# Copy Number Variation Distribution in Six Monozygotic Twin Pairs Discordant for Schizophrenia

Christina A. Castellani,[1] Zain Awamleh,[1] Melkaye G. Melka,[1] Richard L. O'Reilly,[2] and Shiva M. Singh[1]

[1]Department of Biology, The University of Western Ontario, London, Ontario, Canada
[2]Department of Psychiatry, The University of Western Ontario, London, Ontario, Canada

We have evaluated copy number variants (CNVs) in six monozygotic twin pairs discordant for schizophrenia. The data from Affymetrix® Human SNP 6.0 arrays™ were analyzed using Affymetrix® Genotyping Console™, Partek® Genomics Suite™, PennCNV, and Golden Helix SVS™. This yielded both program-specific and overlapping results. Only CNVs called by Affymetrix Genotyping Console, Partek Genomics Suite, and PennCNV were used in further analysis. This analysis included an assessment of calls in each of the six twin pairs towards identification of unique CNVs in affected and unaffected co-twins. Real time polymerase chain reaction (PCR) experiments confirmed one CNV loss at 7q11.21 that was found in the affected patient but not in the unaffected twin. The results identified CNVs and genes that were previously implicated in mental abnormalities in four of the six twin pairs. It included *PYY* (twin pairs 1 and 5), *EPHA3* (twin pair 3), *KIAA1211L* (twin pair 4), and *GPR139* (twin pair 5). They represent likely candidate genes and CNVs for the discordance of four of the six monozygotic twin pairs for this heterogeneous neurodevelopmental disorder. An explanation for these differences is ontogenetic de novo events that differentiate in the monozygotic twins during development.

■ **Keywords:** monozygotic twins, discordance, schizophrenia, genomes, de novo, copy number variations

Schizophrenia (OMIM (Online Mendelian Inheritance in Man®): 181500) is a chronic, severe, and debilitating mental illness that generally presents in early adult life and is characterized by a disruption of perception and thinking. The lifetime prevalence of this disease is ∼1% with little variation in prevalence seen throughout the world (Public Health Agency of Canada Steering Committee on Health — Adjusted Life Expectancy, 2013). The disease clusters in some families and has a high heritability estimate (80%; Sullivan et al., 2003). In fact, the best predictor of the occurrence of this disease is family history. The inheritance pattern of schizophrenia is complex. This complexity is reflected in the observation that monozygotic twins, who are said to share 100%, and dizygotic twins who are said to share 50%, of their genetic makeup are concordant in only 48% and 17% of cases, respectively (McGuffin et al., 1994). These observations suggest a role for non-genetic and random genetic factors (O'Reilly & Singh, 1996; Singh & O'Reilly, 2009), including random developmental events (Singh et al., 2004), epigenetic mechanisms (Singh et al., 2002), and environmental factors (Torrey et al., 1997). Over 30 years of genetic research using linkage and associ-

ation analysis have identified a number of promising linkages (Sullivan, 2005) and candidate genes (Hamilton, 2008; Karayiorgou & Gogos, 2006). Most of these results have been difficult to reliably replicate except in the case of a few variants, which have been associated across multiple studies, typically when large sample sizes are employed (Ripke et al., 2013; Torkamani et al., 2010). This difficulty in identifying causal genes for schizophrenia has been attributed to extensive heterogeneity, including different patients from the same family (Beckmann & Franzek, 2000). Application of genome-wide expression arrays in schizophrenia has identified a long list of genes with altered expression in the brain (McInnes & Lauriat, 2006; Verveer et al., 2007) and blood tissue (Gladkevich et al., 2004; Tsuang et al., 2005).

However, altered expression of these genes cannot always be replicated and may be a secondary effect.

Recent advances in human genomics have helped in the identification of structural variants, termed copy number variants (CNVs), and opened a new direction in schizophrenia genetics research (Kirov, 2010). Copy number variants lead to deletions and duplications of a given segment of the genome. They are common (Conrad et al., 2010) and widespread in the human genome (Iafrate et al., 2004). By virtue of their variable size, they may directly disrupt multiple genes that are co-located (Feuk et al., 2006). In addition to having a direct effect on the expression of the amplified or deleted genes (Stranger et al., 2007), they may have indirect effects on gene expression extending upstream and downstream of the CNV region (Henrichsen et al., 2009). While most CNVs are polymorphic, some are generated de novo (Zogopoulos et al., 2007). The common CNVs in humans are believed to play a role in evolution (Lee & Scherer, 2010). They also underlie a significant proportion of variation in humans, including differences in cognitive, behavioral, and psychological features (Lee & Lupski, 2006). Further, they have been implicated across a wide variety of common disorders (Buchanan & Scherer, 2008; Stankiewicz & Lupski, 2010; Wellcome Trust Case Control Consortium et al., 2010), including mental disorders (Feuk et al., 2006; Lee & Lupski, 2006; McCarroll & Altshuler, 2007), particularly autism (OMIM: 209850; Autism Genome Project Consortium et al., 2007; Glessner et al., 2009; Moessner et al., 2007; Sebat et al., 2007; Wang et al., 2009) and schizophrenia (Glessner et al., 2010; Kirov et al., 2008; Need et al., 2009; Stefansson et al., 2008; Walsh et al., 2008; Xu et al., 2008). The results, generated with increasing genomic coverage and numbers of patients, have identified a set of candidate CNVs. These include rare deletions at 1q21.1, 15q13.3, 15q11.2, and 22q11.2, as well as duplications at 16p11.2, 16p13.1, and 7q36.3 (Kirov et al., 2008). In addition, various gene regions have been associated with copy number variation in schizophrenia, namely deletions of *NRXN1* (Entrez Gene: 9378), *APBA2* (Entrez Gene: 321), and *CNTNAP2* (Entrez Gene: 26047; Friedman et al., 2008; Liu et al., 2002).

The findings in the field also suggest that, with few exceptions, schizophrenia is caused by aberrations in a relatively large number of genes, most with relatively small effects, that cumulatively produce a genetic predisposition. Some of these aberrations may be inherited while others may represent de novo events (Singh et al., 2009). The field is starting to recognize that rare variants likely play a role in the causation of schizophrenia. This model is not compatible with traditional experiments in which a group of patients are compared with an equally large group of unaffected controls. In such an approach, adding more patients will add additional genetic heterogeneity across cases. This complexity is likely better approached by the precise genetic matching of patients with unaffected controls that can be achieved using monozygotic twins. Even if rare variants identified using this approach are limited to a given set of twins or a given family, they are likely to help in identifying the underlying pathways and genes involved in this disorder.

In this research, we used six pairs of monozygotic twins discordant (MZD) for schizophrenia and assessed the CNV differences between twin pairs. The resulting CNV differences are of interest in identifying patient-specific differences, including gene dosage changes that may differ in a MZD pair. Previous studies utilizing monozygotic twins have associated CNV and methylation differences between twins with various diseases. Using monozygotic twins, three somatic CNV events were found to be associated with discordance for congenital heart defects (Breckpot et al., 2012). Similarly, two de novo CNVs — a pre-twinning duplication and a post-twinning deletion were found to be associated with attention problems (Ehli et al., 2012). Another study looking at Rett syndrome in discordant monozygotic twins found differences in deoxyribonucleic acid (DNA) methylation between twins detected in fibroblasts in the upstream region of genes involved in brain function to be associated with the disease (Miyake et al., 2013). The results are twin-specific and trends are not always consistent (Bloom et al., 2013; Halder et al., 2012; Maiti et al., 2011). Some studies call CNV differences (Maiti et al., 2011) while others call no difference in CNVs between MZ twins discordant for schizophrenia (Bloom et al., 2013). In either case, the MZD strategy is effective in the identification of previously undiscovered genes in schizophrenia, particularly when combined with the use of multiple software programs. Given the high heterogeneity of this disorder, we would a priori expect many aberrations to be patient-specific. These patient-specific genetic changes can be best identified using nature's best match for each patient — their monozygotic twin. We have hypothesized that the discordance of monozygotic twins for schizophrenia may involve de novo mutations (DNM; Singh et al., 2009). If that is so, we should be able to identify differences between MZD twins for schizophrenia that are de novo in nature and do not apply to all twin pairs, but instead show twin-pair specificity. In this report, we have employed a stringent copy number variation detection protocol using multiple CNV calling methods, and identified CNV differences between MZD for schizophrenia. The results support the potential presence of de novo CNVs that are compatible with the development of schizophrenia.

## Materials and Methods
### Ethics Statement and Clinical Background
This study received ethics approval by the University of Western Ontario's Committee on research involving human subjects. All subjects provided written informed consent to participate in this study and were interviewed by a psychiatrist (ROR) using the SCID-I (Structured Clinical

Christina A. Castellani, Zain Awamleh, Melkaye G. Melka, Richard L. O'Reilly and Shiva M. Singh

Interview for DSM-IV Axis I Disorders; First et al., 1996) and the SCID-II (Structured Clinical Interview for DSM-IV Axis II Disorders; First et al., 1997). All of the patients were adults at the time of consent. Past clinical notes were obtained to aid diagnosis. Whole blood samples were obtained from each individual. The twin pairs studied ranged in the age from 20 to 53 years at the time of sample collection. Three of the pairs were female and three of the pairs were male. The twins were discordant for schizophrenia (defined as the time of first contact with mental health services because of symptoms of mental illness) for 4 to 31 years. The strategy used to generate and interpret the molecular results is outlined in the flowchart (Figure 1).

### DNA Preparation, Hybridization and CEL File Analysis

Deoxyribonucleic acid was extracted from whole blood using the PerfectPure DNA Blood Kit (http://www.5prime.com) following the manufacturer's protocol. Whole genome microarray analysis using the Affymetrix® Genome-Wide Human SNP Array 6.0™ was performed at the London Regional Genomics Center (LRGC) following the manufacturer's protocol. For downstream analysis of .cel files, Affymetrix® Genotyping Console 4.1.1™ (A), Partek® Genomics Suite™ (P), PennCNV (p), and Golden Helix® SVS Suite 7.0™ (G) were used.

### Calling and Merging of CNVs for Individual Genomes

The Hap Map 270 6.0 Array reference was utilized as a model reference file. Variants were identified as those DNA regions, which were called as copy number state 0, 1, 3, or 4+ by 10 or more consecutive markers on the chip. Recent literature suggests that a baseline of at least seven consecutive probes is necessary for reliable CNV detection (Wineinger & Tiwari, 2012). In addition, only variants that were greater than 1 kb in size were classified as CNVs for the purposes of this study and only those identified by at least three software programs in the same individual were included in subsequent analysis. We used quantile normalization across all four software programs. Overlapping genes were identified using the UCSC (University of California, Santa Cruz) genome browser table view (NCBI36/hg18). Identification of CNVs was followed by merging of CNV calls within software programs and comparison of calls between software programs to identify CNVs identified by three or more software programs. We identified CNVs within each of the four software programs that were likely to be the same event using the following criteria: (1) CNVs had to be adjacent on the same chromosome (no other CNV call between them); (2) had to share the same gain/loss status; (3) adjacent calls were merged together into one single call, using gap ≤20% of total length. That is, if there are three genomic segments, A, B, and C, where A and C are both losses, we divided the length of gap B by the length of A + B + C, and if this fraction is ≤20%, then we merged A + B + C as a single CNV call.

### Identification of Common CNV Calls for Individual Genomes

We utilized a reciprocal overlap (RO) formula. Copy number variants that shared 50% or more similarity with one another were classified as common. This is consistent with the definition of an overlapping CNV identified by other groups (Pang et al., 2010; Wain et al., 2009; Yavas et al., 2009). In other words, if at least half of the first CNV overlapped with the second CNV and vice versa they were considered to be the same event.

### Comparison of CNV Calls Within Monozygotic Twin Pairs

The same RO definition was used to compare calls between monozygotic twin pairs. Copy number variants were compared between affected and unaffected twin pairs to determine which CNVs were shared and unshared between twins. Unshared CNVs between twins were then annotated with gene information and compared to CNVs in the Database of Genomic Variants (DGV). The genes overlapping CNVs that were different between twin pairs and called by at least three software programs were further characterized using Ingenuity Pathway Analysis (IPA; Ingenuity Systems, California) and GeneMania (Toronto, ON) to identify gene networks and canonical pathways. Finally, we compared the genes identified in this study to those genes listed in the Schizophrenia Gene Database (http://www.schizophreniaforum.org/res/sczgene/default.asp) to determine the genes that appear to be most likely to play a role in schizophrenia. Additional searches from PubMed covering the most recent results helped update any connection between genes of interest and disease pathology.

### Confirmation of Unique CNVs

Differences between monozygotic twins were confirmed using TaqMan Quantitative PCR (qPCR) Copy Number Assays from Life Technologies. The control used for comparison of copy number in the TaqMan experiments was RNAse P and the calibrator used was the individual's unaffected co-twin.

## Results

First, the number of unfiltered CNVs identified by each software program in each of the six twin pairs varied by program (Supplementary Table 1). Golden Helix's® SVS™ identified the highest number of CNVs in each individual, with a range of 168 to 209 variants. Affymetrix® Genotyping Console™ and Partek® Genomics Suite™ called similar numbers of CNVs in each individual, with ranges of 41 to 61 and 37 to 72 variants, respectively. PennCNV called the smallest number of CNVs in each individual, with a range of 21 to 53 variants. The smallest number of unfiltered variants called in any one individual was 21 (PennCNV) and
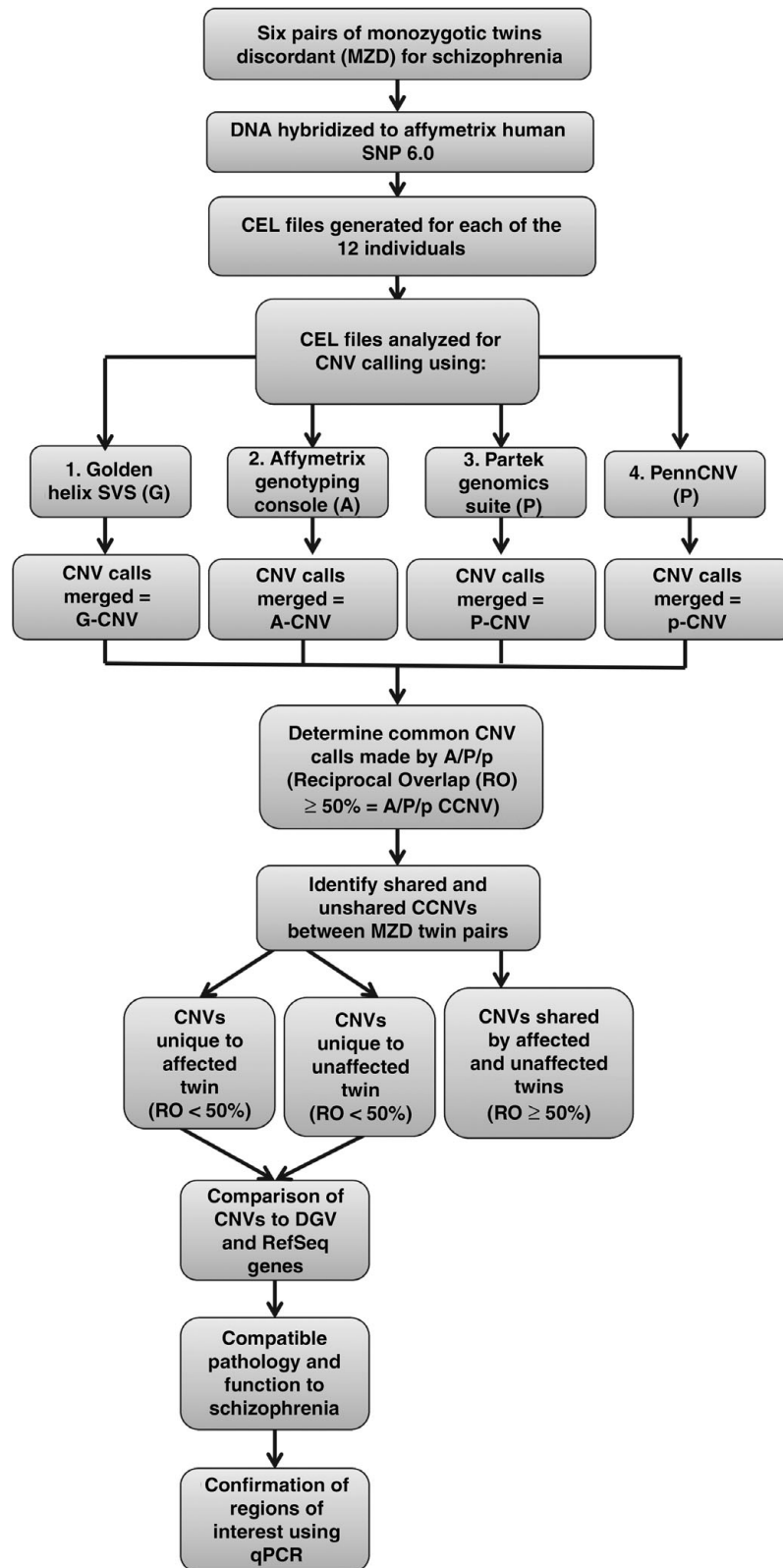
**FIGURE 1**

Flowchart of experimental design CCNV: Consistent Copy Number Variant.

Christina A. Castellani, Zain Awamleh, Melkaye G. Melka, Richard L. O'Reilly and Shiva M. Singh

**TABLE 1**

CNVs Identified by Affymetrix Genotyping Console, Partek Genomics Suite, and PennCNV (A/P/p), as Only Present in Affected Co-Twin

| Sample ID | Chr[a] | Region (build 36) | Start position | End position | Length (kb) | State | Number of Probes | Genes | DGV[b] |
|---|---|---|---|---|---|---|---|---|---|
| 1A | 6 | 6q14.1 | 77016905 | 77027694 | 11 | Loss | 33 | None | No |
|  | 7 | 7q31.1 | 109441200 | 109453946 | 13 | Loss | 13 | None | No |
|  | 14 | 14q21.2 | 41628351 | 41657337 | 29 | Loss | 11 | None | No |
|  | 17 | 17q21.31 | 39423041 | 39430053 | 7 | Loss | 17 | PYY | No |
| 2A | 11 | 11p15.1 | 18949072 | 18961778 | 13 | Gain | 55 | None | Yes |
| 3A | 3 | 3p11.2-3p11.1 | 89394600 | 89419369 | 25 | Loss | 51 | EPHA3 | No |
|  | 7 | 7q11.21 | 64594329 | 64955220 | 361 | Loss | 174 | LOC441242,INTS4L2, CCT6P1,SNORA22 | Yes |
|  | 11 | 11p15.4 | 5789589 | 5809449 | 20 | Loss | 30 | OR52N2 | Yes |
|  | 16 | 16p11.1 | 34459037 | 34624994 | 166 | Gain | 128 | LOC283914,LOC146481, LOC100130700 | Yes |
| 4A | 2 | 2q11.2 | 98858308 | 98879625 | 21 | Loss | 52 | KIAA1211L | No |
|  | 3 | 3q25.1 | 151511085 | 151547185 | 36 | Loss | 30 | None | No |
|  | 16 | 16q23.2 | 78372440 | 78377393 | 5 | Loss | 11 | None | No |
| 5A | 16 | 16p12.3 | 19945650 | 19965863 | 20 | Loss | 30 | GPR139 | No |
|  | 17 | 17q21.31 | 39423041 | 39430053 | 7 | Loss | 17 | PYY | No |
| 6A |  |  |  | NONE |  |  |  |  |  |

Note: [a]Chr = chromosome; [b]DGV = database of genomic variants.

the largest number was 209 (Golden Helix® SVS™). The majority of CNV calls across all four software programs were between 1 kb and 100 kb in size, with at least 56% of the total CNVs in each analysis falling into this range. Partek® Genomics Suite™ yielded the highest number of large calls (>10 Mb) and Golden Helix® SVS™ yielded the largest number of small CNV calls (1–100 kb).

Second, the percentage of Affymetrix® Genotyping Console™ CNV calls detected by three or more algorithms was 23.04%, the percentage of Partek® CNV calls detected by three or more algorithms was 30.30%, the percentage of PennCNV calls detected by three or more algorithms was 33.02% and the percentage of SVS™ CNV calls detected by three or more algorithms was only 0.36%. This summary of overlapping CNVs across four programs strongly suggests that our most reliable calls may represent overlapping CNVs involving Affymetrix® (A), Partek® (P), and PennCNV (p), termed as A/P/p. A combination that includes Golden Helix® along with any other two methods yielded rare CNVs only and was considered too restrictive. More importantly, the overlap generated by A/P/p calls was less restrictive across the twin pairs. Consequently, the CNVs identified by this combination were further assessed in the follow-up analysis involving shared and unshared CNVs between members of the six twin pairs studied (Supplementary Table 2).

Third, we found a total of 38 CNV events called by the three (A/P/p) software programs that were not shared with their co-twin across the six pairs. Specifically, 14 unique CNV events were observed in co-twins affected with schizophrenia (Table 1), while 22 were unique to the six unaffected co-twins of the six MZD pairs (Table 2). Some of the CNVs in both categories contained genes, while others were located in non-coding regions of the genome. In fact, there were a total of 12 unique genes overlapping the

15 CNVs that were found in affected members only. Similarly, there were a total of 28 unique genes overlapping the 23 CNVs that were found in unaffected twins only. The results confirm that monozygotic twins do differ for rare CNVs. They allow us to undertake pair-specific analysis in an effort to explain the discordance of the monozygotic twin pairs for schizophrenia, described below.

**Twin Pair 1**

The affected male patient in twin pair 1 was diagnosed with a psychotic disorder at age 19. He had 41, 53, and 36 raw CNV calls by each of the three (A, P, or p) programs, respectively. In comparison, his unaffected co-twin had 50, 72, and 41 raw CNV calls by the three methods. After CNV merging and discarding of CNVs that were not called by the combination of A/P/p, the affected and unaffected members of this twin pair yielded 14 and 12 CNVs, respectively. These CNVs fell into three categories; shared between the twin pair (10), unique to normal (2), and unique to the affected member of twin pair 1 (4). The two CNVs that were found to be unique to the unaffected twin were a loss at 2q22.1 and a gain at 7q35. These CNVs overlapped the genes *THSD7B* (Entrez Gene: 80731) and *TPK1* (Entrez Gene: 27010), respectively (Table 2). Of the four CNVs that were found to be unique to the affected member, none are reported in the DGV. Three of these CNVs (6q14.1, 7q31.1, 14q21.2) cover no gene overlaps while one (17q21.31) covers the *PYY* (Entrez Gene: 5697) gene (Table 1). Interestingly, the 14q21 region has been previously implicated in bipolar disorder (Liu et al., 2003). Also, *PYY* encodes a protein that has been previously identified to be a potential cerebrospinal fluid marker for mental illness (Widerlov et al., 1988) and autism spectrum disorders (de Krom et al., 2009). This CNV loss for the *PYY* gene has not been previously reported in the

DGV and may be a potential candidate for the discordance of this twin pair for schizophrenia.

### Twin Pair 2

The affected female patient in twin pair 2 was diagnosed with schizoaffective disorder at age 27. She had 48, 66, and 41 raw CNV calls by each of the three (A, P, or p) programs, respectively. In comparison, her unaffected co-twin had 46, 65, and 42 raw CNV calls. After CNV merging and discarding of CNVs that were not called by the combination of A/P/p, 12 and 15 CNVs remained in the affected and unaffected twins, respectively. Of these, 11 were shared between the twin pair, 4 were unique to normal, and 1 was unique to the affected member of this twin pair. The four CNVs that were found to be unique to the unaffected member were a loss at 1q21.1 that overlapped two genes, a loss at 7q21.2 that overlapped no genes, a gain at 17q21.32 that overlapped six genes, and a loss at 19q13.13-19q13.2 that overlapped eight genes (Table 2). The CNV that was found to be unique to the affected member was a gain found in the region 11p15.1 that does not overlap with any known genes and has been previously reported in the DGV (Table 1). No CNVs identified in twin pair 2 have been reported for any neurodevelopmental disorder and the observations do not seem to be likely candidates to explain the discordance for schizophrenia seen in this twin pair.

### Twin Pair 3

The affected female in twin pair 3 was diagnosed with paranoid schizophrenia at age 22. She had 47, 54, and 25 raw CNV calls by each of the three (A, P, or p) programs, respectively. In comparison, her twin sister had 44, 43, and 21 raw CNV calls. The merging and discarding of non-overlapping CNVs yielded 10 and 7 CNVs in the affected and unaffected twins, respectively. Of these, six of the CNVs were shared between twin pair 3, one was unique to normal, and four were unique to the affected member. The CNV that was found to be unique to the unaffected member was a loss at 3q26.1 that did not overlap any gene (Table 2). The four CNVs that were found to be unique to the affected twin overlapped four regions: 3p11.2-3p11.1, 7q11.21, 11p15.4, and 16p11.1 (Table 1). Of these, the loss at 3p11.2-3p11.1 is the only one that was not previously reported in the DGV. This CNV overlapped the *EPHA3* (Entrez Gene: 2042) gene. This gene belongs to the ephrin receptor subfamily of protein-tyrosine kinases that have been implicated in mediating developmental events, particularly in the nervous system, and has been previously associated with neurodegenerative diseases (Martinez & Soriano, 2005). Further, the 16p11 region has been implicated in mental disorders, including psychosis (Steinberg et al., 2014). The other genes unique to the affected twin were *LOC441242* (Entrez Gene: 441242), *INTS4L2* (Entrez Gene: 644619), *CCT6P1* (Entrez Gene: 643253), *SNORA22* (Entrez Gene: 677807), *OR52N2* (Entrez Gene:

390077), *LOC283914* (Entrez Gene: 283914), *LOC146481* (Entrez Gene: 146481), and *LOC100130700* (Entrez Gene: 100130700). Of particular interest is *OR52N2*, an olfactory receptor. Olfactory receptors share a seven-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals (Malnic et al., 2004). Recent reports suggest robust olfactory deficits in schizophrenia patients (Moberg et al., 2013). Yet another identified gene, *SNORA22*, encodes a small nucleolar ribonucleic acid (RNA), which may guide chemical modifications of other RNAs (Kiss, 2001). Interestingly, *CCT6P1* is highly expressed in brain (Velculescu et al., 1995). Also, the *EPHA3* gene is the only CNV identified in this patient that has not been previously reported in DGV and has the potential to explain discordance for schizophrenia in this twin pair.

### Twin Pair 4

The female patient in twin pair 4 was diagnosed with paranoid schizophrenia at age 18. She had 44, 49, and 25 raw CNV calls identified by each of the A, P, or p programs, respectively. In comparison, her unaffected co-twin had 52, 37, and 22 raw CNV calls. After CNV merging and discarding of CNVs that were not called by A/P/p, 10 and 7 CNVs remained in the affected and unaffected members of the twin pair, respectively; seven were shared between the twin pair and three were unique to the affected member. Interestingly, there was no CNV that was unique to the normal twin (Table 2). Of the three CNVs that were found to be unique to the patient, none are reported in the DGV. Two of the CNVs (3q25.1, 16q23.2) cover no genes while one (2q11.2) covers the *KIAA1211L* (Entrez Gene: 343990) gene (Table 1). Interestingly, the 3q25 region has been previously implicated in autism-spectrum disorders (Auranen et al., 2002). Also, *KIAA1211L* is expressed in the brain and has been reported in bipolar disorder (Scott et al., 2009). Also of interest, this CNV has not been previously identified in the DGV, making this CNV loss a potential candidate for the discordance for schizophrenia of this twin pair.

### Twin Pair 5

The affected male in twin pair 5 was diagnosed with undifferentiated schizophrenia at age 20. He had 56, 54, and 45 raw CNV calls and his unaffected co-twin had 54, 58, and 41 such calls identified by the A, P, or p software programs, respectively. After CNV merging and discarding of CNVs that were not called by all three programs, 10 and 18 CNVs remained in the affected and unaffected twins, respectively. These CNVs fell into three categories: shared between the twin pair (8), unique to normal (10), and unique to the affected member of twin pair 5 (2). The 10 CNVs that were found to be unique to the unaffected member were found in the regions 1p36.33, 1p21.1,

Christina A. Castellani, Zain Awamleh, Melkaye G. Melka, Richard L. O'Reilly and Shiva M. Singh

**TABLE 2**

CNVs Identified by Affymetrix Genotyping Console, Partek Genomics Suite, and PennCNV (A/P/p), as Only Present in Unaffected Co-Twin

| Sample ID | Chr[a] | Region (build 36) | Start position | End position | Length (kb) | State | Number of probes | Genes | DGV[b] |
|---|---|---|---|---|---|---|---|---|---|
| 1B | 2 | 2q22.1 | 138045294 | 138065759 | 20 | Loss | 24 | *THSD7B* | No |
| | 7 | 7q35 | 143917601 | 144051219 | 134 | Gain | 31 | *TPK1* | Yes |
| 2B | 1 | 1q21.2 | 149086185 | 149202838 | 117 | Loss | 76 | *ARNT,SETDB1* | Yes |
| | 7 | 7q21.2 | 91031620 | 91040715 | 9 | Loss | 29 | None | Yes |
| | 17 | 17q21.32 | 44535027 | 44739523 | 204 | Gain | 30 | *B4GALNT2,GNGT2,ABI3,PHOSPHO1,FLJ40194,ZNF652* | No |
| | 19 | 19q13.13-19q13.2 | 43294390 | 43538078 | 244 | Loss | 117 | *SIPA1L3,DPF1,PPP1R14A,SPINT2,YIF1B,C19orf33,KCNK6, CATSPERG* | Yes |
| 3B | 3 | 3q26.1 | 165046890 | 165083684 | 37 | Loss | 65 | None | No |
| 4B | | | | | | NONE | | | |
| 5B | 1 | 1p36.33 | 61735 | 86161 | 24 | Loss | 14 | None | Yes |
| | 1 | 1p21.1 | 104155790 | 104248433 | 93 | Loss | 31 | None | Yes |
| | 1 | 1q25.2 | 174796556 | 174801847 | 5 | Loss | 17 | *PAPPA2* | No |
| | 2 | 2p22.3 | 34709689 | 34727855 | 18 | Loss | 21 | None | Yes |
| | 3 | 3p14.1 | 68746577 | 68747401 | 1 | Loss | 11 | None | No |
| | 3 | 3q11.2 | 98944458 | 98949409 | 5 | Loss | 38 | *EPHA6* | No |
| | 4 | 4p15.1 | 34779042 | 34822761 | 44 | Loss | 40 | None | No |
| | 4 | 4q24 | 104742390 | 104761153 | 19 | Loss | 37 | *TACR3* | No |
| | 8 | 8p11.23 | 39301780 | 39384688 | 83 | Loss | 28 | *ADAM5* | Yes |
| | 9 | 9p11.2 | 44247866 | 44705656 | 458 | Loss | 16 | *LOC643648* | Yes |
| 6B | 2 | 2p22.3 | 34709689 | 34727867 | 18 | Loss | 21 | None | Yes |
| | 2 | 2q14.3 | 129638490 | 129640285 | 2 | Loss | 11 | None | Yes |
| | 3 | 3q26.1 | 162551776 | 162619878 | 68 | Gain | 39 | *SPTSSB* | No |
| | 12 | 12p11.1 | 33301406 | 33306843 | 5 | Loss | 10 | None | Yes |
| | 17 | 17q21.32 | 44214888 | 44362186 | 147 | Gain | 89 | *TTLL6, CALCOCO2, ATP5G1,UBE2Z* | Yes |

Note: [a]Chr = chromosome; [b]DGV = database of genomic variants.

1q25.2, 2p22.3, 3p14.1, 3q11.2, 4p15.1, 4q24, 8p11.23, and 9p11.2, and overlapped the *PAPPA2* (Entrez Gene: 60676), *EPHA6* (Entrez Gene: 285220), *TACR3* (Entrez Gene: 6870), *ADAM5* (Entrez Gene: 255926), *LOC643648* (Entrez Gene: 643648), *LOC283914, LOC146481,* and *LOC100130700* genes (Table 2). Of the two CNVs that were found to be unique to the patient (16p12.3 and 17q21.31), the first, a loss at 16p12.3, overlaps the *GPR139* (Entrez Gene: 124274) gene and the second, a loss at 17q21.31, overlaps the *PYY* gene (Table 1). *PYY*, as presented in the pair-specific results shown for twin pair 1 above, is a potential candidate for the discordance for schizophrenia identified in this twin pair as well. *GPR139* is a gene that is an important mediator of signal transduction. G-protein receptors are almost exclusively expressed in brain and are likely to play important roles in the central nervous system. *GPR139* has been previously reported to be associated with attention deficit hyperactivity disorder (ADHD; OMIM: 143465; Ebejer et al., 2013). Both the CNV loss overlapping the *GPR139* gene and the CNV loss overlapping the *PYY* gene have not been listed in the DGV. They may represent de novo events and candidates for the disease discordance of this twin pair.

### Twin Pair 6

The male patient in twin pair 6 was diagnosed with paranoid schizophrenia at age 16. He had 61, 47, and 50 raw CNV calls as identified by the A, P, or p software programs, respectively. In comparison, his unaffected co-twin had 53, 48, and 53 raw CNV calls. After CNV merging and discarding of CNVs that were not called by all three (A/P/p) programs, 8 and 13 CNVs remained in the affected and unaffected members, respectively. These CNVs fell into three categories: shared between the twin pair (8), unique to normal (5), and unique to the affected member of twin pair 6 (0). The five CNVs that were found to be unique to the unaffected member were found at 2p22.3, 2q21.1, 3q26.1, 12p11.1, and 17q21.32. Only the CNV gains at 3q26.1 and 17q21.32 overlapped genes. The CNV at 3q26.1 overlapped the *SPTSSB* (Entrez Gene: 165679) gene and the CNV at 17q21.32 overlapped *TTLL6* (Entrez Gene: 284076), *CALCOCO2* (Entrez Gene: 10241), *ATP5G1* (Entrez Gene: 516), and *UBE2Z* (Entrez Gene: 65264) (Table 2). No CNVs were found to be unique to the affected member of this twin pair. Consequently, no CNVs identified in twin pair 6 seem to represent candidates to explain their discordance for schizophrenia.

The results outlined above have identified rare and pair-specific CNV differences between monozygotic twins in each of the six twin pairs discordant for schizophrenia, studied. Some CNVs involve single or multiple genes and others represent non-coding genomic regions. Also, a number of these are not reported in DGV, specifically 10 of 14 events seen uniquely in affected twins and 9 of 22 events seen uniquely in unaffected members.

## Discussion

It has become apparent that CNVs are common in human populations and play a significant role in the etiology of complex diseases, including schizophrenia (Ahn et al., 2013; St Clair, 2013). However, it is not easy to identify disease-specific CNVs and establish their mode of action in the causation of the disease. Of special concern is the use of arrays with different degrees of genome coverage and the large number of algorithms available to call CNVs. Although the Affymetrix Human Array 6.0 appears to meet most of the platform criteria, including coverage for CNV calling in humans, a gold standard algorithm for the analysis of data has not been established (Zhang et al., 2011). There is a likelihood of false positive results. Despite this, such experiments have generated and continue to generate valuable insights. Reports assessing the use of different software algorithms to analyze the same microarrays have identified a low concordance rate between software programs (Kim et al., 2012; Pinto et al., 2011). This is likely due to the substantial background noise, which contributes to a false discovery rate of variants (Grayson & Aune, 2011). To avoid this, often two programs are used to call for the CNVs and the resulting shared CNVs are considered to be reliable. Although logical, this approach is not totally satisfactory as it may ignore and miss out on some critical results. In this analysis we have focused on more reliable results and used four different software programs to call for CNVs. We found a low percentage of concordance between these calls. This is consistent with findings in the literature (Kim et al., 2012; Pinto et al., 2011) and highlights the necessity for more stringent guidelines for CNV calling from microarrays. A study conducted by Kim et al. (2012) suggested that at least three calling algorithms should be used to ensure the reliability of results.

As stated, we used four CNV calling programs (Golden Helix's® SVS™, Affymetrix® Genotyping Console™, Partek® Genomics Suite™, and PennCNV) and selected CNVs that were called by three methods (Affymetrix® Genotyping Console™, Partek® Genomics Suite™, and PennCNV) referred to as A/P/p. Also, we chose 10 CNVs to confirm the results by Real Time PCR. Our qPCR results established that one CNV, a CNV loss at 7q11.21 in twin pair 3 was significantly different between twins (Figure 2). Further, two CNVs (CNV loss at *PYY* in twin pair 5 and CNV gain at 5q11.2 in twin pair 6) showed the expected trend but were not statistically significant (Table 3). This suggests that the experimental confirmation by qPCR of the CNV calls is at least 10%.

Yet another challenge with genetic studies in schizophrenia is the extensive heterogeneity that may include multiple genetic, epigenetic, and environmental factors (Singh & O'Reilly, 2009; van Dongen & Boomsma, 2013). Two observations are of particular relevance to this discussion. First, a number of genomic regions and genes including CNVs,

Christina A. Castellani, Zain Awamleh, Melkaye G. Melka, Richard L. O'Reilly and Shiva M. Singh
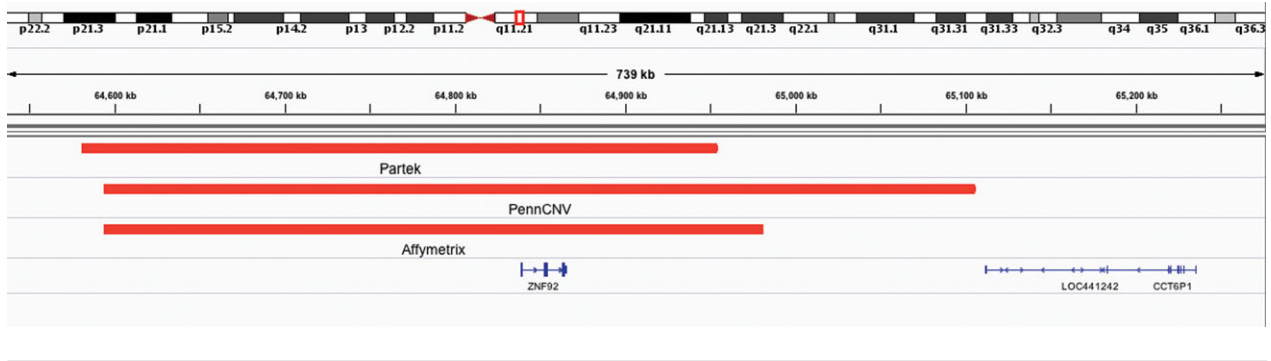
**FIGURE 2**

(Colour online) Copy number variation loss at 7q11.21 in twin 3A.

**TABLE 3**

**TaqMan Real Time PCR Results**

| Gene/region | Individual | qPCR result |
| --- | --- | --- |
| EPHA3 | 3A | 2.29 |
| GPR139 | 5A | 2.04 |
| OR52N2 | 3A | 1.94 |
| CACNB2 | 1B | 1.87 |
| ST8S1A6 | 1B | 2.02 |
| PYY | 1A | 1.99 |
| | 5A | 1.89 |
| KIAA1211L | 4A | 2.07 |
| 5q11.2 | 6A | 2.23 |
| 7q11.21 | 3A | 1.47* |

Note: *a significant difference between twin pair 3 in qPCR confirmation.

both inherited (relatively common) and de novo (extremely rare), have been implicated in this complex neurodevelopmental disease (Kirov, 2010; Maiti et al., 2011; Van Den Bossche et al., 2013). Also, there is less than 100% concordance (48%) between monozygotic twins (McGuffin et al., 1994). Consequently, the genome of the normal twin may provide a near perfect match to the genome of the affected member. Also, in some cases the discordance of monozygotic twins for schizophrenia could be attributed to differences in their CNVs potentially caused by DNM (Maiti et al., 2011; Singh et al., 2009). The published results suggest that DNM are not limited to the germ lines alone. Rather they are ongoing throughout life, including stages of differentiation, development, and aging (Lupski, 2010). The occurrence of DNMs has now been demonstrated using a variety of strategies including MZ twins (Bruder et al., 2008; Singh et al., 2009), trios (Vissers et al., 2010), and MZ twins compared with both parents (Maiti et al., 2011). Quantitatively, Maiti et al. (2011) identified one and two DNMs in two pairs of MZ twins respectively, based on parental genotypes, while Vissers et al. (2010) identified one to two DNMs per trio in eight trios with a mentally retarded proband using family-based exome sequencing. The results suggest that DNMs can account for phenotypic discordance between MZ pairs. Also, the degree of difference may vary from pair to pair.

Their phenotypic impact will depend not only on the genomic region involved but also on the background genotype, and the timing of DNMs during ontogeny and the mechanism that is responsible for genomic discordance of MZ twins may generate mosaics, with or without significant phenotypic manifestation (Ruderfer et al., 2013).

Our results on six MZD pairs show that each MZD pair differs in rare CNVs. Also, the discordance of some of the pairs could be attributed to CNVs identified in this analysis. For this, we have used the following criteria. First, the CNV should be present in the affected member(s) of the twin pair only. Second, the genomic region involved must encode for gene(s). Third, the gene must be expressed in the brain and/or the gene must have relevance to the neurodevelopment and physiological outcomes associated with schizophrenia. Furthermore, the CNV of interest must not have been identified previously in normal healthy individuals (DGV). The use of these criteria has allowed us to identify potential causes for schizophrenia in four of the six pairs studied. The four pairs that do meet our criteria have their own twin-pair-specific CNV patterns. Given extensive heterogeneity and the rare nature of de novo events, these patterns are expected to be variable. Not surprisingly, the observed differences are pair-specific with respect to the genomic region(s) and gene(s) involved. Although the results are patient and pair-specific, we did find some genomic regions and genes that are common across unrelated patients. For example, the region 16p11 was identified as uniquely disrupted in the affected member of twin pair 3 — this is particularly interesting as this region has been previously associated with psychosis (Steinberg et al., 2014). Another example of a region identified in more than one sample was 2p22.3, the same CNV was uniquely identified in this region in the unaffected member of twin pair 5 and twin pair 6 — this may suggest a possible protective or mediating effect on the disease from a copy number variation loss in this region.

In conclusion, the MZD twin-based genomic (CNV) strategy to identify candidate genes that may be involved in schizophrenia is logical and practical. It has the potential

to serve as an effective strategy in identification of genes and genetic mechanisms that may cause complex disorders. Specifically, individuals with schizophrenia have CNV gains and losses that are likely to contribute to the disease. Here, the inherited mutations may provide predisposition that may not be sufficient for disease manifestation. Occurrence of any/some additional de novo event (CNV or mutation) may add to this predisposition and manifest the disease. This two-hit model (Maynard et al., 2001; Singh et al., 2004) of disease development may explain a number of observations on schizophrenia. First, in most cases of discordance in monozygotic twins, even the normal twin may have some or delayed manifestation of some or all symptoms. Second, in most cases of familial schizophrenia, the second hit may or may not be needed depending on the nature of familial predisposition. Also, in some cases environmental components may add to the predisposition or be enough by itself to affect neurodevelopment and result in the disease. What is needed is precise and reliable results that are not always forthcoming. This challenge is apparent from our results where only one of the 10 CNVs could be statistically confirmed, while two CNVs showed the expected trend but failed to reach the level of significance in qPCR. This follows a number of recent reports (Ehli et al., 2012). Despite such limitations, the MZD strategy outlined appears realistic. Specifically, the use of strict criteria for the assessment of copy number variations in monozygotic twin genomes discordant for schizophrenia has identified a novel CNV (7q11.21) that is surrounded by low copy repeats with the potential to undergo mechanisms which generate CNVs de novo. This confirmed CNV was seen exclusively in the affected patient of twin pair 3 and deserves further investigation as a candidate region for schizophrenia and related disorders in this twin pair and beyond.

## Supplementary Material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/thg.2014.6.

## Acknowledgments

## References

Ahn, K., Gotay, N., Andersen, T. M., Anvari, A. A., Gochman, P., Lee, Y., . . . Rapoport, J. L. (2013). High rate of disease-related copy number variations in childhood onset schizophrenia. *Molecular Psychiatry*. Advance online publication.

Auranen, M., Vanhala, R., Varilo, T., Ayers, K., Kempas, E., Ylisaukko-Oja, T., . . . Jarvela, I. (2002). A genomewide screen for autism-spectrum disorders: Evidence for a major susceptibility locus on chromosome 3q25–27. *American Journal of Human Genetics*, *71*, 777–790.

Autism Genome Project Consortium, Szatmari, P., Paterson, A. D., Zwaigenbaum, L., Roberts, W., Brian, J., . . . Meyer, K. J. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genetics*, *39*, 319–328.

Beckmann, H., & Franzek, E. (2000). The genetic heterogeneity of 'schizophrenia'. *The World Journal of Biological Psychiatry*, *1*, 35–41.

Bloom, R. J., Kahler, A. K., Collins, A. L., Chen, G., Cannon, T. D., Hultman, C., & Sullivan, P. F. (2013). Comprehensive analysis of copy number variation in monozygotic twins discordant for bipolar disorder or schizophrenia. *Schizophrenia Research*, *146*, 289–290.

Breckpot, J., Thienpont, B., Gewillig, M., Allegaert, K., Vermeesch, J. R., & Devriendt, K. (2012). Differences in copy number variation between discordant monozygotic twins as a model for exploring chromosomal mosaicism in congenital heart defects. *Molecular Syndromology*, *2*, 81–87.

Bruder, C., Piotrowski, A., Gijsbers, A. A., Andersson, R., Erickson, S., Diaz de Ståhl, T., . . . Dumanski, J. P. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *The American Journal of Human Genetics*, *82*, 763–771.

Buchanan, J. A., & Scherer, S. W. (2008). Contemplating effects of genomic structural variation. *Genetics in Medicine*, *10*, 639–647.

Conrad, D. F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C., . . . Hurles, M. E. (2010). Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nature Genetics*, *42*, 385–391.

de Krom, M., Staal, W. G., Ophoff, R. A., Hendriks, J., Buitelaar, J., Franke, B., . . . van Ree, J. M. (2009). A common variant in DRD3 receptor is associated with autism spectrum disorder. *Biological Psychiatry*, *65*(7), 625–630.

Ebejer, J. L., Duffy, D. L., van der Werf, J., Wright, M. J., Montgomery, G., Gillespie, N. A., . . . Medland, S. E. (2013). Genome-wide association study of inattention and hyperactivity-impulsivity measured as quantitative traits. *Twin Research and Human Genetics*, *16*, 560–574.

Ehli, E. A., Abdellaoui, A., Hu, Y., Hottenga, J. J., Kattenberg, M., van Beijsterveldt, T., . . . Davies, G. E. (2012). De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on attention problems. *European Journal of Human Genetics*, *20*, 1037–1043.

Feuk, L., Marshall, C. R., Wintle, R. F., & Scherer, S. W. (2006). Structural variants: Changing the landscape of chromosomes and design of disease studies. *Human Molecular Genetics*, *15*, R57–R66.

First, M., Gibbon, M., Spitzer, R., Williams, J., & Benjamin, L. (Eds.). (1997). *Structured clinical interview for DSM-IV axis II personality disorders (SCID-II)*. Washington, DC: American Psychiatric Press.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1996). *Clinical interview for DSM-IV axis I disorders: Clinician version (SCID-CV)*. Washington, DC: American Psychiatric Press.

Friedman, J. I., Vrijenhoek, T., Markx, S., Janssen, I. M., van der Vliet, W. A., Faas, B. H., . . . Veltman, J. A. (2008). CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. *Molecular Psychiatry*, 13, 261–266.

Gladkevich, A., Kauffman, H. F., & Korf, J. (2004). Lymphocytes as a neural probe: Potential for studying psychiatric disorders. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 28, 559–576.

Glessner, J. T., Reilly, M. P., Kim, C. E., Takahashi, N., Albano, A., Hou, C., . . . Hakonarson, H. (2010). Strong synaptic transmission impact by copy number variations in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 10584–10589.

Glessner, J. T., Wang, K., Cai, G., Korvatska, O., Kim, C. E., Wood, S., . . . Hakonarson, H. (2009). Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, 459, 569–573.

Grayson, B. L., & Aune, T. M. (2011). A comparison of genomic copy number calls by partek genomics suite, genotyping console and birdsuite algorithms to quantitative PCR. *BioData Mining*, 4, 8.

Halder, A., Jain, M., Chaudhary, I., & Varma, B. (2012). Chromosome 22q11.2 microdeletion in monozygotic twins with discordant phenotype and deletion size. *Molecular Cytogenetics*, 5, 13.

Hamilton, S. P. (2008). Schizophrenia candidate genes: Are we really coming up blank? *The American Journal of Psychiatry*, 165, 420–423.

Henrichsen, C. N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F., . . . Reymond, A. (2009). Segmental copy number variation shapes tissue transcriptomes. *Nature Genetics*, 41, 424–429.

Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., . . . Lee, C. (2004). Detection of large-scale variation in the human genome. *Nature Genetics*, 36, 949–951.

Karayiorgou, M., & Gogos, J. A. (2006). Schizophrenia genetics: Uncovering positional candidate genes. *European Journal of Human Genetics*, 14, 512–519.

Kim, S. Y., Kim, J. H., & Chung, Y. J. (2012). Effect of combining multiple CNV defining algorithms on the reliability of CNV calls from SNP genotyping data. *Genomics & Informatics*, 10, 194–199.

Kirov, G. (2010). The role of copy number variation in schizophrenia. *Expert Review of Neurotherapeutics*, 10, 25–32.

Kirov, G., Gumus, D., Chen, W., Norton, N., Georgieva, L., Sari, M., . . . Ullmann, R. (2008). Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. *Human Molecular Genetics*, 17, 458–465.

Kiss, T. (2001). Small nucleolar RNA-guided posttranscriptional modification of cellular RNAs. *The EMBO Journal*, 20, 3617–3622.

Lee, J. A., & Lupski, J. R. (2006). Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, 52, 103–121.

Lee, C., & Scherer, S. W. (2010). The clinical context of copy number variation in the human genome. *Expert Reviews in Molecular Medicine*, 12, e8.

Liu, H., Abecasis, G. R., Heath, S. C., Knowles, A., Demars, S., Chen, Y. J., . . . Karayiorgou, M. (2002). Genetic variation in the 22q11 locus and susceptibility to schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 16859–16864.

Liu, J., Juo, S. H., Dewan, A., Grunn, A., Tong, X., Brito, M., . . . Baron, M. (2003). Evidence for a putative bipolar disorder locus on 2p13-16 and other potential loci on 4q31, 7q34, 8q13, 9q31, 10q21-24, 13q32, 14q21 and 17q11-12. *Molecular Psychiatry*, 8, 333–342.

Lupski, J. R. (2010). New mutations and intellectual function. *Nature Genetics*, 42, 1036–1038.

Maiti, S., Kumar, K. H., Castellani, C. A., O'Reilly, R., & Singh, S. M. (2011). Ontogenetic de novo copy number variations (CNVs) as a source of genetic individuality: Studies on two families with MZD twins for schizophrenia. *PloS One*, 6(3), e17125.

Malnic, B., Godfrey, P. A., & Buck, L. B. (2004). The human olfactory receptor gene family. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2584–2589.

Martinez, A., & Soriano, E. (2005). Functions of ephrin/eph interactions in the development of the nervous system: Emphasis on the hippocampal system. *Brain Research Reviews*, 49, 211–226.

Maynard, T. M., Sikich, L., Lieberman, J. A., & LaMantia, A. S. (2001). Neural development, cell-cell signaling, and the 'two-hit' hypothesis of schizophrenia. *Schizophrenia Bulletin*, 27, 457–476.

McCarroll, S. A., & Altshuler, D. M. (2007). Copy-number variation and association studies of human disease. *Nature Genetics*, 39(Suppl.), S37–S42.

McGuffin, P., Asherson, P., Owen, M., & Farmer, A. (1994). The strength of the genetic effect. Is there room for an environmental influence in the aetiology of schizophrenia? *The British Journal of Psychiatry*, 164, 593–599.

McInnes, L. A., & Lauriat, T. L. (2006). RNA metabolism and dysmyelination in schizophrenia. *Neuroscience and Biobehavioral Reviews*, 30, 551–561.

Miyake, K., Yang, C., Minakuchi, Y., Ohori, K., Soutome, M., Hirasawa, T., . . . Kubota, T. (2013). Comparison of genomic and epigenomic expression in monozygotic twins discordant for Rett syndrome. *PloS One*, 8(6), e66729.

Moberg, P. J., Kamath, V., Marchetto, D. M., Calkins, M. E., Doty, R. L., Hahn, C. G., . . . Turetsky, B. I. (2013). Meta-analysis of olfactory function in schizophrenia, first-degree family members, and youths at-risk for psychosis. *Schizophrenia Bulletin*, doi:10.1093/schbul/sbt049.

Moessner, R., Marshall, C. R., Sutcliffe, J. S., Skaug, J., Pinto, D., Vincent, J., . . . Scherer, S. W. (2007). Contribution of

SHANK3 mutations to autism spectrum disorder. *American Journal of Human Genetics*, 81, 1289–1297.

Need, A. C., Ge, D., Weale, M. E., Maia, J., Feng, S., Heinzen, E. L., . . . Goldstein, D. B. (2009). A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genetics*, 5(2), e1000373.

O'Reilly, R. L., & Singh, S. M. (1996). Retroviruses and schizophrenia revisited. *American Journal of Medical Genetics*, 67, 19–24.

Pang, A. W., MacDonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., . . . Scherer, S. W. (2010). Towards a comprehensive structural variation map of an individual human genome. *Genome Biology*, 11, R52.

Pinto, D., Darvishi, K., Shi, X., Rajan, D., Rigler, D., Fitzgerald, T., . . . Feuk, L. (2011). Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nature Biotechnology*, 29, 512–520.

Public Health Agency of Canada Steering Committee on Health — Adjusted Life Expectancy. (2013). Report summary — Health-adjusted life expectancy in Canada: 2012 report by the Public Health Agency of Canada. *Chronic Diseases and Injuries in Canada*, 33(2), 103.

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kahler, A. K., Akterin, S., . . . Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, 45, 1150–1159.

Ruderfer, D. M., Chambert, K., Moran, J., Talkowski, M., Chen, E. S., Gigek, C., . . . Ernst, C. (2013). Mosaic copy number variation in schizophrenia. *European Journal of Human Genetics*. Advance online publication.

Scott, L. J., Muglia, P., Kong, X. Q., Guan, W., Flickinger, M., Upmanyu, R., . . . Boehnke, M. (2009). Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 7501–7506.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., . . . Wigler, M. (2007). Strong association of de novo copy number mutations with autism. *Science*, 316, 445–449.

Singh, S. M., Castellani, C. A., & O'Reilly, R. L. (2009). Copy number variation showers in schizophrenia: An emerging hypothesis. *Molecular Psychiatry*, 14, 356–358.

Singh, S. M., McDonald, P., Murphy, B., & O'Reilly, R. (2004). Incidental neurodevelopmental episodes in the etiology of schizophrenia: An expanded model involving epigenetics and development. *Clinical Genetics*, 65, 435–440.

Singh, S. M., Murphy, B., & O'Reilly, R. (2002). Monozygotic twins with chromosome 22q11 deletion and discordant phenotypes: Updates with an epigenetic hypothesis. *Journal of Medical Genetics*, 39, e71.

Singh, S. M., & O'Reilly, R. (2009). (Epi)genomics and neurodevelopment in schizophrenia: Monozygotic twins discordant for schizophrenia augment the search for disease-related (epi)genomic alterations. *Genome*, 52, 8–19.

Stankiewicz, P., & Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annual Review of Medicine*, 61, 437–455.

St Clair, D. (2013). Structural and copy number variants in the human genome: Implications for psychiatry. *The British Journal of Psychiatry*, 202, 5–6.

Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O. P., Ingason, A., Steinberg, S., . . . Stefansson, K. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature*, 455, 232–236.

Steinberg, S., de Jong, S., Mattheisen, M., Costas, J., Demontis, D., Jamain, S., . . . Stefansson, K. (2014). Common variant at 16p11.2 conferring risk of psychosis. *Molecular Psychiatry*, 19, 108–114.

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., . . . Dermitzakis, E. T. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, 315, 848–853.

Sullivan, P. F. (2005). The genetics of schizophrenia. *PLoS Medicine*, 2(7), e212.

Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. *Archives of General Psychiatry*, 60, 1187–1192.

Torkamani, A., Dean, B., Schork, N. J., & Thomas, E. A. (2010). Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Research*, 20, 403–412.

Torrey, E. F., Miller, J., Rawlings, R., & Yolken, R. H. (1997). Seasonality of births in schizophrenia and bipolar disorder: A review of the literature. *Schizophrenia Research*, 28, 1–38.

Tsuang, M. T., Nossova, N., Yager, T., Tsuang, M. M., Guo, S. C., Shyu, K. G., . . . Liew, C. C. (2005). Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: A preliminary report. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 133, 1–5.

Van Den Bossche, M. J., Strazisar, M., Cammaerts, S., Liekens, A. M., Vandeweyer, G., Depreeuw, V., . . . Del-Favero, J. (2013). Identification of rare copy number variants in high burden schizophrenia families. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 162B, 273–282.

van Dongen, J., & Boomsma, D. I. (2013). The evolutionary paradox and the missing heritability of schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 162B, 122–136.

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, 270, 484–487.

Verveer, J., Huizer, K., Fekkes, D., & van Beveren, N. J. (2007). Gene-expression profiling in schizophrenia: An overview. [Genexpressieprofilering bij schizofrenie: Een overzicht] *Tijdschrift Voor Psychiatrie*, 49, 7–19.

Vissers, L. E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., . . . Veltman, J. A. (2010). A de novo paradigm for mental retardation. *Nature Genetics*, 42, 1109–1112.

Wain, L. V., Pedroso, I., Landers, J. E., Breen, G., Shaw, C. E., Leigh, P. N., . . . Al-Chalabi, A. (2009). The role of copy number variation in susceptibility to amyotrophic lateral

sclerosis: Genome-wide association study and comparison with published loci. *PloS One*, 4, e8175.

Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., . . . Sebat, J. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, 20, 539–543.

Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J. T., Abrahams, B. S., . . . Hakonarson, H. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*, 459, 528–533.

Wellcome Trust Case Control Consortium, Craddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., Plagnol, V., . . . Donnelly, P. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, 464, 713–720.

Widerlov, E., Lindstrom, L. H., Wahlestedt, C., & Ekman, R. (1988). Neuropeptide Y and peptide YY as possible cerebrospinal fluid markers for major depression and schizophrenia, respectively. *Journal of Psychiatric Research*, 22, 69–79.

Wineinger, N. E., & Tiwari, H. K. (2012). The impact of errors in copy number variation detection algorithms on association results. *PloS One*, 7(4), e32396.

Xu, B., Roos, J. L., Levy, S., van Rensburg, E. J., Gogos, J. A., & Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. *Nature Genetics*, 40, 880–885.

Yavas, G., Koyuturk, M., Ozsoyoglu, M., Gould, M. P., & LaFramboise, T. (2009). An optimization framework for unsupervised identification of rare copy number variation from SNP array data. *Genome Biology*, 10, R119.

Zhang, D., Qian, Y., Akula, N., Alliey-Rodriguez, N., Tang, J., Bipolar Genome Study, . . . Liu, C. (2011). Accuracy of CNV detection from GWAS data. *PloS One*, 6, 2034–2043.

Zogopoulos, G., Ha, K. C., Naqib, F., Moore, S., Kim, H., Montpetit, A., . . . Gallinger, S. (2007). Germline DNA copy number variation frequencies in a large North American population. *Human Genetics*, 122, 345–353.