:299-303.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462-467.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818-821.

Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ (2003) The UCSC Genome Browser Database. Nucl Acids Res 31:51-54.

Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12:656-664.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler aD (2002) The Human Genome Browser at UCSC. Genome Res 12:996-1006.

King RC, Stansfield WD (1997) A Dictionary of Genetics. New York: Oxford University Press.

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392:605-608.

Klein OD, Cotter PD, Albertson DG, Pinkel D, Tidyman WE, Moore MW, Rauen KA (2004) Prader-Willi syndrome resulting from an unbalanced translocation: characterization by array comparative genomic hybridization. Clin Genet 65:477-482.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385-389.

Korner M, Burckhardt E, Mazzucchelli L (2005) Different proportions of aneusomic cells in ovarian inclusion cysts associated with serous borderline tumours and serous high-grade carcinomas support different pathogenetic pathways. J Pathol 207:20-26.

Krzywinski M, Bosdet I, Smailus D, Chiu R, Mathewson C, Wye N, Barber S, Brown-John M, Chan S, Chand S, Cloutier A, Girn N, Lee D, Masson A, Mayo M, Olson T, Pandoh P, Prabhu AL, Schoenmakers E, Tsai M, Albertson D, Lam W, Choy CO, Osoegawa K, Zhao S, de Jong PJ, Schein J, Jones S, Marra MA (2004) A set of BAC clones spanning the human genome. Nucleic Acids Res 32:3651-3660.

Levinson DF, Holmans PA, Laurent C, Riley B, Pulver AE, Gejman PV, Schwab SG, Williams NM, Owen MJ, Wildenauer DB, Sanders AR, Nestadt G, Mowry BJ, Wormley B, Bauche S, Soubigou S, Ribble R, Nertney DA, Liang KY, Martinolich L, Maier W, Norton N, Williams H, Albus M, Carpenter EB, DeMarchi N, Ewen-White KR, Walsh D, Jay M, Deleuze JF, O'Neill FA, Papadimitriou G, Weilbaecher A, Lerer B, O'Donovan MC, Dikeos D, Silverman JM, Kendler KS, Mallet J, Crowe RR, Walters M (2002) No major schizophrenia locus detected on chromosome 1q in a large multicenter sample. Science 296:739-741.

Li T, Ma X, Sham PC, Sun X, Hu X, Wang Q, Meng H, Deng W, Liu X, Murray RM, Collier DA (2004) Evidence for association between novel polymorphisms in the PRODH gene and schizophrenia in a Chinese population. Am J Med Genet B Neuropsychiatr Genet 129:13-15.

Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, Pinkel D, Eichler EE (2003) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. Genome Res 13:347-357.

Lu R, Schuster VL (1998) Molecular cloning of the gene for the human prostaglandin transporter hPGT: gene organization, promoter activity, and chromosomal localization. Biochem Biophys Res Commun 246:805-812.

Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M (2003) Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 13:2291-2305.

MacIntyre DJ, Blackwood DH, Porteous DJ, Pickard BS, Muir WJ (2003) Chromosomal abnormalities and mental illness. Mol Psychiatry 8:275-287.

Mason PJ, Bautista JM, Vulliamy TJ, Turner N, Luzzatto L (1990) Human red cell glucose-6-phosphate dehydrogenase is encoded only on the X chromosome. Cell 62:9-10.

Matsumine H, Saito M, Shimoda-Matsubayashi S, Tanaka H, Ishikawa A, Nakagawa-Hattori Y, Yokochi M, Kobayashi T, Igarashi S, Takano H, Sanpei K, Koike R, Mori H, Kondo T, Mizutani Y, Schaffer AA, Yamamura Y, Nakamura S, Kuzuhara S, Tsuji S, Mizuno Y (1997) Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27. Am J Hum Genet 60:588-596.

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM (2006) Common deletion polymorphisms in the human genome. Nat Genet 38:86-92.

McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A (2003) The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch Gen Psychiatry 60:497-502.

Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, Clair DM, Muir WJ, Blackwood DH, Porteous DJ (2000) Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 9:1415-1423.

Miyamoto S, Duncan GE, Marx CE, Lieberman JA (2005) Treatments for schizophrenia: a critical review of pharmacology and mechanisms of action of antipsychotic drugs. Mol Psychiatry 10:79-104.

Murphy KC, Jones LA, Owen MJ (1999) High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry 56:940-945.

Nolan KA, Bilder RM, Lachman HM, Volavka J (2004) Catechol O-methyltransferase Val158Met polymorphism in schizophrenia: differential effects of Val and Met alleles on cognitive stability and flexibility. Am J Psychiatry 161:359-361.

Owen MJ, Williams NM, O'Donovan MC (2004) The molecular genetics of schizophrenia: new findings promise new insights. Mol Psychiatry 9:14-27.

Payen E, Verkerk T, Michalovich D, Dreyer SD, Winterpacht A, Lee B, De Zeeuw CI, Grosveld F, Galjart N (1998) The centromeric/nucleolar chromatin protein ZFP-37 may function to specify neuronal nuclear domains. J Biol Chem 273:9099-9109.

Polymeropoulos MH, Higgins JJ, Golbe LI, Johnson WG, Ide SE, Di Iorio G, Sanges G, Stenroos ES, Pho LT, Schaffer AA, Lazzarini AM, Nussbaum RL, Duvoisin RC (1996) Mapping of a Gene for Parkinson's Disease to Chromosome 4q21-q23. Science 274:1197-1199.

Rao SK, Huynh C, Proux-Gillardeaux V, Galli T, Andrews NW (2004) Identification of SNAREs involved in synaptotagmin VII-regulated lysosomal exocytosis. J Biol Chem 279:20471-20479.

Rehen SK, McConnell MJ, Kaushal D, Kingsbury MA, Yang AH, Chun J (2001) Chromosomal variation in neurons of the developing and adult mammalian nervous system. Proc Natl Acad Sci U S A 98:13361-13366.

Rosa A, Peralta V, Cuesta MJ, Zarzuela A, Serrano F, Martinez-Larrea A, Fananas L (2004) New evidence of association between COMT gene and prefrontal neurocognitive function in healthy individuals from sibling pairs discordant for psychosis. Am J Psychiatry 161:1110-1112.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general use and for biologist programmers. In: Bioinformatics Methods and Protocols: Methods in Molecular Biology (Krawetz SM, S., ed), pp 365-386. Totowa: Humana Press.

Saito M, Matsumine H, Tanaka H, Ishikawa A, Shimoda-Matsubayashi S, Schaffer AA, Mizuno Y, Tsuji S (1998) Refinement of the gene locus for autosomal recessive juvenile parkinsonism (AR-JP) on chromosome 6q25.2-27 and identification of markers exhibiting linkage disequilibrium. J Hum Genet 43:22-31.

Schaid DJ (1998) Transmission disequilibrium, family controls, and great expectations. Am J Hum Genet 63:935-941.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525-528.

Shifman S, Bronstein M, Sternfeld M, Pisante-Shalom A, Lev-Lehman E, Weizman A, Reznik I, Spivak B, Grisaru N, Karp L, Schiffer R, Kotler M, Strous RD, Swartz-Vanetik M, Knobler HY, Shinar E, Beckmann JS, Yakir B, Risch N, Zak NB, Darvasi A (2002) A highly significant association between a COMT haplotype and schizophrenia. Am J Hum Genet 71:1296-1302.

Shih Ie M, Sheu JJ, Santillan A, Nakayama K, Yen MJ, Bristow RE, Vang R, Parmigiani G, Kurman RJ, Trope CG, Davidson B, Wang TL (2005) Amplification of a chromatin remodeling gene, Rsf-1/HBXAP, in ovarian carcinoma. Proc Natl Acad Sci U S A 102:14004-14009.

Shprintzen RJ, Goldberg R, Golding-Kushner KJ, Marion RW (1992) Late-onset psychosis in the velo-cardio-facial syndrome. Am J Med Genet 42:141-142.

Shprintzen RJ, Higgins AM, Antshel K, Fremont W, Roizen N, Kates W (2005) Velo-cardio-facial syndrome. Curr Opin Pediatr 17:725-730.

Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K (2003) alpha-Synuclein locus triplication causes Parkinson's disease. Science 302:841.

Sklar P (2002) Linkage analysis in psychiatric disorders: the emerging picture. Annu Rev Genomics Hum Genet 3:371-413.

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D, Albertson DG (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. Nat Genet 29:263-264.

Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E, Gunnarsdottir S, Walker N, Petursson H, Crombie C, Ingason A, Gulcher JR, Stefansson K, St Clair D (2003) Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. Am J Hum Genet 72:83-87.

Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, Gunnarsdottir S, Ivarsson O, Chou TT, Hjaltason O, Birgisdottir B, Jonsson H, Gudnadottir VG, Gudmundsdottir E, Bjornsson A, Ingvarsson B, Ingason A, Sigfusson S, Hardardottir H, Harvey RP, Lai D, Zhou M, Brunner D, Mutel V, Gonzalo A, Lemke G, Sainz J, Johannesson G, Andresson T, Gudbjartsson D, Manolescu A, Frigge ML, Gurney ME, Kong A, Gulcher JR, Petursson H, Stefansson K (2002) Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 71:877-892.

Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B, Cesare AJ, Gibberman A, Wang X, O'Neill FA, Walsh D, Kendler KS (2002) Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. Am J Hum Genet 71:337-348.

Walker E, Kestler L, Bollini A, Hochman KM (2004) Schizophrenia: etiology and course. Annu Rev Psychol 55:401-430.

Wilson GM, Flibotte S, Chopra V, Melnyk BL, Honer WG, Holt RA (2006a) DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling. Hum Mol Genet.

Wilson GM, Flibotte S, Missirlis PI, Marra MA, Jones S, Thornton K, Clark AG, Holt RA (2006b) Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. Genome Res 16:173-181.

Zar JH (1999) Biostatistical Analysis, 4th Edition. Upper Saddle River, NJ: Prentice Hall.

Zhou X, Mok SC, Chen Z, Li Y, Wong DT (2004) Concurrent analysis of loss of heterozygosity (LOH) and copy number abnormality (CNA) for oral premalignancy progression using the Affymetrix 10K SNP mapping array. Hum Genet 115:327-330.

APPENDIX A. PROBAND DETAILS.

TABLE A1. DETAILS OF PROBAND SAMPLES.

Trio Identification	Gender	Age	Ethnicity
163	Male	22	Caucasian
187	Male	23	Caucasian
196	Male	26	Caucasian
201	Female	20	Caucasian
220	Female	21	Caucasian
231	Male	23	Caucasian
239	Male	19	Caucasian
257	Female	18	European Caucasian
261	Female	24	Caucasian
297	Male	20	Caucasian
309	Male	21	Caucasian
311	Female	18	N/A
341	Male	18	Caucasian
395	Male	19	Asian
408	Male	20	N/A
414	Female	21	N/A
446	Male	20	European Caucasian
456	Female	N/A	N/A
462	Male	21	N/A
601	Female	N/A	N/A

APPENDIX B. MYSQL DATABASE.

TABLE B1. DESCRIPTION OF TRIOS TABLE IN MYSQL DATABASE.

Field	Type	Null	Key	Default	Extra
row_id	int(11)		PRI	NULL	auto_increment
clone_id	varchar(25)	YES		NULL	
trio_id	varchar(5)	YES		NULL	
cy3_dna	varchar(5)	YES		NULL	
cy5_dna	varchar(5)	YES		NULL	
slide	tinyint(4)	YES		NULL	
array	tinyint(4)	YES		NULL	
log2_ratio	decimal(7,4)	YES		NULL	
num_reps	int(11)	YES		NULL	
log2_int	decimal(7,4)	YES		NULL	
log2_stdev	decimal(7,4)	YES		NULL	
log2_uncert	decimal(7,4)	YES		NULL	
away	decimal(7,4)	YES		NULL	

TABLE B2. DESCRIPTION OF TRIOS_HD TABLE IN MYSQL DATABASE.

Field	Type	Null	Key	Default
row_id	int		PRI	NULL
clone_name	varchar(25)	YES		NULL
trio_id	varchar(5)	YES		NULL
test_dna	varchar(5)	YES		NULL
ref_dna	varchar(5)	YES		NULL
position	int	YES		NULL
chromosome	int	YES		NULL
start_pos	int	YES		NULL
stop_pos	int	YES		NULL
num_replicates	int	YES		NULL
log2_intensity	decimal(7,4)	YES		NULL
log2_ratio	decimal(7,4)	YES		NULL
log2_stdev	decimal(7,4)	YES		NULL
log2_uncertainty	decimal(7,4)	YES		NULL
away	decimal(7,4)	YES		NULL

QUERY B1. CRITERIA FOR FINDING ABERRANT CLONES FROM HUMARRAY 2.0 DATA.

```
SELECT c.clone_name, c.chromosome, c.position, t.trio_id, t.cy3_dna, t.c y5_dna, t.away,
t.log2_ratio
FROM trios t, clone c
WHERE away > 3
AND log2_stdev < 0.2
AND log2_int > 10
AND chromosome < 23
AND t.clone_id = c.clone_name;
```

QUERY B2. CRITERIA FOR CALCULATING INTER-ARRAY VARIABILITY.

```
SELECT clone_id, std(log2_ratio)
FROM trios
WHERE clone_id = 'RP11-122N11' or clone_id = 'RP4-693L23' or clone_id = 'RP11-82D16'
or clone_id = 'RP11-60B6' or clone_id = 'RP11-1P3' or clone_id = 'RP11-88C10' or clone_id =
'RP11-43B19' or clone_id = 'RP11-95J4' or clone_id = 'RP11-88L18' or clone_id = 'RP11-
122P17' or clone_id = 'RP11-31N23' or clone_id = 'RP11-11C17' or clone_id = 'RP11-182E4'
or clone_id = 'RP11-205J24' or clone_id = 'RP11-234M13' or clone_id = 'RP11-189K9'
GROUP BY clone_id;
```

QUERY B3. CRITERIA FOR CALCULATING INTRA-ARRAY VARIABILITY.

```
SELECT trio_id, slide, array, Cy3_dna, Cy5_dna, std(log2_ratio)
FROM trios
GROUP BY trio_id, slide, array
ORDER BY trio_id, slide, array;
```

QUERY B4. CRITERIA FOR FINDING ABERRANT CLONES FROM TRIO 456 HIGH DENSITY ARRAY DATA.

```
SELECT clone_name, position, trio_id, test_dna, ref_dna, log2_ratio, away, chromosome,
start_pos, stop_pos
FROM trios_hd
WHERE away > 3
AND log2_stdev < 0.2
AND log2_intensity > 10
AND chromosome < 23;
```

APPENDIX C. QUANTITATIVE PCR.

>chr1:2144215-2144770
CAGATCCCCAGACGACTCAGATGCACGGACACCCAGATGACATGGATGCA
CGGACGACTCAGATCCACAGATGACTCAAACGCACAGATGACTGGGATGC
TCAGATGACCGCTGCTGTGCTGCCACCCCAGGGCTCTCTGGGCGTTTCCT
GTGGTTTGGGCACCAGGAGCCTGGGAGTCCCATGCTGCCCCCAGGGCACT
ACCTCCTGGGCCCAGCCCTGCATCCGGTGGCAGGGCTCACCGTCATCACC
*CCAACAGTGCAGGGTGGTCTCAGGGACCTCCTCTCATCATTGCC*AAGaac
tggctccaggatgtttccatgtggccggctagtatggccaaagtggaccc
tggcgtgctgtccccttggacgcctccaggccctgcccagcacgtggGGC
TCGTCCATTCTGTGCCTGACCATGCTCTGCCATGCGGGGCCTAGCCCAGC
CCCCCAGCCCTGCTGCTTCTCCCCACCCCACCCCCGCCACCTCCACCAAG
CCACCAGCATCCTGCCTGGCCCTACGGACAGCAGGGTCGTCCTGTGTCCA
AAAGCC

FIGURE C1. REGION OF CHROMOSOME 1 (RP11-82D16, *PRKCZ*) FOR RTQPCR ANALYSIS. Underlined regions are 4 mono-nucleotide stretches, highlighted regions are >4 mono-nucleotide stretches and **boldface regions** are 100% identical to another genomic segment. SUB-REGION 1. *SUB-REGION 2.* sub-region 3.

TABLE C1. PRIMER AND PROBE SEQUENCES. ALL SEQUENCES ARE WRITTEN AS 5'→3'.

Locus	Probe (6FAM NNNNN MGB NFQ)	Forward Primer	Reverse Primer
PRKCZ	CTGGAGCCAGTTCTTG	GGGACCTCCTCTCATCATTG	CCGGCCACATGGAAACATC
LPA	TCAGGTGGGAGTACTGC	CTTCTGCGTCTGAGCATTGC	GTGGCAGCTCCTTATTGTTATACGA
ZFP37	TCTCCATGGGCACAGTTG	GCTCCTGCAGCCTATACTATATTACC	GTTGGACATCTTGGTTTGAGTCAGA
SYT7	CAGTACCTCCACAACCC	GACCCTTTTCTCAACCCACAGA	CTGTGGCTGGAGCTGCTA
PGT	CCATCCATGTCCTCATCTC	ATCCCCAAAGCACCTGGTTT	AGAGGCCAAGATAGTCCTGGTAA
G6PD	CCGTGAGGCCTGGCGTA	GGGCTTCTCCAGCTCAATCTG	GCCTCCCAAGCCATACTATGTC

TABLE C2. CALCULATION OF *PRKCZ* AND *G6PD* COPY NUMBER FOR TRIO 187.

Sample	*PRKCZ* C_t	*PGT* C_t	*G6PD* C_t	*PGT-PRKCZ*	*PGT-G6PD*	ddC_t*PRKCZ*	ddC_t*G6PD*	$2^{-ddC_t PRKCZ}$	$2^{-ddC_t G6PD}$	*PRKCZ* CN	*PRKCZ* SD	*G6PD* CN	*G6PD* SD
187_F	27.21	27.76	28.36	0.55	-0.6	-0.34	-0.45	1.27	1.37	1.31	0.05	1.27	0.14
	27.12	27.76	28.58	0.64	-0.83	-0.43	-0.22	1.34	1.17				
187_M	27.03	27.57	27.41	0.54	0.16	-0.33	-1.21	1.25	2.31	1.17	0.12	2.33	0.03
	27.18	27.52	27.33	0.33	0.19	-0.12	-1.24	1.09	2.36				
187_P	27.33	27.52	28.54	0.19	-1.03	0.02	-0.02	0.98	1.01	0.96	0.04	1.03	0.03
	27.28	27.39	28.37	0.11	-0.98	0.11	-0.07	0.93	1.05				
Male	27.27	27.48	28.53	0.21	-1.05	0	0	1	1	1		1	
Female	27.55	27.94	27.86	0.39	0.07	-0.18	-1.12	1.13	2.17	0.99	0.12	2.01	0.15
	27.76	27.84	27.97	0.07	-0.14	0.14	-0.91	0.91	1.88				
	27.74	27.85	27.91	0.11	-0.06	0.1	-0.98	0.93	1.98				

APPENDIX D. STATISTICAL ANALYSIS

TABLE D1. COEFFICIENT OF VARIATION (V) FOR THREE REFERENCE HYBRIDIZATIONS.

Hybridization	Standard Deviation (s)	Mean (X)	Sample Size (n)	Coefficient of Variation (V)
Array 1	0.141	0.0857	2464	1.645
Array 2	0.114	0.0725	2464	1.572
**Array 3	0.102	0.0706	2464	1.445

**reference hybridization showing chromosome 13 aberration

$$V_1 = \frac{s_1}{\bar{X}_1} = \frac{0.141}{0.0857} = 1.645$$

$$Z_{1,2} = \frac{V_1 - V_2}{\sqrt{\left(\frac{V_p{}^2}{n_1 - 1} + \frac{V_p{}^2}{n_2 - 1}\right)\left(0.5 + V_p{}^2\right)}} = \frac{1.645 - 1.572}{\sqrt{\left(\frac{1.609}{2464 - 1} + \frac{1.609}{2464 - 1}\right)(0.5 + 1.609)}} = 1.391$$

$$V_p{}^2 = \frac{(n_1 - 1)V_1 + (n_2 - 1)V_2}{(n_1 - 1) + (n_2 - 1)} = \frac{(2464 - 1)(1.645) + (2464 - 1)(1.572)}{(2464 - 1) + (2464 - 1)} = 1.609$$

$Z_{0.05(2)} = 1.96$

Because $Z_{1,2} = 1.609 < Z_{0.05(2)}$, the null hypothesis ($H_o$: the coefficients of variation of array 1 and array 2 are the same) cannot be rejected.

Because $Z_{1,3} = 3.949 > Z_{0.05(2)}$, the null hypothesis ($H_o$: the coefficients of variation of array 1 and array 3 are the same) can be rejected.

Because $Z_{2,3} = 2.560 > Z_{0.05(2)}$, the null hypothesis($H_o$: the coefficients of variation of array 2 and array 3 are the same) can be rejected.