Published in final edited form as:

J Child Neurol. 2015 December ; 30(14): 1947–1953. doi:10.1177/0883073815602066.

The Influence of Microdeletions and Microduplications of 16p11.2 on Global Transcription Profiles

Mary Kusenda, PhD[1,6,*], **Vladimir Vacic, PhD**[1], **Dheeraj Malhotra, PhD**[1,4], **Linda Rodgers**[1], **Kevin Pavon, BS**[1], **Jennifer Meth, BS**[1], **Ravinesh A. Kumar, PhD**[2], **Susan L. Christian, PhD**[2], **Hilde Peeters, PhD**[3], **Shawn S. Cho, BS**[4], **Anjene Addington, MPH, PhD**[5], **Judith L. Rapoport, MD**[5], and **Jonathan Sebat, PhD**[4]

[1]Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA [2]Department of Human Genetics, University of Chicago, Chicago, IL, USA [3]Laboratory for Genetics of Human Development, Department of Human Genetics, Faculty of Medicine, Katholieke Universiteit Leuven, Leuven, Netherlands [4]Beyster Center for Genomics of Psychiatric Diseases, Department of Psychiatry, and Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA [5]Child Psychiatry Branch, National Institute of Mental Health, Bethesda, MD, USA [6]Department of Biology, Chemistry and Environmental Studies, Molloy College, Rockville Centre, New York 11571, USA

Abstract

Copy number variants (CNVs) of a 600 kb region on 16p11.2 are associated with neurodevelopmental disorders and changes in brain volume. The authors hypothesize that abnormal brain development associated with this CNV can be attributed to changes in transcriptional regulation. The authors determined the effects of 16p11.2 dosage on gene expression by transcription profiling of lymphoblast cell lines derived from 6 microdeletion carriers, 15 microduplication carriers and 15 controls. Gene dosage had a significant influence on the transcript abundance of a majority (20/34) of genes within the CNV region. In addition, a limited number of genes were dysregulated *in trans*. Genes most strongly correlated with patient head circumference included *SULT1A*, *KCTD13*, and *TMEM242*. Given the modest effect of 16p11.2 copy number on global transcriptional regulation in lymphocytes, larger studies utilizing

Corresponding Author: Jonathan Sebat, PhD, Beyster Center for Genomics of Psychiatric Diseases, Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. jsebat@ucsd.edu.
*Current affiliation

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

This study is exempt under 45 CFR 46.101 (b) (4).

Author Contributions

Study concept and design (MK, VV, JS), acquisition of data (MK, VV, JS), analysis and interpretation of data (MK, VV, JS), drafting of the manuscript (MK, VV, JS), critical revision of the manuscript for important intellectual content (MK, VV, RK, SC, JS), donating samples and sample data (RK, SC, HP, AA, JR). Approval of final manuscript (MK, VV, JS), statistical analysis (VV, SC, JS), obtained funding (JS).

neuronal cell types may be needed in order to elucidate the signaling pathways that influence brain development in this genetic disorder.

Keywords

autism spectrum disorders (ASD); schizophrenia (SCZ); copy number variation (CNV); gene expression; 16p11.2

Large (~600 kb) copy number variants (CNVs) of the 16p11.2 region (BP4-BP5) are strongly associated with psychiatric disorders and cognitive deficits.[1–5] Reciprocal deletions and duplications of 16p11.2 result in contrasting clinical and anatomical phenotypes. Deletion is associated primarily with pediatric neurodevelopmental disorders,[3,5] and duplication is associated with a broader range of psychiatric phenotypes, which include bipolar disorder, schizophrenia and autism.[1] Reciprocal deletions and duplications also have differential effects on head size[1] and body mass index (BMI).[6] A variety of other clinical characteristics have been frequently reported including hypotonia,[7] seizures, and bone malformations.[8] Both deletions and duplications at this locus are rare in the general population, occurring at the rate of 0.01% and 0.03% respectively.[1]

The CNV impacts approximately 34 genes (see Supplementary Figure 1 and Supplementary Table 1). Genes responsible for some of the clinical phenotypes observed in 16p11.2 CNV carriers have been identified. A recent study by Golzio et al. in zebrafish[9] determined that the gene *KCTD13* is a key driver of neuronal proliferation and brain growth abnormalities that have been reported in patients.[1,10] Other genes may also influence neurodevelopment, and several candidate genes have been proposed, including *TAOK2*,[11] *TBX6*,[8,12] *SEZ6L2*,[13,14] *MAPK3*,[13] *DOC2A*,[7,13] *QPRT*,[7] and *MVP*.[12,14] However, a clear mechanism is still lacking and the molecular effects of the CNV at the transcriptional level are not well understood.

Gene expression profiling reveals complex biological processes found *in vivo*, and when pathogenesis is mediated through changes in gene expression, it may contribute to discovery of indicators of disease.[15–20] One or multiple genes located within the copy variable region are dosage sensitive and may contribute to the phenotypes seen in 16p11.2 CNV carriers. In addition, dosage sensitive genes within this region may include genes involved in cell signaling and gene regulation, and thereby, the 16p11.2 genotype could impact the expression of genes *in trans* at other locations in the genome. To elucidate the biological mechanisms behind disease pathogenesis and progression, the authors sought to define the set of genes that are dysregulated in patients who carry copy number variants of 16p11.2.

The authors analyzed genomewide expression profiles of lymphoblastoid cell lines (LCLs) of 35 subjects previously genotyped for the 16p11.2 CNV in studies of ASD[21] and SCZ.[1] Subjects included 6 carriers of the deletion, 15 carriers of the duplication and 15 diploid controls. The use of lymphocyte cells to evaluate the transcriptional effects of CNVs on developmental pathways has clear advantages and disadvantages. The primary rationale for this study and for numerous other studies that have used peripheral tissues to study gene expression in ASD[11] and other neurodevelopmental disorders[15,18,22–26] is that blood

samples are readily obtainable and genes involved in neurodevelopmental disorders are often expressed in multiple tissues including blood. Also, genetic disorders that impact neurodevelopment can produce clinical phenotypes in other organ systems. For example, clinical phenotypes observed in some 16p11.2 CNV carriers include scoliosis and a variety of other malformations in bone.[8] In addition, parental reports of recurrent infections and a reported case of severe immunodeficiency[27] suggest a possible influence of the 16p11.2 genes on the immune response. Genes dysregulated in LCLs could identify pathways that directly underlie clinical phenotypes observed in peripheral tissues.

The authors investigated copy number-dependent dosage effects on genes transcribed from the 16p11.2 locus as well as genes outside of the CNV region that are putatively regulated *in trans* by the dosage sensitive 16p11.2 genes. Here the authors show that the copy number of the 16p11.2 region has a strong influence on the expression of genes within the CNV region and a comparatively modest influence on the expression of genes *in trans*.

Materials and Methods

Lymphoblast Cell Line Sample Collection and Maintenance

The authors collected the Epstein-Barr virus immortalized lymphocyte cell lines (LCLs) from 21 16p11.2 CNV carriers with either an autism or schizophrenia diagnosis, and 15 diploid healthy controls. Cryopreserved lymphocyte cell lines were acquired from the Autism Genetic Resource Exchange (AGRE) sample repository. A summary of samples by genotype and phenotype is shown in Supplementary Tables 2 and 3. LCLs were cultured in RPMI1640 medium (Invitrogen, Carlsbad, CA) supplemented with 15% fetal bovine serum (Atlas Biologicals, Ft. Collins, CO), 1% Penicillin/Streptomycin (Invitrogen), 1% MEM nonessential amino acid solution (Invitrogen), and 1% L-Glutamax (Invitrogen). Cells are then placed into 37°C incubator at 5% CO_2. Cells in lag phase were seeded as biological triplicates at a 5×10^6 cells per flask with 10 ml media.

RNA Extraction and Hybridization and Processing of Microarray Data

Cells were pelleted 24 hours postseeding and homogenized using Qiashredder (Qiagen, Valencia, CA). RNA extraction was performed using the RNeasy Mini kit (Qiagen) following the protocol for RNA extraction of animal cells. Samples of cDNA were prepared and hybridized at the Microarray Shared Resource at Cold Spring Harbor Laboratory using the Human Genome U133 Plus 2.0 array (Affymetrix, Santa Clara, CA). Raw data was processed and normalized via robust medioid averaging (RMA) using Gene Spring GX software (Agilent Technologies, Santa Clara, CA) and the averages of 3 replicates were computed for each sample.

Gene Expression Analysis

Gene expression measurements were analyzed using the “samr” package for R (version 1.26), an implementation of the significance analysis of microarrays (SAM) method.[28] The authors performed SAM quantitative analysis, which treated gene expression as the response variable and regressed it on the genotype (copy numbers 1, 2, 3). The SAM score corresponds to the slope of the regression line corrected by a variance penalty term. Delta

parameter, which specifies the maximum tolerated difference between observed and expected SAM scores (dashed lines in the QQ plot in Figure 1), was chosen to capture array probes with scores that significantly deviated from the null distribution and at the same time exclude probes consistent with the null or whose association to the copy number may be spurious (close to the x = y line in Figure 1). The chosen delta parameter corresponds to an empirical FDR of 9.89%. Details of the expression analysis and selection of parameters are provided in Supplementary Materials and Methods. Differential expression between 16p11.2 deletions and copy normal controls, and separately between 16p11.2 duplications and controls, was analyzed using the Limma[29] package in R, following the previously published analysis.[11]

Real-time PCR of Lymphoblast Cell Line RNA

The authors sought to validate expression of *BMP7*, *PTGS1*, and *SIAE*, 3 genes that were observed to be dysregulated *in trans*, using real-time PCR. For each of the 6 deletion and 15 duplication carriers, biological triplicates from the sample were processed in (technical) triplicate, for a total of 9 replicates per cell line. RNA extracted from lymphoblast cell lines was converted to cDNA using the Taqman RT-PCR reagents (Applied Biosystems, Foster City, CA). Sequences of the real-time PCR primers can be found in Supplementary Table 4. Real-time PCR reaction was carried out using 2X SYBR Green PCR master mix (Applied Biosystems) on a Real-time PCR ABI Prism 7000 Sequence Detection System (Applied Biosystems). Results were processed using ABI Prism 7000 SDS software (Applied Biosystems). The effects of concentration and replicate were adjusted using β-actin as the endogenous control, that is, for each of the replicates $\Delta Ct = Ct_{gene} - Ct_{\beta\text{-actin}}$. ΔCt value for each sample was determined as median of reactions for that sample. Statistical significance of ΔΔCt, the difference in means of ΔCt for copy number 1 and copy number 3, was calculated using the Wilcoxon rank sum test. Finally the fold difference was computed as $2^{\Delta\Delta Ct}$.

Replication Gene Expression Data Sets

The authors sought to replicate their results in 2 independent data sets. One replication data set consisted of a subset of 7 deletion carriers, 6 duplication carriers and 120 controls (unaffected siblings of ASD patients) from the Simons Simplex Collection, previously analyzed as a part of an ASD transcriptome study.[11] Data from LCL-derived RNA scanned on the Illumina Whole Human Genome Array HumanRef-8 version 3.0 was processed as described in Luo et al.[11] Association of gene expression to copy number was analyzed using linear regression as in Luo et al.[11] Affymetrix and Illumina probes were matched, when possible, using transcript identifiers and overlapping probe genomic coordinates as provided in the respective annotation files, and when not, using gene identifiers. The second replication data set consisted of published data by Blumenthal et al.[35] No samples overlapped between the authors' data set and the other 2.

Results

Linear Effects of 16p11.2 Genotype on Gene Expression Levels

Based on the authors' previous findings that the regulation of brain growth correlates inversely with 16p11.2 copy number,[1,30] the authors reasoned that transcriptional effects related to clinical phenotypes would follow a similar quantitative relationship with 16p11.2 genotype. Thus, the authors applied a regression-based analysis, SAM, to the microarray data set to identify genes whose expression is correlated with the 16p11.2 copy number.

A Manhattan plot of scores from the SAM analysis revealed a very distinct peak corresponding to gene dosage effects of the 16p11.2 region (Figure 1a). No peaks of a similar magnitude were evident elsewhere in the genome. SAM analysis identified 38 probes in 20 genes that are associated with the copy number, significant at 9.89%FDR (see Table 1). All significant probes were positively correlated with the copy number. The quantile-quantile (QQ) plot of observed versus expected SAM scores (Figure 1b) clearly outlines the significant probes as the group of data points in the top right corner that deviate from the x = y line. The negative correlation tail of the QQ plot did not significantly deviate from the background distribution. The set of probes with FDR ≤ 2.9% captures expression of 18 genes located within the 16p11.2 CNV and no genes outside of the CNV (see Figure 1 and Table 1). At FDR ≤ 9.89%, a total of 20 genes from the CNV region were significant. In addition, SAM scores were high (ranked among the top 0.2% of probes) but nonsignificant for 4 additional 16p11.2 genes, including *YPEL3*, *ASPHD1*, *TAOK2*, and *MAPK3*. Nine genes (*DOC2A*, *SEZ6L2*, *TBX6*, *FAM57B*, *C16orf92*, *GDPD3*, *LOC440356*, *ZG16*, *PPRT2*) showed weak or no dosage effects in lymphocytes. All of the 16p11.2 CNV genes with low SAM scores were well within the boundaries of the CNV, so the apparent lack of correlation with the copy number cannot be attributed to variable CNV breakpoints in different carriers.

Reasoning that the strongest gene dosage effects might have greater phenotypic importance, the authors examined the relative strength of the gene-dosage effects. The highest scoring genes, were *ALDOA*, *CORO1A* and *CDIPT*; however, given that there is a clear relationship between the SAM score of a 16p11.2 gene and its absolute expression levels in lymphoblasts (Spearman ρ = .68; see Figure 2), the highest scores do not necessarily represent the strongest dosage effects. After regressing SAM scores on absolute expression levels and accounting for the variance component due to the absolute expression levels alone, the genes with the largest residuals were *HIRIP3*, *KIF22*, and *KCTD13*.

Specific quantitative clinical phenotypes are influenced by 16p11.2 dosage, including orbital frontal circumference (OFC). The authors reasoned that the genes directly influencing neurodevelopment might be those with expression levels that correlate most strongly with OFC, a question that had been addressed in a previous study.[11] After examining the correlation of gene expression levels with head size for a subset of 7 subjects for which measurements of OFC were available (Supplementary Table 5). The probes for genes *SULT1A3/4*, *KCTD13*, and pseudogene *IMAA* on 16p11.2 and *TMEM242* on 6q25.3 had the strongest linear dependency with head size (Supplementary Table 6). However, these results did not achieve statistical significance after Bonferroni correction for the 89 probes that were analyzed.

The 16p11.2 genotype did not significantly influence the expression of genes adjacent to the CNV; that is, the authors did not observe any positional effects. *Trans*-effects were observed but were weaker than gene dosage effects and did not affect a large number of genes. The authors initially identified 13 significant probes corresponding to genes located outside of the 16p11.2 CNV that were positively correlated with the copy number. The authors applied stringent filtering criteria to remove probes that had low mean expression (<5.8 $\log_2$ scale) and those with high sequence similarity to genes in the 16p11.2 region, in which case significant correlation with genotype might be explained by cross-hybridization. Final analysis pointed to 7 genes: *PTGS1, TMEM242, BMP7, ARL17, SIAE, URB1,* and *IZUMO4* (in bold in Table 1). Genes affected *in trans* did not cluster together, but were all on different chromosomes. The authors attempted to identify molecular associations between the 16p11.2 genes and the significant trans effects via pathway analysis using PANTHER, Consensus DB, Innate DB, Ingenuity Pathway Analysis, and Gather, however no solid links between the 2 groups of genes were found.

Allele-specific Effects of 16p11.2 Genotype

Deletions and duplications of 16p11.2 could impact the expression of different sets of genes. Such allele-specific effects are not captured optimally using linear regression. Therefore, the authors applied an analysis strategy used previously[11] where gene expression levels of 6 deletion and 15 duplications carriers were each compared to the 15 copy normal control subjects. Results of the differential expression analysis showed strong gene-dosage effects associated with the deletion. A total of 16 genes were significantly dysregulated, 14 of which were located within the CNV region and 2 genes, *CCRL1* and *IGF2BP1*, were located on chromosomes 3 and 17 respectively. The authors did not detect significant differential expression associated with the duplication.

Validation of Gene Expression Findings

The authors compared their results to the association of gene expression with 16p11.2 copy number in 2 orthogonal data sets, both assaying expression in lymphoblastoid cell lines derived from 16p11.2 deletion and duplication carriers. The first was recently published in Luo et al[11] as a part of an ASD CNV transcriptome study, and the second is part of a published study by Blumenthal et al.[35] While this study and these agree that gene dosage effects are evident for a majority of genes within the CNV region (see Supplementary Table 7), the significant trans effects in the authors' study do not replicate consistently in Luo et al and Talkowski et al. Replication *P* values for the 3 (out of 7) trans effect genes that were also assayed in Luo et al were .015 (*PTGS1*), .24 (*BMP7*), and .79 (*SIAE*), and none of these genes were significantly correlated with 16p11.2 gentoype in Talkowski et al. The authors note that the alternative transcript of BMP7 identified in the authors' analysis (Affymetrix probe 211260_at) was not assayed on the Illumina platform; therefore the authors were unable to replicate this result in Luo et al. Similarly, the authors were unable to replicate the allele-specific effects of 16p11.2 genotype, since *CCRL1* and *IGF2BP1* were not assayed on the Illumina chip used in Luo et al.[11]

The authors further sought to validate the trans effects of *BMP7, SIAE* and *PTGS1*, using real-time quantitative PCR (qPCR) of the RNA samples used in this study. All 3 genes

showed a trend toward lower expression in deletion than in duplication carriers in the authors' samples (fold differences 1.52, 1.20 and 1.28 for *BMP7*, *PTGS1*, and *SIAE* respectively), However, results were statistically significant for *BMP7* only (Wilcoxon rank sum $P = .0124$; see Figure 3).

Discussion

The authors analyzed the effects of the reciprocal deletions and duplications at the 16p11.2 locus, using genomewide expression profiling of LCLs derived from blood samples of CNV carriers and copy normal controls. The authors have shown that dosage effects dominate the expression differences observed in CNV carriers: probes significant at 9.89% FDR include 20 out of 34 genes located within the 16p11.2 CNV. The authors identified 7 genes outside of the 16p11.2 locus whose expression was positively correlated with copy number; however, the authors did not find strong evidence supporting the same trans effects in independent gene expression data sets and the authors did not obtain strong experimental confirmation in the authors' study sample by qPCR. In summary, the authors' findings demonstrate that strong dosage effects are present among the set of genes located within the CNV, and genetic effects elsewhere in the genome are modest and difficult to reproduce. It will be of interest to compare these results to those of other large CNVs associated with neurodevelopmental phenotypes to determine if the same is true for other disease-associated CNVs.

After accounting for the effects of absolute expression levels, gene dosage effects were greatest for *HIRIP3*, *KIF22*, and *KCTD13*. The relative correlation of gene expression with head size was greatest for *SULT1A3/4*, *IMAA*, and *KCTD13*. Importantly, similar results were observed for one of these genes (*KCTD13*) in an independent study.[11] The comparatively strong dosage effect observed for the *KCTD13* gene and the strong correlation of *KCTD13* expression with head size is particularly notable in light of a recent study that identified this gene as a key driver of brain growth abnormalities in a zebra fish model system.[9] The authors' transcriptome results are consistent with the possible role of this gene in influencing molecular and clinical phenotypes in humans.

Copy number change of the 16p11.2 region had a genetic effect on the expression of a small number of genes *in trans*. The observed trans effects were weak and differed considerably from the results of a previous study.[11] The contrasting results between studies could be explained by a combination of factors including the very subtle transcriptional effects on genes *in trans*, differences between the array platforms and probe sets used in each study, and the smaller number of controls (and consequently reduced power) in this study.

The few genes that were found to be dysregulated *in trans* could constitute downstream components of signaling pathways that are altered in patient lymphocytes. Bone morphogenic protein 7 (*BMP7*) belongs to the transforming growth factor (TGF-β) super family of secreted signaling molecules,[31] which play a role in early neuronal and epidermal development[32] and in the formation of bones and cartilage. Prostaglandin-endoperoxide synthase 1 (*PTGS1/COX1*) is involved in mediating the inflammatory response and maintenance of the normal stomach lining.[33] However, substantially larger sample sizes will

be needed to determine whether these and other genes are in fact dysregulated in 16p11.2 CNV carriers.

The authors' results suggest that copy number variation of 16p11.2 has a modest effect on global patterns of gene expression in lymphocytes. This observation highlights the perils of using samples of peripheral tissue to investigate the molecular mechanism of a neurodevelopmental disorder. Assuming that the neurodevelopmental phenotypes associated with the 16p11.2 CNV are mediated at the transcriptional level, the relevant signaling pathways may not be active to the same extent in peripheral tissues and in brain. In neuronal tissues, transcriptional effects of 16p11.2 may differ and the impact of the CNV on patterns of gene expression could be strong in specific cell types and at specific developmental time points. For instance, 1 of the genes within the CNV region, *TBX6*, is a transcription factor with expression that is restricted to the primitive streak in the gastrulation stage embryo and later in development to the neural tube.[34]

The results of this study suggest that, even when transcriptome analysis is focused on a single mutation of large effect, understanding the molecular mechanisms of disease remains a challenge. If scholars wish to better understand the molecular pathophysiology of this and other genetic disorders, molecular studies in neuronal tissues are needed. Patient-derived induced pluripotent stem cells and neuronal cell cultures are likely to yield greater insight into transcriptional pathways relevant to neurodevelopment. Further efforts must also focus on understanding cell-signaling pathways directly under the regulation of 16p11.2 genes at the protein level.

Acknowledgments

We gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by NIMH Grant 1U24MH081810 to Clara M. Lajonchere (PI). The AGRE Consortium: Dan Geschwind, MD, PhD, UCLA, Los Angeles, CA; Maja Bucan, PhD, University of Pennsylvania, Philadelphia, PA; W. Ted Brown, MD, PhD, FACMG, NYS, Institute for Basic Research in Developmental Disabilities, Staten Island, NY; Rita M. Cantor, PhD, UCLA School of Medicine, Los Angeles, CA; John N. Constantino, MD, Washington University School of Medicine, St. Louis, MO; T. Conrad Gilliam, PhD, University of Chicago, Chicago, IL; Martha Herbert, MD, PhD, Harvard Medical School, Boston, MA; Clara Lajonchere, PhD, Autism Speaks, Los Angeles, CA; David H. Ledbetter, PhD, Emory University, Atlanta, GA; Christa Lese-Martin, PhD, Emory University, Atlanta, GA; Janet Miller, JD, PhD, Autism Speaks, Los Angeles, CA; Stanley F. Nelson, MD, UCLA School of Medicine, Los Angeles, CA; Gerard D. Schellenberg, PhD, University of Washington, Seattle, WA; Carol A. Samango-Sprouse, EdD, George Washington University, Washington, DC; Sarah Spence, MD, PhD, UCLA, Los Angeles, CA; Matthew State, MD, PhD, Yale University, New Haven, CT; Rudolph E. Tanzi, PhD, Massachusetts General Hospital, Boston, MA; Rui Luo, UCLA, Los Angeles, CA. We wish to thank Dan Geschwind, Rui Luo, and Mike Talkowski for sharing with us expression data on independent samples of subjects with 16p11.2 CNVs.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this study was provided grants to JS from the Simons Foundation (SFARI 178088) and the National Institutes of Health (MH076431).

References

1. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 2009; 41:1223–1227. [PubMed: 19855392]

2. Mefford HC, Eichler EE. Duplication hotspots, rare genomic disorders, and common disease. Curr Opin Genet Dev. 2009; 19:196–204. [PubMed: 19477115]
3. Kumar RA, KaraMohamed S, Sudi J, et al. Recurrent 16p11.2 microdeletions in autism. Hum Mol Genet. 2008; 17:628–638. [PubMed: 18156158]
4. Mefford HC, Cooper GM, Zerr T, et al. A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. Genome Res. 2009; 19:1579–1585. [PubMed: 19506092]
5. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 2008; 358:667–675. [PubMed: 18184952]
6. Jacquemont S, Reymond A, Zufferey F, et al. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature. 2011; 478:97–102. [PubMed: 21881559]
7. Bedoyan JK, Kumar RA, Sudi J, et al. Duplication 16p11.2 in a child with infantile seizure disorder. Am J Med Genet A. 2010; 152A:1567–1574. [PubMed: 20503337]
8. Shimojima K, Inoue T, Fujii Y, Ohno K, Yamamoto T. A familial 593-kb microdeletion of 16p11.2 associated with mental retardation and hemivertebrae. Eur J Med Genet. 2009; 52:433–435. [PubMed: 19770079]
9. Golzio C, Willer J, Talkowski ME, et al. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature. 2012; 485:363–367. [PubMed: 22596160]
10. Shinawi M, Liu P, Kang SH, et al. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioral problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet. 2010; 47:332–341. [PubMed: 19914906]
11. Luo R, Sanders Stephan J, Tian Y, et al. Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am J Hum Genet. 2012; 91:38–55. [PubMed: 22726847]
12. Shen Y, Chen X, Wang L, et al. Intra-family phenotypic heterogeneity of 16p11.2 deletion carriers in a three-generation Chinese family. Am J Med Genet B. 2011; 156:225–232.
13. Kumar RA, Marshall CR, Badner JA, et al. Association and mutation analyses of 16p11.2 autism candidate genes. PLOS ONE. 2009; 4:e4582. [PubMed: 19242545]
14. Crepel A, Steyaert J, De la Marche W, et al. Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. Am J Med Genet B. 2011; 156:243–245.
15. Bowden NA, Weidenhofer J, Scott RJ, et al. Preliminary investigation of gene expression profiles in peripheral blood lymphocytes in schizophrenia. Schizophr Res. 2006; 82:175–183. [PubMed: 16414245]
16. Curtis RK, Oresic M, Vidal-Puig A. Pathways to the analysis of microarray data. Trends Biotechnol. 2005; 23:429–435. [PubMed: 15950303]
17. de Vries BB, Pfundt R, Leisink M, et al. Diagnostic genome profiling in mental retardation. Am J Hum Genet. 2005; 77:606–616. [PubMed: 16175506]
18. Hu VW, Frank BC, Heine S, Lee NH, Quackenbush J. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. BMC Genom. 2006; 7:118.
19. Whitney AR, Diehn M, Popper SJ, et al. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci USA. 2003; 100:1896–1901. [PubMed: 12578971]
20. Yoon H, Liyanarachchi S, Wright FA, et al. Gene expression profiling of isogenic cells with different TP53 gene dosage reveals numerous genes that are affected by TP53 dosage and identifies CSPG2 as a direct target of p53. Proc Natl Acad Sci USA. 2002; 99:15632–15637. [PubMed: 12438652]
21. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. Science. 2007; 316:445–449. [PubMed: 17363630]
22. Traynor J, Agarwal P, Lazzeroni L, Francke U. Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. BMC Med Genet. 2002; 3:12. [PubMed: 12418965]
23. Baron CA, Liu SY, Hicks C, Gregg JP. Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. J Autism Dev Disord. 2006; 36:973–982. [PubMed: 16845580]

24. Nishimura Y, Martin CL, Vazquez-Lopez A, et al. Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. Hum Mol Genet. 2007; 16:1682–1698. [PubMed: 17519220]
25. Baron CA, Tepper CG, Liu SY, et al. Genomic and functional profiling of duplicated chromosome 15 cell lines reveal regulatory alterations in UBE3A-associated ubiquitin-proteasome pathway processes. Hum Mol Genet. 2006; 15:853–869. [PubMed: 16446308]
26. Slonimsky A, Levy I, Kohn Y, et al. Lymphoblast and brain expression of AHI1 and the novel primate-specific gene, C6orf217, in schizophrenia and bipolar disorder. Schizophr Res. 2010; 120:159–166. [PubMed: 20452750]
27. Shiow LR, Paris K, Akana MC, et al. Severe combined immunodeficiency (SCID) and attention deficit hyperactivity disorder (ADHD) associated with a Coronin-1A mutation and a chromosome 16p11.2 deletion. Clin Immunol. 2009; 131:24–30. [PubMed: 19097825]
28. Tusher VG, Tibshirani R, Chu G. Significance analysis of micro-arrays applied to the ionizing radiation response. Proc Natl Acad Sci. 2001; 98:5116–5121. [PubMed: 11309499]
29. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3:3.
30. Ricard G, Molina J, Chrast J, et al. Phenotypic consequences of copy number variation: insights from smith-magenis and potocki-lupski syndrome mouse models. PLOS Biol. 2010; 8:e1000543. [PubMed: 21124890]
31. Chen D, Zhao M, Mundy GR. Bone morphogenetic proteins. Growth Factors. 2004; 22:233–241. [PubMed: 15621726]
32. Suzuki A, Kaneko E, Ueno N, Hemmati-Brivanlou A. Regulation of epidermal induction by BMP2 and BMP7 signaling. Dev Biol. 1997; 189:112–122. [PubMed: 9281341]
33. Baumgartner HK, Starodub OT, Joehl JS, Tackett L, Montrose MH. Cyclooxygenase 1 is required for pH control at the mouse gastric surface. Gut. 2004; 53:1751–1757. [PubMed: 15542509]
34. Chapman DL, Garvey N, Hancock S, et al. Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. Dev Dyn. 1996; 206:379–390. [PubMed: 8853987]
35. Blumenthal I, Ragavendran A, Erdin S, et al. Transcriptional consequences of 16p11.2 deletion and duplication in mouse cortex and multiplex autism families. Am J Hum Genet. 2014; 94:870–883. [PubMed: 24906019]

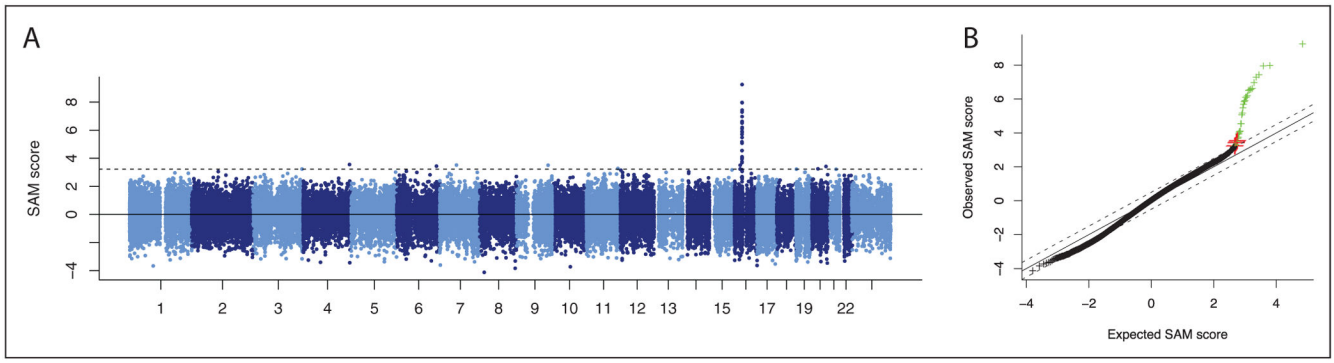

Figure 1.

Results of the SAM analysis displayed as A) Manhattan and B) quantile-quantile (QQ) plots highlight the strong signal at the 16p11.2 locus and weak signals elsewhere in the transcriptome. QQ plot of observed vs. expected SAM scores highlights genes that have expression significantly correlated to the copy number. The first 28 significant probes are positively correlated with the copy number and are 16p11.2 genes, while the next 9 significant probes include the 7 trans effects denoted by the larger points deviating from the background distribution (x = y line). The dashed lines correspond to an empirical 9.88% false discovery rate (FDR) significance cut off.

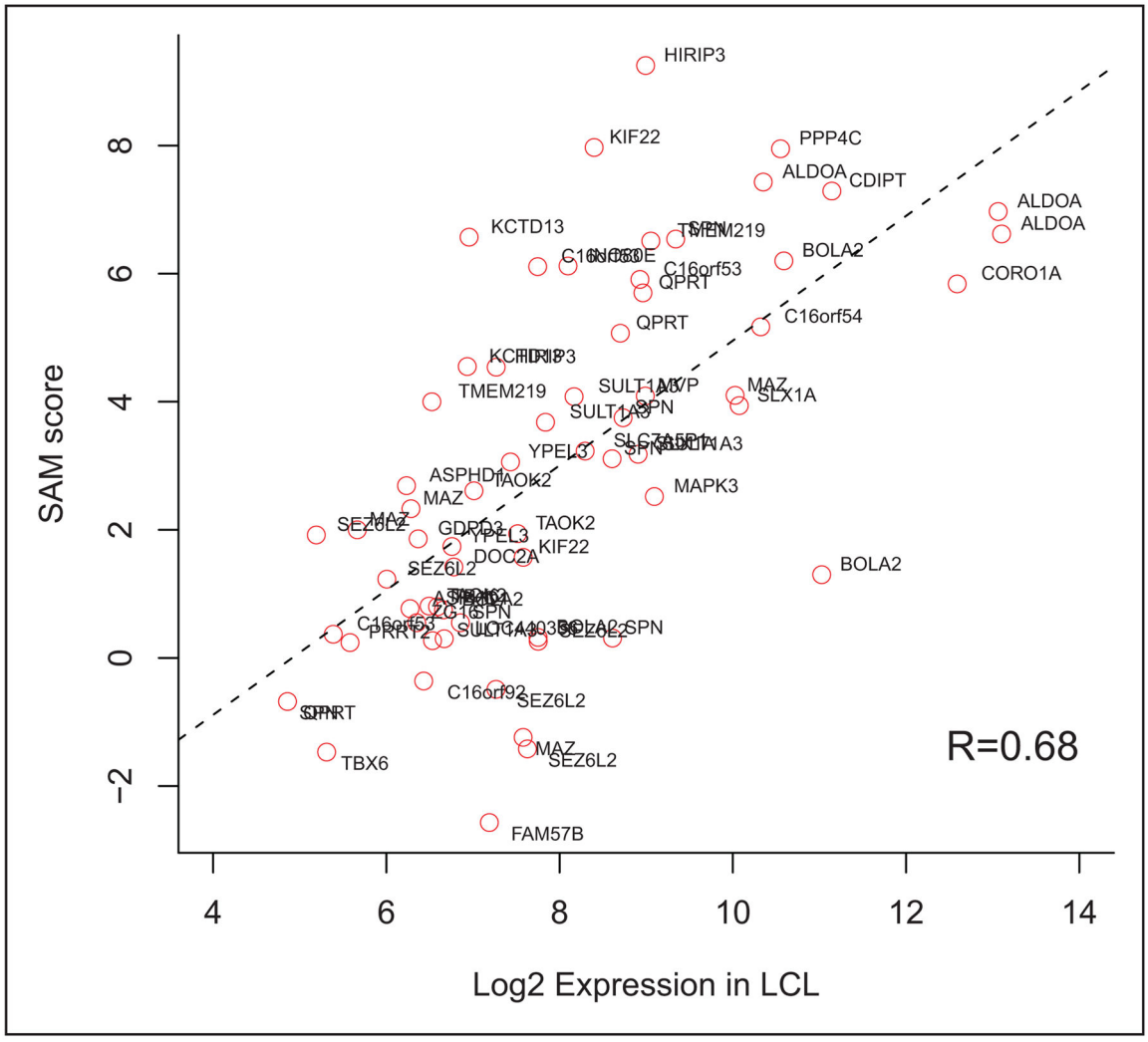

Figure 2.

The relative influence of gene dosage on the expression of different genes is driven by the transcript level in lymphocytes. Scatter plot of $\log_2$ gene expression in lymphoblasts against SAM scores for 16p11.2 CNV genes shows a strong correlation between the 2. Expression is the median $\log_2$ signal intensity across all samples. The dashed line is the linear fit between expression and SAM score. Spearman correlation coefficient was .68.

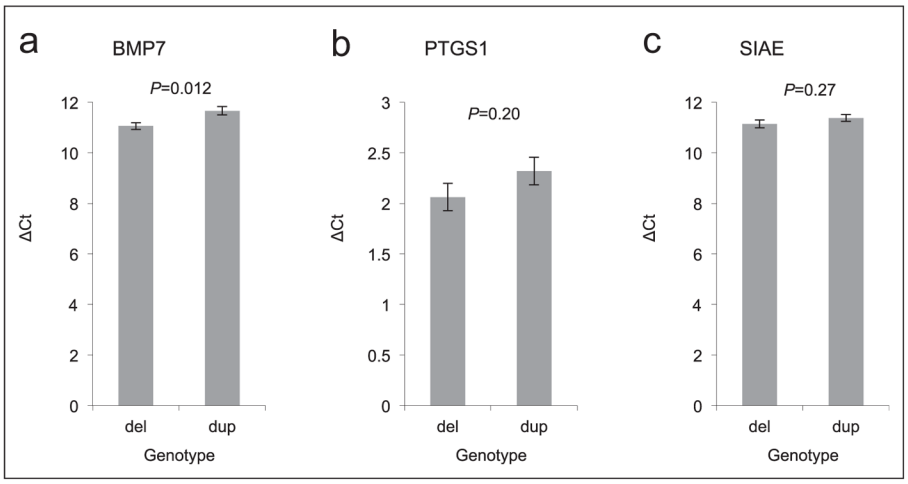

Figure 3.

Mean and standard errors of RT-PCR gene expression levels for (a) BMP7, (b) PTGS1, and (c) SIAE in 6 deletion and 15 duplication carriers.

Table 1

List of Top Probes From Genomewide SAM Expression Analysis, Ordered by False Discovery Rate (FDR).

	Affymetrix probe ID	Gene symbol	Cytoband	FDR (%)
1	204504_s_at	HIRIP3	16p11.2	0
2	202183_s_at	KIF22	16p11.2	0
3	208932_at	PPP4C	16p11.2	0
4	238996_x_at	ALDOA	16p11.2	0
5	201253_s_at	CDIPT	16p11.2	0
6	200966_x_at	ALDOA	16p11.2	0
7	214687_x_at	ALDOA	16p11.2	0
8	221889_at	KCTD13	16p11.2	0
9	1568964_x_at	SPN	16p11.2	0
10	224981_at	TMEM219	16p11.2	0
11	209836_x_at	BOLA2	16p11.2	0
12	231878_at	C16orf53	16p11.2	0
13	227286_at	INO80E	16p11.2	0
14	218300_at	C16orf53	16p11.2	0
15	209083_at	CORO1A	16p11.2	0
16	242414_at	QPRT	16p11.2	0
17	238142_at	*retired	16p11.2	0
18	1559584_a_at	C16orf54	16p11.2	0
19	204044_at	QPRT	16p11.2	0
20	229697_at	HIRIP3	16p11.2	0
21	45653_at	KCTD13	16p11.2	0
22	210580_x_at	SULT1A3	16p11.2	0
23	202180_s_at	MVP	16p11.2	0
24	212064_x_at	MAZ	16p11.2	0
25	228513_at	TMEM219	16p11.2	0
26	218317_x_at	SLX1A	16p11.2	0
27	206057_x_at	SPN	16p11.2	2.90
28	209607_x_at	SULT1A3	16p11.2	2.90
29	**215813_s_at**	**PTGS1**	**9q32-q33.3**	**5.07**
30	244766_at	SMG1 homolog	16p11.2,12.2,12.3	5.07
31	**218453_s_at**	**TMEM242**	**6q25.3**	**7.15**
32	**211260_at**	**BMP7**	**20q13**	**7.15**
33	**1554245_x_at**	**ARL17**	**17q21.31**	**9.27**
34	**223744_s_at**	**SIAE**	**11q24**	**9.89**
35	208118_x_at	IMAA	16p11.2, 12.2	9.89
36	1556103_at	*retired probe	—	9.89
37	**217633_at**	**URB1**	**21q22.11**	**9.89**
38	**215734_at**	**IZUMO4**	**19p13.3**	**9.89**

The top 28 significant probes are within the 16p11.2 variation, followed by probes corresponding to genes outside of the CNV (trans effects, in bold). Genes associated with probes marked with an asterisk have been retired from the latest human genome annotations.