Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes

Loviglio, M N; Leleu, M; Männik, K; Passeggeri, M; Giannuzzi, G; van der Werf, I; Waszak, S M; Zazhytska, M; Roberts-Caldeira, I; Gheldof, N; Migliavacca, E; Alfaiz, A A; Hippolyte, L; Maillard, A M [**and 18 more**]

How to cite

LOVIGLIO, M N et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. In: Molecular Psychiatry, 2017, vol. 22, n° 6, p. 836–849. doi: 10.1038/mp.2016.84

This publication URL: https://archive-ouverte.unige.ch/unige:112825
Publication DOI: 10.1038/mp.2016.84

Molecular Psychiatry (2017) 22, 836–849
www.nature.com/mp

OPEN

ORIGINAL ARTICLE

Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes

MN Loviglio[1,24], M Leleu[2,3,24], K Männik[1,4], M Passeggeri[5], G Giannuzzi[1], I van der Werf[1], SM Waszak[2,3,6], M Zazhytska[1], I Roberts-Caldeira[5], N Gheldof[1], E Migliavacca[1,2], AA Alfaiz[1,2], L Hippolyte[5], AM Maillard[5], 2p15 Consortium[25], 16p11.2 Consortium[26], A Van Dijck[7], RF Kooy[7], D Sanlaville[8], JA Rosenfeld[9,10], LG Shaffer[11], J Andrieux[12], C Marshall[13], SW Scherer[14,15], Y Shen[16,17,18], JF Gusella[19,20], U Thorsteinsdottir[21], G Thorleifsson[21], ET Dermitzakis[2,22], B Deplancke[2,3], JS Beckmann[2,5,23], J Rougemont[2,3], S Jacquemont[5,27,28] and A Reymond[1,28]

Copy number variants (CNVs) are major contributors to genomic imbalance disorders. Phenotyping of 137 unrelated deletion and reciprocal duplication carriers of the distal 16p11.2 220 kb BP2-BP3 interval showed that these rearrangements are associated with autism spectrum disorders and mirror phenotypes of obesity/underweight and macrocephaly/microcephaly. Such phenotypes were previously associated with rearrangements of the non-overlapping proximal 16p11.2 600 kb BP4-BP5 interval. These two CNV-prone regions at 16p11.2 are reciprocally engaged in complex chromatin looping, as successfully confirmed by 4C-seq, fluorescence *in situ* hybridization and Hi-C, as well as coordinated expression and regulation of encompassed genes. We observed that genes differentially expressed in 16p11.2 BP4-BP5 CNV carriers are concomitantly modified in their chromatin interactions, suggesting that disruption of chromatin interplays could participate in the observed phenotypes. We also identified *cis*- and *trans*-acting chromatin contacts to other genomic regions previously associated with analogous phenotypes. For example, we uncovered that individuals with reciprocal rearrangements of the *trans*-contacted 2p15 locus similarly display mirror phenotypes on head circumference and weight. Our results indicate that chromosomal contacts' maps could uncover functionally and clinically related genes.

Molecular Psychiatry (2017) **22,** 836–849; doi:10.1038/mp.2016.84; published online 31 May 2016

INTRODUCTION

Long-range chromatin contacts that bring genes and regulatory sequences in close proximity are necessary for co-transcription of biologically related and developmentally co-regulated genes.[1,2] Correspondingly, genomic structural changes were associated with disruption of the organization of chromatin compartments by shifting regulatory elements between domains and/or modifying domain boundaries, which resulted in ectopic interactions, gene misexpression and disease.[3,4] In the last 15 million years the 16p11.2–12.2 region rapidly integrated segmental duplications contributing to profound modifications of these chromosomal bands in hominoids.[5,6] It allowed the emergence of new transcripts[7] and placed the whole region at risk for various recurrent rearrangements[8–10] through non-allelic homologous recombination[11] (Figure 1). These rearrangements include a recurrent interstitial deletion of ~600 kb defined by 16p11.2 breakpoints 4–5 (BP4-BP5; OMIM#611913), which encompasses 28 'unique' genes and four genes with multiple copies[12] (Figure 1). With a population prevalence of ~0.05% this variant is one of the most frequent known etiologies of autism spectrum disorder (ASD).[9,13–15] It impacts adaptive behavior and language skills and predisposes to a highly penetrant form of obesity and macrocephaly.[15–19] A mirror phenotype is observed in carriers of the reciprocal duplication (OMIM#614671), who present a high risk of schizophrenia (SCZ), Rolandic epilepsy, being underweight and microcephalic.[18–22] Case series have reported variable expressivity;

[1]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; [2]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland; [3]School of Life Sciences, EPFL (Ecole Polytechnique Fédérale de Lausanne), Lausanne, Switzerland; [4]Estonian Genome Center, University of Tartu, Tartu, Estonia; [5]Service of Medical Genetics, Lausanne University Hospital (CHUV), Lausanne, Switzerland; [6]Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany; [7]Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; [8]Service de Génétique, Hôpital Femme Mère Enfant, Centre de Biologie et Pathologie Est, Hospices Civils de Lyon, CRNL, CNRS UMR5292, INSERM U1028, Université Claude Bernard Lyon I, Bron, France; [9]Signature Genomic Laboratories, PerkinElmer, Inc., Spokane, WA, USA; [10]Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA; [11]Paw Print Genetics, Genetic Veterinary Sciences, Inc., Spokane, WA, USA; [12]Institut de Génétique Médicale, CHRU de Lille - Hôpital Jeanne de Flandre, Lille, France; [13]Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, USA; [14]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, USA; [15]McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, ON, USA; [16]Department of Pathology, Harvard Medical School, Boston, MA, USA; [17]Department of Laboratory Medicine, Boston Children's Hospital, Boston, MA, USA; [18]Shanghai Jiaotong University, School of Medicine, Shanghai, China; [19]Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; [20]Departments of Genetics and Neurology, Harvard Medical School, Boston, MA, USA; [21]deCODE Genetics, Amgen, Reykjavik, Iceland; [22]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland and [23]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. Correspondence: Professor A Reymond, Center for Integrative Genomics, University of Lausanne, Genopode Building, Lausanne CH-1015, Switzerland.
E-mail: alexandre.reymond@unil.ch
[24]Co-first authors.
[25]The list of the 16p11.2 Consortium members is specified before conflict of interest, while their affiliations are detailed in Supplemental Data.
[26]The list of the 2p15 Consortium members is specified before conflict of interest, while their affiliations are detailed in Supplemental Data.
[27]Current address: CHU Sainte Justine, University of Montreal, Montreal, QC, Canada.
[28]Co-senior authors.
Received 4 November 2015; revised 18 March 2016; accepted 18 April 2016; published online 31 May 2016

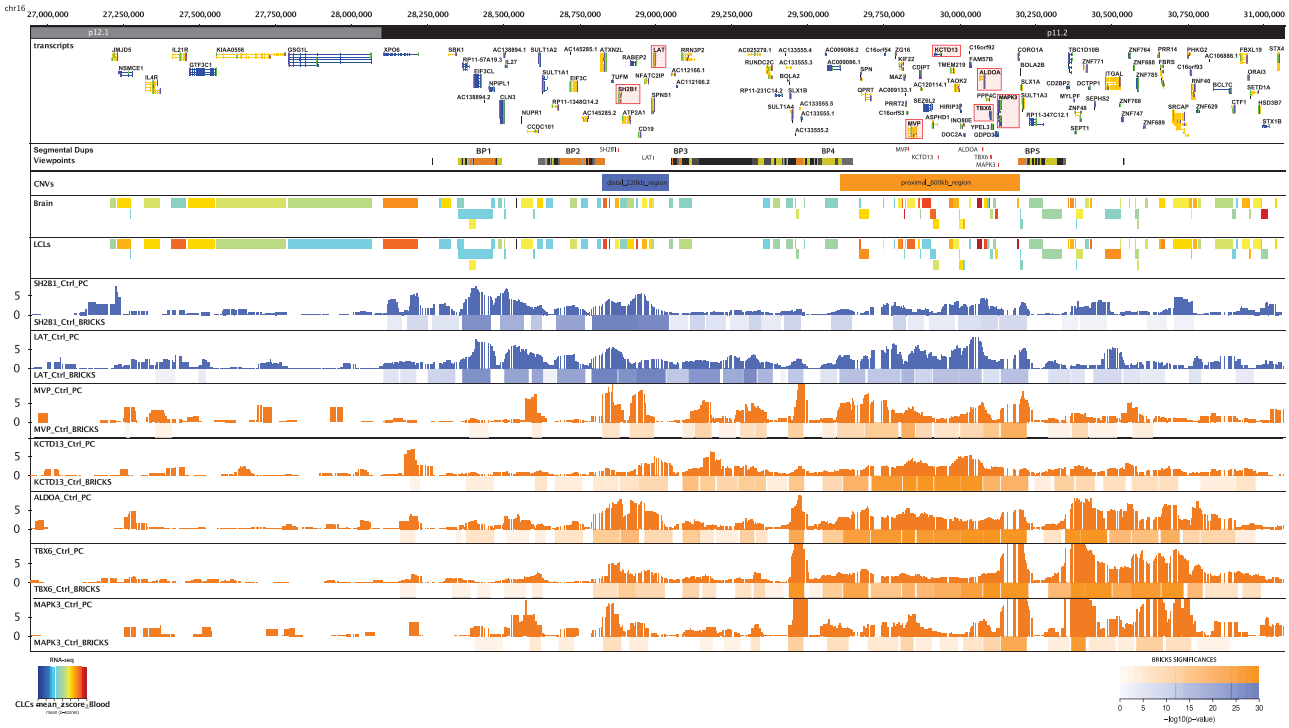

Figure 1. The 16p11.2 region and its 4C interactions profile (panels from top to bottom). *Transcripts:* The transcripts mapping within the human chromosome 16 GRCh37/hg19 27–31 Mb region are indicated. The 4C-targeted *SH2B1, LAT, MVP, KCTD13, ALDOA, TBX6* and *MAPK3* genes are highlighted in red. *Segmental duplications/viewpoints:* The duplicated regions containing the low-copy repeats (LCR) that flank these rearrangements telomerically and centromerically are shown, whereas the position of the restrictions fragments used as viewpoints are marked with red ticks. *CNVs:* The position of the 600 kb BP4-BP5 (orange) and 220 kb BP2-BP3 intervals (blue) are depicted. *Brain/LCLs:* The mean *z*-score for transcript expression per group (Brain or LCLs) from GTEx is displayed. The corresponding RNA-seq heatmap color legend is showed at the bottom left corner. *PC/BRICKs:* Smoothed and profile-corrected 4C signal (upper part of each panel) and BRICKs (lower part) identified for each of the seven 4C viewpoints within the 16p11.2 cytoband, that is, from top to bottom *SH2B1, LAT, MVP, KCTD13, ALDOA, TBX6* and *MAPK3*. The corresponding BRICKs significance heatmap color legend is showed at the bottom right corner.

systematic phenotyping showed that deletion and duplication lead to an average IQ decrease of 26 and 16 points in proband compared with non-carrier family members.[15,23] Correspondingly, the phenotypes of carriers identified in unselected populations are reminiscent of those described for carriers of 16p11.2 rearrangements ascertained in clinical cohorts.[24] Deletions and duplications show a mirroring impact on brain volume and specific cortico-striatal structures implicated in reward, language and social cognition.[25] Changes in copy numbers of this interval are associated with significant modifications of the mRNA levels of ciliopathy and ASD-associated genes in humans and mice.[12,26] Correspondingly, mouse models engineered to have three copies of the 7qF3 orthologous region showed reduced cilia length in the CA1 hippocampal region, whereas modulation of the expression of ciliopathy-associated genes rescued phenotypes induced by *KCTD13* (MIM#608947) under- and overexpression,[12] one of the key drivers of the 16p11.2 600 kb BP4-BP5 CNV genomic-interval associated traits.[27] Distal to BP4-BP5, the deletion of 16p11.2 220 kb BP2-BP3 interval was similarly associated with obesity, developmental delay, intellectual disability and SCZ.[16,28–35] However, detailed data about the phenotypes associated with the reciprocal duplication are still lacking.

We hypothesized that copy number modification of the 16p11.2 600 kb BP4-BP5 interval alters the three-dimensional positioning of these genes resulting in expression alterations of pathways involved in its phenotypic manifestation. We used chromatin conformation capture to explore the chromosome-wide effects of the 16p11.2 600 kb BP4-BP5 structural rearrangements on chromatin structure and assessed how these underlay the associated phenotypes. This region engages in multi-gene complex structures that are disrupted when its copy number changes. The implicated genes are known to be linked to 16p11.2-associated phenotypes, such as primary cilium alteration, energy imbalance, head circumference (HC) and ASD. We also demonstrate that our approach could be used to identify additional loci, whose copy number changes are associated with strikingly similar phenotypic manifestations.

MATERIALS AND METHODS

Recruitment and phenotyping of patients

The institutional review board of the University of Lausanne, Switzerland approved this study. Participants were enrolled in the study after signing an informed consent form and being clinically assessed by their respective physicians. For the data collected through questionnaires, information was gathered retrospectively and anonymously by physicians who had ordered chromosomal microarray analyses performed for clinical purposes only. Consequently, research-based informed consent was not required by the institutional review board of the University of Lausanne, which granted an exemption for this part of the data collection. Overall cognitive functioning was assessed as published.[15]

To better assess the phenotypic features associated with the 16p11.2 220 kb BP2-BP3 rearrangements, we recruited and phenotyped 110 and 57 carriers of the 220 kb BP2-BP3 deletion (OMIM#613444) and duplication from 88 and 49 families, respectively (Supplementary Table S1). Whereas these structural variants were previously reported to be among the CNVs most frequently harboring a possibly deleterious second genetic lesion (29 and 13% of the time, respectively),[32] we do not confirm such propensity. Indeed, second-site structural variants were identified in 'only' 7% (6/88) and 4% (2/49) of the enrolled BP2-BP3 deletion and duplication carrier probands, respectively. Deleterious CNVs were defined as: (i) known

recurrent genomic disorder, (ii) CNV encompassing published critical genomic region or disrupting a gene that is a known etiology of neurodevelopmental disorders or (iii) >500 kb CNV with AF < 0.001. We compared available data on weight, height, body mass index (BMI) and HC for 77 and 39 unrelated deletion and duplication carriers, respectively (including published cases). The mean age of this group of patients was 16 years (range 0.42–78 years, with 34 cases older than 18 years). The prevalence of the 16p11.2 deletion and duplication were inferred from six European population-based genome-wide association studies cohorts, sets of chromosomal microarray-genotyped control individuals and clinical cohorts.[9,30,33,35–39] CNV analyses were carried out as described in Jacquemont *et al*.[18]

We similarly enrolled 26 and 9 unrelated carriers of 2p15 deletions and duplications, including 12 deletion cases from the literature.[40–49] The Signature Genomics cases were recently described in Jorgez *et al*.[50] Patients were identified through routine etiological work-ups of patients ascertained for developmental delay/intellectual disability in cytogenetic centers. The coordinates of the rearrangements' breakpoints (Supplementary Table S2) were recognized by different chromosomal microarray platforms.

Lymphoblastoid cell lines and transcriptome profiling

We had previously established by Epstein-Barr virus transformation lymphoblastoid cell lines (LCLs) from 16p11.2 BP4-BP5 patients, as well as controls. The LCL transcriptome of 50 deletion and 31 duplication carriers, as well as 17 control individuals was previously profiled with Affymetrix GeneChips Human Genome U133+ PM 24 array plates (Affymetrix, Santa Clara, CA, USA). The results are deposited in the NCBI Gene Expression Omnibus under accession number GSE57802. The Robust Multi-array Average approach was used for the creation and normalization of the summarized probe set signals. We applied a nonspecific filter to discard probe sets with low variability and low signal, that is, detectable expression levels. Specifically, probe sets with both (i) signal SD > median of signal SD of all probe sets and (ii) larger signal > median of larger signal of all probe sets were retained as described in Migliavacca *et al*.[12] This selection yielded a total of 23,602 probe sets. To reduce a potential bias toward genes with multiple probe sets, for the modular analysis, only one probe set with the highest variance per gene was kept, for a total of 15,112 probe sets. Using a dosage effect model and moderated *t*-statistics, we identified 1188 and 2209 significantly differentially expressed genes (false discovery rate (FDR) ⩽ 1 and 5%, respectively; uniquely mapping probes).[12] We used Geneprof to access data pertaining to gene expression and co-regulation.

We are well aware of the limitations of the study of LCLs, for instance for genes whose expression specificity resides in other cell lineages. These experiments are nevertheless worth pursuing simply because (i) the primary human target tissues remain often beyond reach; (ii) we cannot exclude a broad to ubiquitous expression pattern and chromatin contacts for the genes involved in these disease processes; and (iii) the pattern of expressions in peripheral tissue may be used as a biomarker in translational project. Similar limitations apply to the use of embryonic stem cells-derived material, while animal tissues have a different set of shortcomings.

Quantitative reverse transcriptase-PCR

For qPCR, 100 ng of high-quality total RNA was converted to cDNA using Superscript VILO (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Primers were designed using PrimerExpress 2.0 software (Applied Biosystems, Foster City, CA, USA), with default parameters except for the primer- and minimal amplicon lengths, which were set at 17–26 and 60 bp respectively. The amplification factor of each primer pair was tested using a cDNA dilution series and only assays with amplification factors between 1.75 and 2.00 were retained. A representative set of samples was tested for genomic contamination. qPCR experiments were performed in triplicate using SYBR-Green (Roche, Basel, Switzerland) as reporter. The reaction mixtures were prepared in 384-well plates using a Freedom Evo robot (Tecan, Männedorf, Switzerland) and run in an ABI 7900HT sequence detection system (Applied Biosystems) using the following conditions: 50°C for 2 min, 95°C for 10 min, followed by 45 cycles of 95°C for 15 s and then 60°C for 1 min, after which dissociation curves were established. Applicable normalization genes were included in each experiment to enable compensation for fluctuations in expression levels between experiments. Using SDS v2.4 software (Applied Biosystems) the threshold and baseline values were adjusted when necessary to obtain raw cycle threshold (Ct) values. The Ct values were further analysed using qBase plus software (Biogazelle, Zwijnaarde, Belgium), which calculates relative expression values per sample per tested gene upon designation of the normalization genes and corrects for the amplification efficiency of the performed assay. We assessed by qPCR the RNA levels of seven DE genes belonging to the ciliopathy or PTEN pathway (*BBS4* (MIM#600374), *BBS7* (MIM#607590), *BBS10* (MIM#610148), *XPOT* (MIM#603180), *NUP58* (MIM#607615), *PTPN11* (MIM#176876) and *SMAD2* (MIM#601366)), and five others that map either to the BP4-BP5 (*ALDOA* (MIM#103850), *KCTD13, MAPK3* (MIM#601795) and *MVP* (MIM#605088)) or the BP2-BP3 interval (*SH2B1* (MIM#608937)) in LCLs from eight carriers of the 220 kb BP2-BP3 deletion, eight carriers of the 600 kb BP4-BP5 deletion and 10 control individuals. In particular, we identified a significant diminution of the hemizygote gene *SH2B1* but not of the neighboring normal-copy *KCTD13*, *MVP* and *MAPK3* in BP2-BP3 deletion carriers.

Viewpoint selection

We used an adaptation of the 4C method,[51–53] the high-resolution Chromosome Conformation Capture Sequencing technology (4C-seq),[54] to identify chromosomal regions that physically associate with the promoters of *MVP, KCTD13, ALDOA, TBX6* (MIM#602427) and *MAPK3*, five of the 28 'unique' genes of the BP4-BP5 interval selected according to their potential role in the described phenotype. Reduction by ~50% of the RNA levels of the ortholog of *ALDOA* (Aldolase A) was associated with a change in brain morphology in zebrafish, suggesting that this gene is dosage sensitive.[55] In humans, recessive *ALDOA* deficiency is associated with glycogen storage disease XII (OMIM#611881).[56] Morpholino-driven reduction of the expression level of the *KCTD13* (Potassium Channel Tetramerization Domain containing protein 13) ortholog resulted in macrocephaly in zebrafish, while its depletion in the brain of mouse embryos resulted in an increase of proliferating cells. The mirroring microcephaly was seen upon overexpression of human *KCTD13* cDNA in zebrafish embryos' heads, a phenotype further amplified upon concomitant overexpression of either *MAPK3* (mitogen-activated protein kinase 3) or *MVP* (Major Vault Protein).[27] *TBX6* (T-Box Transcription Factor 6) is a candidate gene for the vertebral malformations observed in some deletion carriers since (i) mice homozygous for a *Tbx6* mutation showed rib and vertebral body anomalies;[57] (ii) *TBX6* polymorphisms were associated with congenital scoliosis in the Han population;[58] (iii) a stoploss variant in *TBX6* segregates with congenital spinal defects in a three-generation family[59] (OMIM#122600); and (iv) carriers of 16p11.2 600 kb BP4-BP5 deletions and a common hypomorphic *TBX6* allele suggest a compound inheritance in congenital scoliosis.[60] *TBX6* was selected as a viewpoint even though this gene is not expressed in LCLs (or only at extremely low level), as studies have shown that the contacted domains are stable across cell lines and tissues regardless of expression status.[3] Within the group of genes chromatin-contacted by the above viewpoints we selected two more viewpoints within the 16p11.2 220 kb BP2-BP3 region (Figure 1), that is, the promoters of *SH2B1* and *LAT*. The *SH2B1* gene was suggested to be a crucial candidate for the obesity phenotype associated with this genomic interval[16,28] as it encodes an Src homology adaptor protein involved in leptin and insulin signaling.[61,62] Common variants in this locus were repeatedly associated with BMI, serum leptin and body fat in genome-wide association studies,[63–66] while rare dominant mutations were reported to cause obesity, social isolation, aggressive behavior, and speech and language delay.[67] In a recent large-scale association study, the deletion was also significantly linked with SCZ.[35] The *LAT* (linker for activation of T cells) adaptor molecule participates in AKT activation and plays an important role in the regulation of lymphocyte maturation, activation and differentiation.[68,69] Its inactivation could be circumvented by Ras/MAPK constitutive activation.[70]

4C-seq

4C libraries were prepared from LCLs of two control individuals and two carriers each of the 16p11.2 BP4-BP5 deletion and duplication, sex- and age-matched (Supplementary Table S3). Briefly, LCLs were grown at 37°C. 5×10^7 exponentially growing cells were harvested and crosslinked with 1% formaldehyde, lysed and cut with DpnII, a 4-cutter restriction enzyme that allows higher resolution.[53] After ligation and reversal of the crosslinks, the DNA was purified to obtain the 3C library. This 3C library was further digested with *Nla*III and circularized to obtain a 4C library. The inverse PCR primers to amplify 4C-seq templates were designed to contain Illumina adaptor tails, sample barcodes and viewpoint-specific sequences. Viewpoints were selected at the closest suitable DpnII fragment relative to the

transcriptional start sites of the targeted genes. The sequence of the 4C-seq primers is reported in Supplementary Table S4. For all viewpoints, we amplified at least 1.6 μg of 4C template (using about 100 ng of 4C template per inverse PCR reaction, for a total number of 16 PCRs). We multiplexed the 4C-seq templates in equimolar ratios and analyzed them on a 100-bp single-end Illumina HiSeq flow cell. The numbers of raw, excluded, and mapped reads for each viewpoint and LCL sample are detailed in Supplementary Table S5.

4C-seq data analysis

4C-seq data were analyzed as described in Noordermeer *et al.*[53] and Gheldof *et al.*[54] through the 4C-seq pipeline available at http://htsstation.epfl.ch/)[71] and visualized with gFeatBrowser. Briefly, the multiplexed samples were separated, undigested and self-ligated reads removed. Remaining reads were aligned and translated to a virtual library of DpnII fragments. Read counts were then normalized to the total number of reads and replicates combined by averaging the resulting signal densities (Supplementary Figure S1). The local correlation between the profiles of the two samples per viewpoint was calculated ($0.46 \leqslant r^2 \leqslant 0.74$ for controls, $0.29 \leqslant r^2 \leqslant 0.67$ for deletions and $0.22 \leqslant r^2 \leqslant 0.61$ for the duplications). The combined profiles were then smoothed with a window size of 29 fragments. The region directly surrounding the viewpoint is usually highly enriched and can show considerable experimental variation, thereby influencing overall fragment count. To minimize these effects, the viewpoint itself and the directly neighboring 'undigested' fragment were excluded during the procedure. In addition to this filtering, we modeled the data to apply a profile correction similar to the one described in Tolhuis *et al.*[72] using a fit with a slope −1 in a log–log scale.[73] Significantly interacting regions were detected by applying a domainogram analysis as described.[74] We selected BRICKS (Blocks of Regulators In Chromosomal Kontext) with a *P*-value threshold smaller than 0.01 for both '*cis*' and '*trans*' interactions. To determine differentially interacting regions between the 16p11.2 600 kb BP4-BP5 deletion (Del), duplication (Dup) and control (Ctrl), we considered all non-null BRICKS found by a domainogram analysis[74] in either condition and quantified both signals in each BRICK. The resulting table was scaled to the sample with the largest interquartile range and the difference of signals was compared with random in order to associate a *P*-value (FDR) with each BRICK. Finally, only BRICKS with a *P*-value < 0.01 were considered.

All the viewpoints mapping on the BP4-BP5 interval, except *KCTD13*, contact the 146 and 147 kb long low copy repeats that flank the 16p11 600 kb BP4-BP5 rearrangements. To unravel whether the signal was reflecting the interaction with the centromeric, the telomeric or both low copy repeats and given the high similarity (99.5% identity) of the two blocks, we separately treated the reads mapping within these regions (chr16: 29460515–29606852 and chr16: 30199854–30346868 according to GRCh 37/*hg19* assembly, February 2009) using different and more stringent criteria, that is, no mismatch and unique site mapping. All values were normalized to the total number of reads mapping to the two regions (per thousands of reads). We observed a higher proportion of contacts occurring with the centromeric segmental duplication compared with the telomeric one for *MAPK3* and *TBX6*, while the trend was reversed for *MVP*, in agreement with their proximity to the centromeric and telomeric low copy repeat blocks, respectively. No conclusive results were obtained for *ALDOA*.

Hi-C data

Hi-C matrices from Rao *et al.*[75] were prepared by first applying a KR normalization to the 5 and 100 kb resolution observed matrices, and then by dividing each normalized score by the expected one extracted from the KR expected file (as described in section II.c of the Extended Experimental Procedures of reference[75]). KR expected values less than 1 were set to 1 to avoid long-distance interaction biases.

Enrichment analyses

Gene annotation was obtained through BioScript (http://gdv.epfl.ch/bs). Protein interaction networks for the genes selected by BRICKS calling and from the list of interacting regions affected by the rearrangements were determined using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v9.1 (http://string-db.org/).[76] We exploited Bioscript for Gene Ontology analysis (topGO) (http://gdv.epfl.ch/bs), DAVID GO and KEGG, OMIM Disease and KEGG coupled with Enrichr to assess if the chromatin-contacted genes were enriched in specific pathways and genes associated with Mendelian diseases[77–80] (http://www.omim.org/downloads). The OMIM gene-set library was obtained directly from the NCBI's OMIM Morbid Map.[81] We exploited the SFARI Gene lists and scores (https://sfari.org/; March 2014 release), the union of the genes cataloged in Girard *et al.*[82] and Xu *et al.*[82–84] and the genome-wide association studies hits for BMI[85] to assess enrichment for ASD, SCZ and BMI genes, respectively. We also used the *de novo* 'high confidence' ASD targets (selected with FDR < 0.1 in De Rubeis *et al.*[86] and likely gene disrupting recurrent mutations target genes in Iossifov *et al.*[87]) to assess enrichment of ASD-associated genes. Ciliary genes enrichment was computed merging the SYSCILIA gold standard (SCGS) and potential ciliary gene lists (genes with no additional evidence for ciliary function were excluded).[88] We used Enrichr Chromosome Location tool and BRICKS count in different window sizes (5 Mb, 1 Mb and 500 kb) to determine whether any cytogenetic band other than 16p11.2 was enriched for BRICKS. Other than 16p11.2, we identified enrichments at 16p12, 16p13, 16q13, 1p36, 11q13, 16q22, 7q31, 15q15 and 1q32 cytobands. As a large proportion of the 16p12.2 BRICKS maps to segmental duplications that are highly similar to the low copy repeats flanking the 16p11 600 kb BP4-BP5 rearrangements we conservatively did not consider this region.

All seven tested viewpoints showed enrichment for contacts with loci that encode proteins that interact together (all $P < 0.01$). A single process, focal adhesion assembly, was shared between the BP2-BP3 and the BP4-BP5 groups of viewpoints (GO:0048041, $P = 6.02e^{-03}$ for the BP2-BP3 group and $P = 1.12e^{-03}$ for the BP4-BP5). Focal adhesion links the internal actin cytoskeleton to the extracellular matrix; it is used by cells to explore their environment, and depends strongly on microtubule dynamics[89] in coordination with the primary cilium.[90,91] Genes participating in focal adhesion (*GSK3B* (MIM#605004)*; PAK7* (MIM#608038)), axon guidance (*ROBO1* (MIM#602430)*; EPHA6* (MIM#600066)*; PAK7; GSK3B*) and Golgi apparatus-related processes (*SND1* (MIM#602181)*; FRMD5* (MIM#616309)) are among the 24 genes *trans*-contacted by both the BP4-BP5 and BP2-BP3 viewpoints. The gene *MPZL1* (MIM#604376), contacted by both *KCTD13* and *SH2B1*, is associated with SCZ.[92] It is a downstream target of *PTPN11*.[93]

Fluorescence *in situ* hybridization

Interphase nuclei were prepared from a LCL of a control individual. Fluorescence *in situ* hybridization experiments were performed using fosmid clones (300 ng) directly labeled by nick-translation with Cy3-dUTP and fluorescein-dUTP as previously described[94] with minor modifications. Fosmid and BAC clones (G248P86150B3 for the *ALDOA* locus, G248P800063B6 for the *SH2B1* locus; G248P86115A10 for the *KIAA0556* locus; RP11-301D18 for the *KCTD13* and *MVP* loci; RP11-383D9 for *PTEN* (MIM#601728); RP11-477N2 for *USP34* (MIM#615295)/*XPO1* (MIM#602559) and RP11-43E18 for *MARK4* (MIM#606495)) were obtained from the CHORI BACPAC Resources Center (https://bacpac.chori.org/). We picked the *MARK4*-encompassing BAC as 'control BAC' because it maps to a gene-rich region on chromosome 19, a centrally-positioned chromosome within the nucleus. Hybridization was performed at 37 °C in 2 × SSC, 50% (v/v) formamide, 10% (w/v) dextran sulfate, 3 μg C0t-1 DNA and 3 μg sonicated salmon sperm DNA in a volume of 10 μl. Post-hybridization washing was at 60 °C in 0.1 × SSC, three times. Nuclei were DAPI-stained and digital images were obtained using a Zeiss Imager A1 fluorescence microscope (Carl Zeiss, Oberkochen, Germany). We considered 50–60 cells per experiment (i.e. at least 100 distances) and co-localization was defined if the distance between signals was ⩽ 0.3 um. The contact between *SH2B1* and *ALDOA*, versus the control *KIAA0556,* was estimated by calculating the distance between BAC probes (median *SH2B1-ALDOA* and *-KIAA0556* distances = 0.43 and 1.24 μm, respectively; Wilcoxon rank-sum test, $P = 1.45e^{-17}$). The contact of *MVP* and *KCTD13* with, respectively, *PTEN* and *USP34/XPO1*, compared with the control *MARK4,* was estimated as the percentage of co-localization (25 and 14% co-localization versus 2% with the control locus; Fisher's test enrichment: $P = 6.9e^{-05}$ and $P = 0.01$, respectively; median *MVP/KCTD13-USP34/XPO1* distances = 1.76, *MVP/KCTD13-PTEN* = 2.61 and *MVP/KCTD13-MARK4* = 4.96 μm; Wilcoxon rank-sum test, $P = 5.4e^{-10}$ and $P = 9.3e^{-05}$, respectively).

ChIP-seq and RNA-seq molecular associations

Detailed experimental procedures and results are described in Waszak *et al.*[95] Briefly, ChIP-seq (chromatin immuno-precipitation coupled with sequencing) and mRNA-seq data were produced from LCLs of 54 individuals of European origin from the 1000 Genomes Project.[96] ChIP-

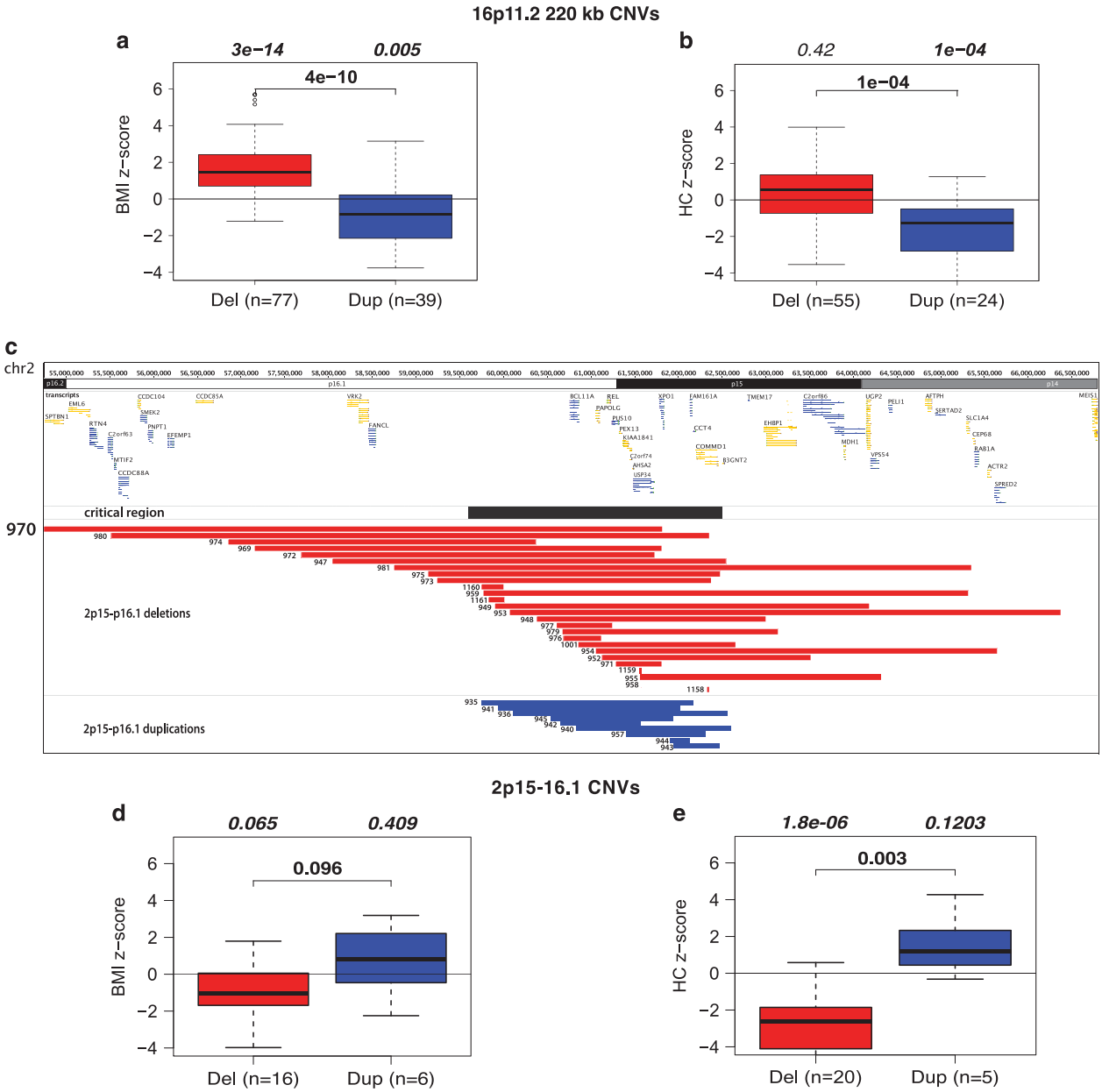

Figure 2. Phenotypic characterization of carriers of 16p11.2 BP2-BP3 and 2p15 rearrangements. Distribution of *Z*-score values of BMI (**a**) and head circumference (**b**) in unrelated carriers of the 16p11.2 220 kb BP2-BP3 deletion (red) and duplication (blue) taking into account the normal effect of age and gender observed in the general population as described in Jacquemont *et al*.[18] The general population has a mean of zero. (**c**) Comparison of the genomic breakpoints of 2p15 deletions (red bars) and duplications (blue bars) in 26 and 9 unrelated carriers, respectively. The breakpoints' coordinates are detailed in Supplementary Table S2. The genes mapping within the interval and cytobands' positions are shown above, while the extent of the critical region is indicated by a black bar. Distribution of *Z*-score values of BMI (**d**) and head circumference (**e**) in carriers of the 2p15 deletion (red) and duplication (blue).

seq with antibodies recognizing H3K4me1, H3K4me3, H3K27ac, PU.1 and RNA Pol2 binding, as well as mRNA-seq gene expression profiling, were carried out from a single growth of LCLs as previously described.[97] Genotypes were obtained from the GEUVADIS consortium.[98] To map associations between pairs of ChIP-seq and/or RNA-seq peaks, we retained 47 individuals after data quality control and proceeded as follows for each of the 15 possible unordered pairs of distinct molecular phenotypes (A_1, A_2). First, we measured inter-individual Pearson correlation between every possible pair of normalized quantifications at peaks (p_1, p_2) within the 16p11.2 interval (28.1–34.6 Mb) such that p_1 and p_2 belong to A_1 and A_2, respectively. Note that the distances here were measured between the respective peak centers, excepted for mRNA for which we used the transcription start site. Then, we assessed to what extent the correlations significantly differed from zero by calculating *P*-values using the R function *cor.test* and corrected them for multiple-testing by using the Benjamini and Hochberg procedure as implemented in the R function *p.adjust* (FDR 5 and 10%).

RESULTS

Distinct and non-overlapping loci at 16p11.2 are associated with mirror phenotypes on BMI and HC and autism susceptibility

To comprehensively assess phenotypic features associated with the distal 16p11.2 220 kb BP2-BP3 CNVs (Figure 1), we collected de-identified data on 137 unrelated carriers (88 deletions and 49

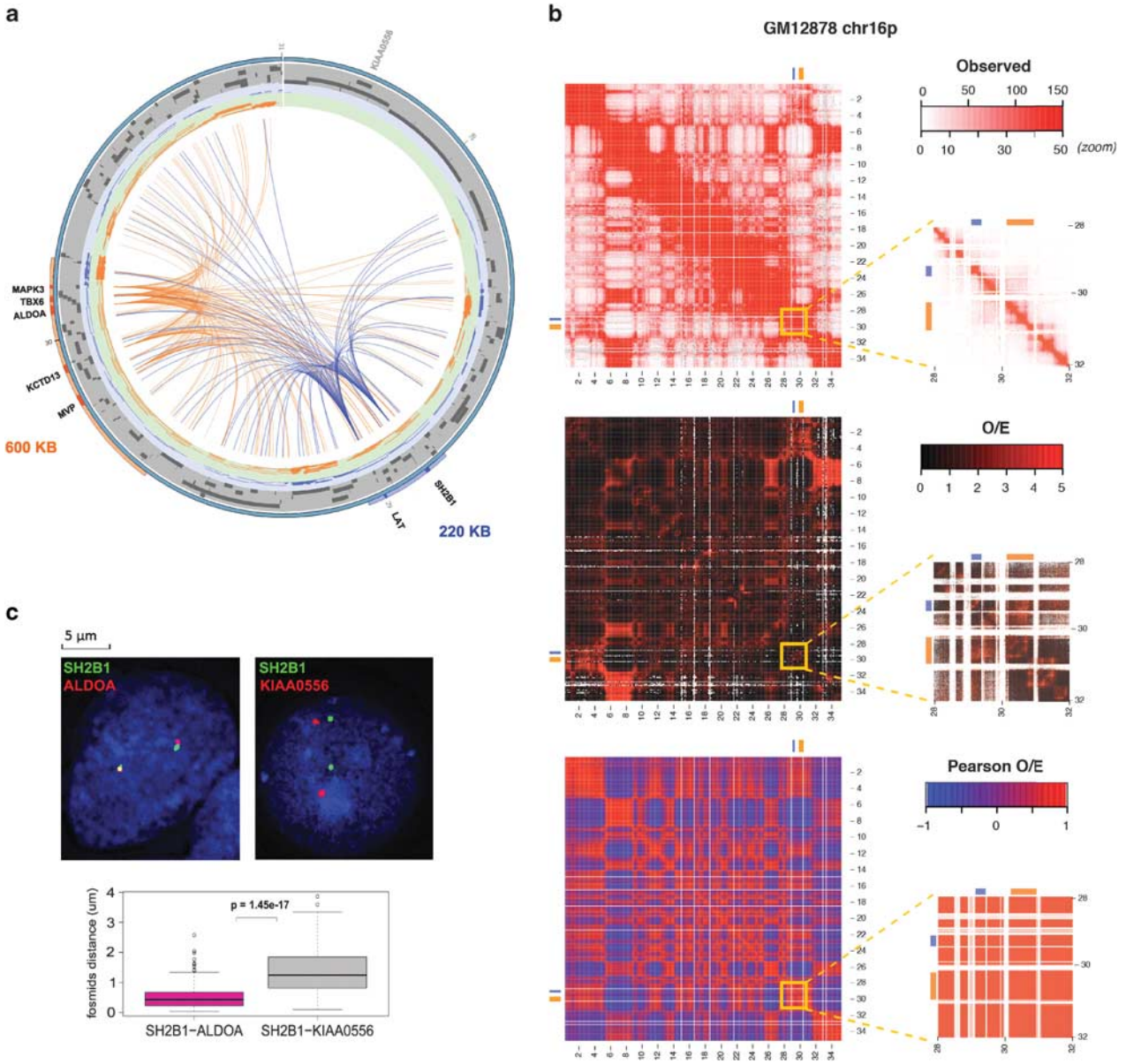

Figure 3. Chromatin interactions between the 16p11.2 600 kb BP4-BP5 and 220 kb BP2-BP3 genomic intervals. (**a**) Circos plot representation of the chromatin loops identified in the human chromosome 16 27.5–31.0 Mb window. The 220 kb BP2-BP3 and 600 kb BP4-BP5 intervals are depicted by blue and orange bars on the peripheral circle, respectively. Darker sections indicated the positions of the viewpoints. Central blue and orange lines indicate the chromatin interactions corresponding to BP2-BP3 and BP4-BP5 viewpoints, respectively. Note the quasi absence of loops between the BP2-BP3 viewpoints (*LAT* and *SH2B1*) and the 27.5-28.4 Mb region. The mapping position of the *KIAA0556* gene, used as control locus in fluorescence *in situ* hybridization experiments, is indicated. (**b**) High-resolution Hi-C chromosome conformation capture results obtained in reference[75] with the GM12878 LCL within the chromosome 16 0–34 Mb window (left panels) and zoom in within the 28–31 Mb region encompassing the two CNVs (5 kb resolution; right panels). The positions of the 220 kb BP2-BP3 and 600 kb BP4-BP5 intervals are shown by blue and orange bars, respectively. Observed (top panel), observed/expected (central panel) and Pearson correlation results are presented (bottom panel). (**c**) Fluorescence *in situ* hybridization experiments show colocalization of *SH2B1* foci (green) that map to the 220 kb BP2-BP3 interval with *ALDOA* foci (red) that map to the 600 kb BP4-BP5 genomic interval (left panel) but not with the equidistant *KIAA0556* (red) foci (central panel). The distribution of interphase nuclei distances between the *SH2B1* and *ALDOA* (deep pink) and *SH2B1* and *KIAA0556* foci (gray) are shown in the lower panel. The mapping positions of *ALDOA*, *SH2B1* and *KIAA0556* are indicated in (**a**)**.**

duplications; Supplementary Table S1) and compared BMI and HC with gender-, age- and geographical location-matched reference population as described[18] (Figures 2a and b). The BMI mean *Z*-score of deletion carriers deviated significantly from that of the general population (*t*-test, $P = 3.1e^{-14}$), replicating the earlier described association of the deletion with obesity.[16,28] We observed a trend towards increased HC in deletion carriers. The duplication carriers showed a mirroring decrease of BMI and HC values when compared with those of the control population (*t*-test, $P = 0.005$ and $1.1e^{-4}$, respectively). We also observed an increase in ASD prevalence in both deletion (23/88; 26%) and duplication (11/49; 22%) carriers compared with the general population (5,338/363,749; 1.5%)[99] (Fisher's enrichment test: OR = 23.7, $P = 2.5e^{-22}$; OR = 19.4, $P = 1.2e^{-10}$) in agreement with

published results.[29–35] Thus, genomic rearrangements at 600 kb BP4-BP5 and 220 kb BP2-BP3, two loci 650 kb apart, present similar clinical patterns: large effect sizes on BMI and HC, as well as association with ASD.

Cis-acting chromatin loops that link the 16p11.2 BP4-BP5 and BP2-BP3 genomic intervals are perturbed in BP4-BP5 CNV carriers

We posited that the remarkable overlap of phenotypic features associated with the BP2-BP3 and BP4-BP5 CNVs might derive from the rearrangement-mediated disruption of the 3D chromatin structure within the 16p11.2 cytoband. To challenge this hypothesis, we assessed the pattern of chromosomal interactions of selected 'viewpoints' from both loci in two LCLs derived from control individuals using an adapted version of the 4C method (4C-seq: circularized chromosome conformation capture combined with multiplexed high-throughput sequencing)[51,52,54,100] (Materials and Methods, Supplementary Table S3). Despite the limitations of the study of LCLs (Materials and Methods), these experiments are worth pursuing as studies have shown that chromatin contacts are stable across cell lines and tissues regardless of contacted-gene expression status[3] and that LCL transcriptome profiles can be recapitulated in other tissues and species.[12] Specifically, our previous analyses of LCL transcriptomes showed that genes whose expression correlated with the dosage of the 16p11.2 locus are significantly enriched in genes associated with ASD and ciliopathies both in human LCLs and mouse cortex.[12] In particular, we identified chromosomal regions that physically associate with the promoters of *MVP, KCTD13, ALDOA, TBX6* and *MAPK3,* five genes mapping to the BP4-BP5 interval, and *SH2B1* and *LAT,* two genes mapping to the BP2-BP3 one. These were investigated based on their potential role in the phenotype (Supplementary Figure S2).[27,55,57–59]

Genome-wide we identified an average of 265 BRICKs (FDR $\leqslant$ 1%), i.e., three-dimensionally interacting genomic fragments, for the seven viewpoints (range: 168–442; Supplementary Tables S6–S12). In particular, we observed complex chromatin looping between genes located in the proximal BP4-BP5 and those mapping both to the distal BP2-BP3 region and the equidistant downstream region rich in Zn-finger genes (Figures 1 and 3a). For instance each of the nine genes of the BP2-BP3 interval (*ATXN2L* (MIM#607931), *TUFM* (MIM#602389), *SH2B1, ATP2A1* (MIM#108730), *RABEP2* (MIM#611869), *CD19* (MIM#107265), *NFATC2IP* (MIM#614525), *SPNS1* (MIM#612583) and *LAT*) is contacted by at least one of the five assessed viewpoints in the BP4-BP5 interval (Figures 1 and 3a). We reciprocally validated these chromatin interactions using the promoters of *SH2B1* and *LAT* as viewpoints (e.g. the chromatin loops of *MVP, KCTD13, ALDOA, TBX6* and *MAPK3* with *SH2B1* are all recapitulated using *SH2B1* as viewpoint; Figure 3a and Supplementary Tables S11 and S12). The preferential contacted domain of the BP2-BP3 viewpoints extends proximally to the BP4-BP5 and Zn-finger gene-rich regions (Figures 3a and b and Supplementary Figure S3). Inversely, significantly less interactions are called in the gene-rich and equidistant distal region (*t*-test $P = 0.011$, Supplementary Figure S3), suggesting that these interactions do not merely reflect the spatial clustering of gene-dense regions.

We confirmed the genomic interaction between the 600 kb BP4-BP5 and 220 kb BP2-BP3 intervals using fluorescence *in situ* hybridization. This independent method showed that the BP2-BP3-mapping *SH2B1* locus was significantly closer to the BP4-BP5-encompassed *ALDOA* locus than to a control region, the *KIAA0556* locus, situated equidistantly on its telomeric side (median *SH2B1-ALDOA* and *SH2B1-KIAA0556* distances = 0.43 and 1.24 μm, respectively; Wilcoxon rank-sum test, $P = 1.45e^{-17}$) (Figure 3c, Supplementary Figure S4). We also examined published Hi-C (genome-wide conformation capture) and high-resolution Hi-C data from LCLs. Although they cannot confirm our chromatin connections given their limited resolution, they support a preferential three-dimensional proximity of these two regions[73,75] (Figure 3b). Concordant results were found in both human IMR90 fibroblasts and embryonic stem cells[26,101] and mouse cortex and embryonic stem cells, suggesting the conservation of the structure of this topological-associated domain across species and tissues,[101] as recently shown for the topological-associated domain spanning the *WNT6/IHH/EPHA4/PAX3* (MIM#604663; #600726; #602188; #606597) locus.[3]

As chromatin interactions were determined in normal diploid context, we next assessed the effect of BP4-BP5 CNVs on these chromatin loops. We identified genomic fragments that interact with the same seven viewpoints in LCLs of two BP4-BP5 deletion patients and two reciprocal duplication patients (Materials and Methods). We observed a genome-wide decrease in the number of BRICKS per viewpoint ranging from 27 to 84%, suggesting that both rearrangements triggered dramatic reorganizations. Consistent with this hypothesis, the *SH2B1* viewpoint, whose copy number is not affected by the proximal BP4-BP5 CNV, shows a 36% reduction in the amount of interacting regions (all BRICKS listed in Supplementary Tables S13–S26). We compared the 4C-seq results from control individuals and the four patients and identified, across all conditions and considering all viewpoints, 1193 genes with significantly modified chromosomal contacts (FDR $<$ 1%; Materials and Methods, Supplementary Table S27). These results support the idea that large structural rearrangements perturb the 3D genomic structure by modifying both *cis* and *trans* contacts.

Perturbations of the chromatin interactions' landscape at 16p11.2 are associated with gene expression modification

Our results show that the gene-rich BP2-BP3 and BP4-BP5 16p11.2 intervals, whose CNVs are linked to overlapping phenotypes, are reciprocally engaged in complex chromatin looping as determined by 4C, fluorescence *in situ* hybridization and Hi-C. The recent discovery of multigene complexes where chromosomal loops orchestrate co-transcription of interacting genes[2,102] is suggestive of functional implications for the chromosomal contacts between the BP2-BP3 and BP4-BP5 intervals.

To assess this possibility, we first used our recent association analyses of population-wide transcription factor DNA binding (PU.1 and RPB2—the second largest subunit of RNA polymerase II), histone modification enrichment patterns (H3K4me1, H3K4me3 and H3K27ac) and gene expression measured by ChIP-seq and RNA-seq in LCLs derived from 47 European unrelated individuals whose genomes were sequenced in the frame of the 1000 Genomes Project.[95] We measured the extent of quantitative coordination of natural inter-individual variation between pairs of these six molecular phenotypes at putative regulatory regions mapping within cytoband 16p11.2 and identified coordinated behavior in terms of mapping enrichment. For example, we found association between active regulatory regions mapping within the BP4-BP5 interval and expression of BP2-BP3 genes (Supplementary Figure S5),[95] consistent with the notion that some of the chromatin loops uncovered between these two intervals might bring together regulatory elements and genes.

Secondly, we examined if genes involved in primary cilium function and related pathways, which are modified in BP4-BP5 deletion patients' cells,[12] are also changed in BP2-BP3 deletion carriers (Materials and Methods). We found that the ciliary genes *BBS4, BBS7, SMAD2*, *XPOT* and *NUP58* are correspondingly modified in LCLs derived from both BP4-BP5 and BP2-BP3 deletion carriers (Supplementary Figure S6). The interplay between the 600 kb BP4-BP5 and the 220 kb BP2-BP3 interval is further substantiated: (i) by published data showing perturbed expression of genes mapping within the BP2-BP3 distal interval (i.e. *LAT, SPNS1* and *ATP2A1*) in cells derived from BP4-BP5 patients;[26] as well as (ii) by the observation that, within the top-10 genes correlated with *SH2B1*

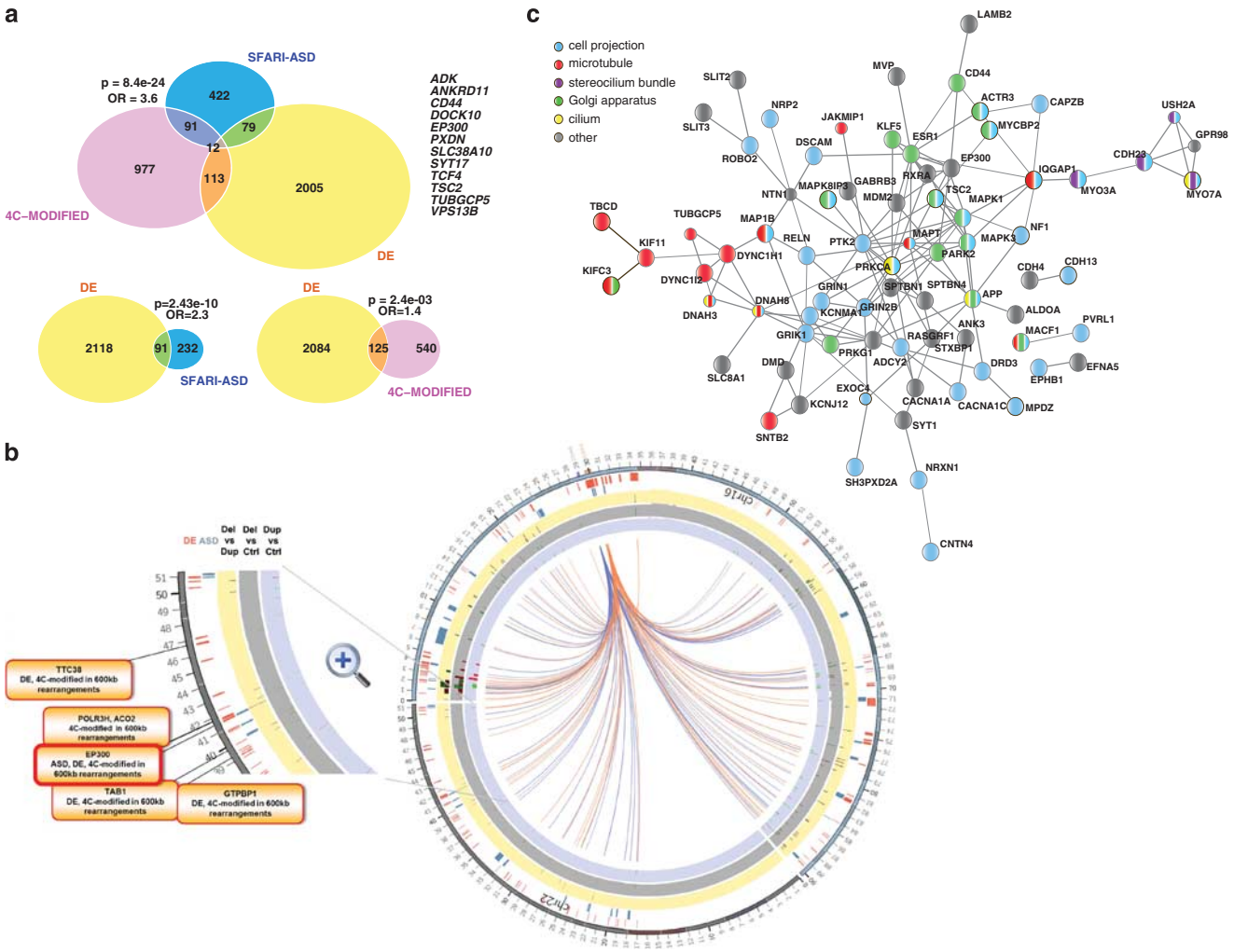

Figure 4. Extensive overlap between differentially expressed genes and loci that show modified chromatin interactions. (**a**) Top panel: weighted Venn diagram showing the overlap between the 2209 genes that are differentially expressed in 16p11.2 rearrangement carriers (DE, yellow disk; FDR $\leqslant$ 5%[12]), the 1193 genes that show modified chromatin interactions in 16p11.2 rearrangement carriers (4C-modified, purple disk; only 665 with detectable expression are considered for the DE enrichment; see Supplementary Table S31) and the 604 genes listed in SFARI Gene (https://sfari.org/; 323 expressed), an annotated list of candidate genes for ASD (ASD; blue disk). The numbers of common genes are indicated and the 12 4C-modified, DE and ASD-SFARI genes are specified on the right. Bottom panels: weighted Venn diagrams showing the overlap between the DE genes and the LCLs-expressed ASD and 4C-modified genes (lower left and right, respectively). (**b**) Circos plot representation of the modified chromatin loops identified in human chromosomes 16 and 22 (right-hand panel). The 220 kb BP2-BP3 and 600 kb BP4-BP5 intervals are depicted by blue and orange bars on the peripheral circle, respectively. Central blue and orange lines indicate the CNVs-modified chromatin interactions corresponding to BP2-BP3 and BP4-BP5 viewpoints, respectively. Ticks on the three internal rings indicate BRICKS with significantly modified interactions between 16p11.2 600 kb BP4-BP5 duplication and control samples (light blue ring), between 16p11.2 600 kb BP4-BP5 deletion and control samples (dark gray), and between 16p11.2 600 kb BP4-BP5 deletion and duplication samples (yellow). Blue and red ticks on the most external rings denote genes differentially expressed in 16p11.2 patients (DE) and SFARI-ASD-associated genes (ASD), respectively. A zoomed-in view with examples of genes with modified chromatin interactions mapping within the 22q13 cytoband is presented in the left-hand panel. (**c**) The 1193 genes that show modified chromatin interaction in 16p11.2 cells with 16p11.2 600 kb BP4-BP5 rearrangements encode proteins that interact. The confidence view interaction network of the encoded proteins corresponding to the enriched GO terms (GO:0030030 cell projection organization, GO:0042995 cell projection, GO:0030173 integral to Golgi membrane, GO:0000138 Golgi trans cisterna, GO:0034504 protein localization to nucleus and GO:0005874 microtubule) is visualized with STRING. Proteins belonging to cell projection (blue), microtubule (red), Golgi apparatus (green), stereocilium bundle (purple) and cilium (yellow) process/cell components are highlighted by colored beads. Disconnected nodes are not shown. FDR, false discovery rate.

expression according to GeneProf (www.geneprof.org), three (*ZNF500*; *CDAN1* (MIM#607465); *LRRC14*) are contacted by viewpoints located in the BP4-BP5 region and two (*PIGO* (MIM#614730); *TUBGCP6* (MIM#610053)) are differentially expressed in BP4-BP5 CNV patients.

We then assessed whether the structural chromatin changes identified in the 600 kb BP4-BP5 CNVs carriers (i.e. 1193 genes with significantly modified chromosomal contacts, FDR < 1%) are paralleled by transcriptome modifications identified in Migliavacca *et al.*[12] (2209 significantly differentially expressed genes, FDR $\leqslant$ 5%; uniquely mapping probes)). We find a significant overlap between the two genes lists as 125 genes with modified chromatin loops are concomitantly differentially expressed in 16p11.2 600 kb BP4-BP5 CNV carriers' cells (125/665 4C-modified genes with detectable expression (see Materials and Methods); Fisher's enrichment test: OR = 1.4, $P = 0.002$; Supplementary Tables S28–S30 and Figure 4a). There is no correlation between quality of the change in 4C contact (increased or decreased) and the sense

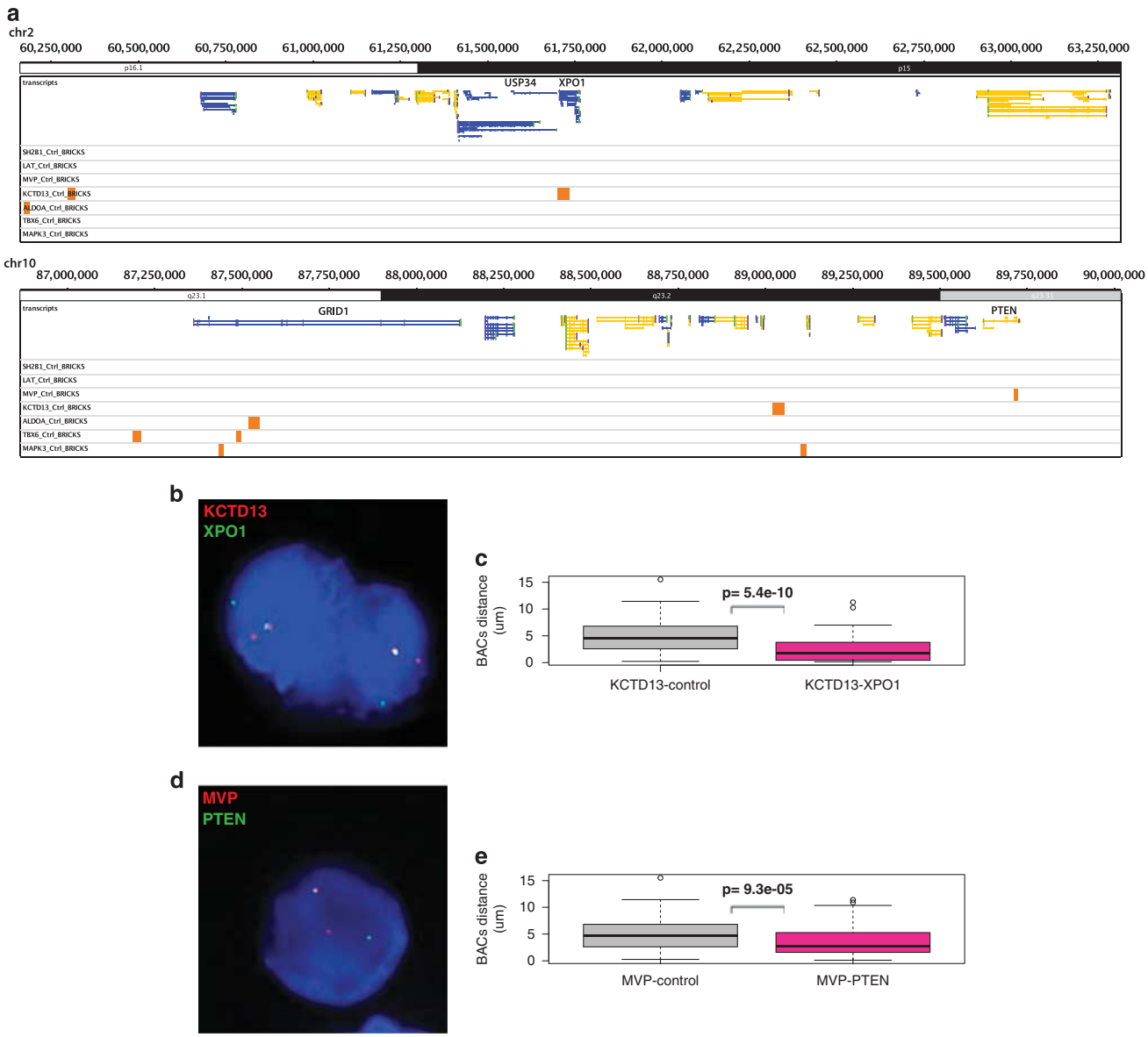

Figure 5. Examples of 16p11.2 viewpoints chromatin-contacted regions. (**a**) Examples of regions (BRICKS) interacting with 16p11.2 viewpoints showing some of the contacted genes, that is, *GRID1*, *PTEN* and *USP34/XPO1*. Other examples (*CHD1L* and *EP300*) are shown in Supplementary Figure S8). Fluorescence *in situ* hybridization experiments show colocalization of the 600 kb BP4-BP5 interval-encompassed *KCTD13* (red) and 2p15-mapping *XPO1* foci (green) (**b**) and the 600 kb BP4-BP5 interval-encompassed *MVP* (red) and 10q23.31-mapping *PTEN* foci (green) (**d**). The distribution of interphase nuclei distances between *KCTD13* and *XPO1* (**c**) and between *MVP* and *PTEN* (**e**) foci are compared with to those between *KCTD13/MVP* and *MARK4* (control) foci (25 and 14% co-localization versus 2% with the control locus; Fisher's test enrichment: $P = 6.9e^{-05}$ and $P = 0.01$, respectively; median *MVP/KCTD13-USP34/XPO1* distances = 1.76, *MVP/KCTD13-PTEN* = 2.61 and *MVP/KCTD13-MARK4* = 4.96 μm; Wilcoxon rank-sum test, $P = 5.4e^{-10}$ and $P = 9.3e^{-05}$, respectively).

of the perturbation in gene expression, that is, a loss of contact does not necessarily imply a decrease in expression and vice versa (Supplementary Figure S7 and Supplementary Table S31).

These results show that the BP2-BP3 and BP4-BP5 loci are linked by chromatin loops, coordinated molecular phenotypes and co-regulation of genes.

Genomic regions contacted by 16p11.2 viewpoints are associated with autism, BMI and HC phenotypes and enriched in ciliary genes

Consistent with the notion that chromatin contacts connect biologically related genes, BRICKs genes are enriched for genes that encode proteins that interact ($P < 0.01$ for all viewpoints; Supplementary Table S32, Materials and Methods). Processes overrepresented within BRICKs genes are listed in Supplementary Table S33. They are also enriched for genes listed in SFARI Gene (https://sfari.org/), an annotated list of candidate genes for ASD (union of the SFARI Syndromic, High Confidence and Strong Candidate Gene categories (Categories S+1-2), 13/76, OR = 2.9, $P = 0.0014$) and for ASD-associated genes identified by whole-exome studies (8/50 of the 'high confidence *de novo*' ASD-associated genes;[86,87] OR = 2.7, $P = 0.016$; Supplementary Table S34). The BP4-BP5 and BP2-BP3 viewpoints contacted genes include *GRID1* (MIM#610659) and *PTEN* at 10q23.2–q23.31, *USP34/XPO1* at 2p15, but also genes linked to HC phenotypes like *CHD1L*

(MIM#613039) at 1q21.1[103–112] and *EP300* (MIM#602700) at 22q13[113] (Figure 5a, Supplementary Figure S8 and Supplementary Table S35). To validate these interactions, we verified a subset by fluorescence *in situ* hybridization (e.g. *MVP-PTEN* and *KCTD13-USP34/*XPO1; Figures 5b–e and Materials and Methods). The comparison of distributions of Hi-C scores in selected versus non-selected BRICKS for each of our 4C viewpoints (Materials and Methods) further demonstrates the reproducibility of the 4C results (Supplementary Figure S9).

Reminiscent of our transcriptome findings,[12] enrichment analysis of the 1193 genes with modified chromosomal contacts showed significant over-representation of ciliary genes[88] (40/493, OR = 1.47, $P = 0.030$, Supplementary Tables S29–30), OMIM terms associated with dysfunction of ciliary structures (Supplementary Table S28) and candidate genes for ASD (Supplementary Tables S29–S30). We showed previously[12] that differentially expressed genes were similarly enriched for SFARI-ASD-associated genes (91/323 with detectable expression; OR = 2.3, $P = 2.43e^{-10}$; Figure 4a). Notably five (*TCF4* (MIM#602272), *EP300, ADK* (MIM#102750), *TUBGCP5* (MIM#608147), *VPS13B* (MIM#607817)) of the 12 genes that are concurrently SFARI-ASD, differentially expressed and modified in 4C contacts (OR = 5.01, $P = 1.5e^{-05}$) were previously associated with head circumference changes[114–117] (Figures 4a and b and Supplementary Tables S29 and S30).

Phenotypes associated with 2p15-16.1 CNVs

Our results suggest that chromatin interactions captured by the BP4-BP5 and BP2-BP3 viewpoints and their perturbations can be exploited to identify additional genes/loci, which when genetically perturbed, are associated with similar pathways, diseases and phenotypes. We challenged this hypothesis by assessing whether the phenotypic features of 2p15-p16.1 deletion and duplication carriers overlap with those of carriers of the 600 kb BP4-BP5 and 220 kb BP2-BP3 rearrangements. Haploinsufficiency of the chromatin-contacted *USP34/XPO1* was suggested to be responsible for the 2p15-p16.1 deletion syndrome (MIM#612513) phenotypes that included intellectual disability, ASD, microcephaly, dysmorphic facial features and a variety of congenital organ defects.[111,112]

We collected clinical data on 26 and 9 unrelated 2p15-p16.1 deletion and duplication carriers, respectively (Figure 2c and Supplementary Table S2). Comparison of data on BMI and HC of both variants pinpoints mirror phenotypes for these two traits (Figures 2d and e). Whereas we do not formally demonstrate a mechanistic and functional link between the 600 kb BP4-BP5 and the 2p15 interval, it should be noted that the *KCTD13–USP34/XPO1* interactions are present in controls LCLs, but neither in 16p11.2 deletion nor duplication LCLs (see below). Furthermore, five (*C2orf74, COMMD1* (MIM#607238), *FAM161A* (MIM#613596), *PEX13* (MIM#601789), *PUS10* (MIM#612787)) and 11 (*AHSA2, BCL11A* (MIM#606557), *PAPOLG, REL* (MIM#164910), *USP34, XPO1*) of the 13 genes mapping within the 2p15-16.1 syndrome minimal overlapping interval[111,112] show perturbed expression levels in 16p11.2 CNV patients with cutoffs of 5 and 15% FDR, respectively. Thus, the *cis*- and *trans*-chromatin contacts we uncovered bridge genomic regions, whose rearrangements are associated with ASD and mirror phenotypes on BMI and HC.

DISCUSSION

The 16p11.2 600 kb BP4-BP5 rearrangements allow investigation of molecular mechanisms underlying the co-morbidity triad of neurodevelopmental disorders, energy imbalance and HC alterations, all associated with changes in gene dosage. To identify pathways that are perturbed when the dosage of this region is modified we cataloged the chromosomal contacts of genes mapping within this genomic interval. Using chromosome conformation capture we uncovered a network of chromatin loops with genes previously associated with ASD and HC demonstrating the pertinence of this approach. We show, for example, that the 16p11.2 phenotype drivers *MVP* and *MAPK3* promoters have long-range chromatin interactions with *PTEN* and *CHD1L*, respectively. The MVP protein regulates the intracellular localization of PTEN,[118] a dual-specificity phosphatase that antagonizes PI3K/AKT and Ras/MAPK signaling pathways. Both *PTEN* germline mutations in humans and targeted inactivation in mice are associated with macrocephaly/ASD syndrome (MIM#605309).[103–106,119] Congruently, germline mutations in the Ras/MAPK pathway cause a group of syndromes frequently regrouped under the term RASopathies, recently shown to affect social interactions.[120,121] We corroboratingly revealed that expression of *PTEN* pathway members is sensitive to gene dosage at the 16p11.2 locus.[12] *CHD1L* was suggested to be a major driver of the phenotypes associated with 1q21.1 rearrangements (OMIM#612474; #612475).[107,108] Analogous to 16p11.2, deletions and duplications of this interval are linked to micro- and macrocephaly, respectively.[109]

These results suggest that chromatin interactions, even when tested in peripheral tissues (such as LCLs) not considered to play a central role in the resulting neurodevelopmental phenotype, could reveal genes or pathways, which are co-regulated and associated with similar phenotypes. Several studies have shown that the contacted domains can be highly stable across species and cell lines (even when the contacted genes are not expressed),[3,101] supporting the notion that patient-derived samples can provide direct insight into regulatory abnormalities and that LCLs still contain valuable information for the study of the patients phenotype.[12] Consistent with this hypothesis, we established that dosage perturbation of two chromatin-contacted loci, the *cis*-contacted 16p11.2 220 kb BP2-BP3 and the *trans*-interacting 2p15 intervals, are associated with mirror phenotypes on BMI and HC. Similar regulatory cores engaged in multiple physical interactions were recently described, for example, the 8q24 oncogenetic locus.[122]

The physiological relevance of the underlying chromatin architecture is further exemplified by the extensive overlap between differentially expressed genes and the loci that show modified loops upon dosage changes of the 16p11.2 600 kb BP4-BP5 interval. Together with the observed enrichments in various pathology-relevant gene ontology terms and pathways, this sheds light on a possible 'chromatin hub' role of the 16p11.2 locus in the observed phenotypes. The 12 genes that are concurrently SFARI-ASD, differentially expressed and modified in 4C contacts include (i) *VPS13B*, whose mutations cause Cohen syndrome (OMIM#216550), an autosomal recessive disorder characterized by intellectual disability, microcephaly, retinal dystrophy and truncal obesity;[123] (ii) *TCF4*, whose haploinsufficiency is associated with Pitt-Hopkins syndrome (OMIM#610954) typified among other traits by intellectual disability, recurrent seizures and microcephaly (Of note, the expression level of this transcription factor of the Wnt/β-catenin signaling cascade was shown to be altered in the cortex of mice models engineered to carry one or three copies of the 16p11.2 orthologous region[26]); (iii) the *TSC2* (MIM#191092) tuberous sclerosis (OMIM#613254) gene, which encodes an inhibitor of mTORC1 signaling that limits cell growth and was linked to ciliary dysfunction;[124] and (iv) *EP300*, which is mutated in a form of Rubinstein-Taybi syndrome 2 (OMIM#613684) associated with a more severe microcephaly.[114,125]

Whereas the functions and natures of the detected contacts remain to be elucidated, recent reports show that transcription of co-regulated genes occurs in the context of spatial proximity, which is lost in knockout studies.[2,102] While we cannot exclude that the observed proximal positions of genes is brought about

by proteins that do not directly interact with any of them[126] or occur through their use of identical transcription factory for example,[127,128] we *de facto* witness that expression of multiple genes of converging pathways is modified and that the chromatin-contacted genes are encoding proteins of overlapping interactomes (Figure 4c and Supplementary Figures S10–S11). For example, *MVP* loops with genes implicated in maintenance of cell polarity ($P = 2.75e^{-03}$) (Supplementary Table S33).

Chromatin spatial organization is conserved, to some extent at least, through evolution.[129] The distal and proximal 16p11.2 regions are physically interacting in both human and mouse cells.[73,101] This chromatin crosstalk is preserved despite modifications in the human lineage of the orientation of both the BP2-BP3 and the BP4-BP5 regions, as well as doubling of the size of the intervening region. While we cannot rule out that similar studies on the clinically relevant tissues might uncover additional important partners, these studies demonstrate that maintenance of chromatin crosstalk across tissues (from fibroblasts to cortical neurons) and in different lineages lends credence to the use of LCLs and animal models as proxies to study chromatin properties of the human central nervous system, the more likely tissue determining the phenotypes associated with 16p11.2 600 kb BP4-BP5 and BP2-BP3 CNVs.

The identified *cis*- and *trans*-chromatin contacts bridge loci whose rearrangements result in mirror phenotypes on BMI and HC, as well as involve known ASD candidate genes. While investigations of the 3D genomic structures of additional regions are warranted, the results present here support the idea that the elucidation of chromatin contacts can be proposed as a new and effective tool to unravel genes participating in similar pathways or disease mechanisms, and identify loci associated with overlapping phenotypic manifestations. Our study also suggests that modifications of chromatin interplays play a crucial role in the observed phenotypes.

16P11.2 CONSORTIUM MEMBERS

Maria Nicla Loviglio, Katrin Männik, Ilse van der Werf, Giuliana Giannuzzi, Marianna Zazhytska, Nele Gheldof, Eugenia Migliavacca, Ali A Alfaiz, Inês Roberts-Caldeira, Loyse Hippolyte, Anne M Maillard, Alessandra Ferrarini, Florence Niel Butschi, Bernard Conrad, Marie-Claude Addor, Marco Belfiore, Katharina Roetzer, Anke Van Dijck, Bettina Blaumeiser, Frank Kooy, Filip Roelens, Annelies Dheedene, Barbara Delle Chiaie, Björn Menten, Ann Oostra, Jean-Hubert Caberg, Melissa Carter, Barbara Kellam, Dimitri J Stavropoulos, Christian Marshall, Stephen W Scherer, Rosanna Weksberg, Cheryl Cytrynbaum, Anne Bassett, Chelsea Lowther, Jane Gillis, Sara MacKay, Iben Bache, Lilian B Ousager, Maja Patricia Smerdel, Jesper Graakjaer, Susanne Kjaergaard, Andres Metspalu, Michele Mathieu, Dominique Bonneau, Agnes Guichet, Philippe Parent, Claude Férec, Marion Gerard, Ghislaine Plessis, James Lespinasse, Alice Masurel, Nathalie Marle, Laurence Faivre, Patrick Callier, Valerie Layet, Nathalie Le Meur, Céline Le Goff, Bénédicte Duban-Bedu, Sylvie Sukno, Odile Boute, Joris Andrieux, Patricia Blanchet, David Geneviève, Jacques Puechberty, Anouck Schneider, Bruno Leheup, Philippe Jonveaux, Sandra Mercier, Albert David, Cédric Le Caignec, Loic de Pontual, Eva Pipiras, Aurelia Jacquette, Boris Keren, Brigitte Gilbert-Dussardier, Frederic Bilan, Alice Goldenberg, Pascal Chambon, Annick Toutain, Marianne Till, Damien Sanlaville, Barbara Leube, Brigitte Royer-Pokora, Hans Jörgen Grabe, Carsten Oliver Schmidt, Claudia Schurmann, Georg Homuth, Gudmar Thorleifsson, Unnur Thorsteinsdottir, Laura Bernardini, Antonio Novelli, Lucia Micale, Giuseppe Merla, Marcella Zollino, Francesca Mari, Caterina Lo Rizzo, Alessandra Renieri, Margherita Silengo, Anneke T Vulto-van Silfhout, Meyke Schouten, Rolph Pfundt, Nicole de Leeuw, Fleur Vansenne, Saskia M Maas, Daniela QCM Barge-Schaapveld, Alida C Knegt, Barbro Stadheim, Olaug Rodningen, Gunnar Houge, Sue Price, Lara Hawkes, Carolyn Campbell, Usha Kini, Julie Vogt, Robin Walters, Alexandra Blakemore, James F Gusella, Yiping Shen, Daryl Scott, Carlos A Bacino, Karen Tsuchiya, Roger Ladda, Susan Sell, Alexander Asamoah, Aline I Hamati, Jill A Rosenfeld, Lisa G Shaffer, Elyse Mitchell, Jennelle C Hodge, Jacques S Beckmann, Sébastien Jacquemont, Alexandre Reymond. The affiliations and email addresses of the members of the 16p11.2 Consortium are listed in Supplementary Table S36.

2P15 CONSORTIUM MEMBERS

Alexandre Reymond, Lisa J Ewans, David Mowat, Jan Walker, David J Amor, Hilde Van Esch, Patricia Leroy, Jean-Hubert Caberg, John-Steven Bamforth, Deepti Babu, Marianne Till, Damien Sanlaville, David Geneviève, Jacques Puechberty, Bertrand Isidor, Nataliya DiDonato, Karl Hackmann, Marzia Passeggeri, Arie van Haeringen, Jill A Rosenfeld, Lisa G Shaffer, Rosemarie Smith, Sara Ellingwood, Darren M Farber, Vinay Puri, Neda Zadeh, David D Weaver, Mandy Miller, Timothy Wilks, Carolina J Jorgez, DeeDee Lafayette, Sébastien Jacquemont. The affiliations and email addresses of the members of the 2p15 Consortium are listed in Supplementary Table S37.

ACCESSION NUMBERS

GEO Seriegs accession number: GSE57802.

THE WEB RESOURCES SECTION

4C-seq pipeline: http://htsstation.epfl.ch/BioScript: http://gdv.epfl.ch/bsGeneProf: http://www.geneprof.orgGTEx: http://www.gtexportal.org/home/gFeatBrowser: http://www.gfeatbrowser.com. Online Mendelian Inheritance in Man: http://www.omim.orgSFARI: https://sfari.org/STRING: http://string-db.org/.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank the members of the Lausanne Genomic Technologies Facility and Jacqueline Chrast for technical help and Catia Attanasio for comments on the manuscript. NG is a grantee of the Marie Heim Vögtlin and the Pro-Women programs of the SNSF and the Faculty of Biology and Medicine, University of Lausanne, respectively. KM was awarded a fellowship from the Swiss Scientific Exchange NMS Program. IvdW was recipient of scholarships from the Nora Baart Foundation and Erasmus European Union Lifelong Learning Programme. AAA was awarded a scholarship from the Saudi Arabian National Guard Health Affairs. SMW was supported by SNSF (Swiss National Science Foundation)(P2ELP3_155365) and EMBO (ALTF 755-2014) fellowships. SJ is recipient of a SNSF Bursary Professor fellowship. This work is supported by grants from the Simons Foundation (SFARI274424 to AR), the Swiss National Science Foundation 31003A_160203 (AR), SNSF Sinergia grants CRSII33-133044 (AR) and CRSI33-130326 (ETD, BD, AR), the Leenaards Foundation Prize (SJ, AR) and the SystemsX.ch initiative grant 3826 (ETD and AR). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

MNL, NG and GG performed the 4C and FISH experiments. ML, MNL, EM, AAA and JR conducted the statistical analyses. KM, MNL and IvdW established cell lines. IvdW, KM and MZ performed the qPCR experiments. SMW, ETD and BD run the molecular associations. IRC, KM, IvdW, MP, LH, AMM, RFK, DS, JAR, LGS, JA, CM, SWS, YS, JFG, UT, GT, JSB, SJ and all other members of the 16p11.2 and 2p15 consortia phenotyped and genotyped individuals. MNL and AR wrote the manuscript with contributions from GG, JSB and SJ. AR designed the study and obtained the necessary funding. All authors commented on and approved the manuscript.

REFERENCES

1 de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 2013; **502**: 499–506.
2 Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. Chromosomal contact permits transcription between coregulated genes. *Cell* 2013; **155**: 606–620.
3 Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015; **161**: 1012–1025.
4 Lupianez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet* 2016; **32**: 225–237.
5 Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* 2001; **413**: 514–519.
6 Antonacci F, Kidd JM, Marques-Bonet T, Teague B, Ventura M, Girirajan S *et al.* A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nat Genet* 2010; **42**: 745–750.
7 Giannuzzi G, Migliavacca E, Reymond A. Novel H3K4me3 marks are enriched at human- and chimpanzee-specific cytogenetic structures. *Genome Res* 2014; **24**: 1455–1468.
8 Ballif BC, Hornor SA, Jenkins E, Madan-Khetarpal S, Surti U, Jackson KE *et al.* Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat Genet* 2007; **39**: 1071–1073.
9 Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008; **358**: 667–675.
10 Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010; **42**: 203–209.
11 Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nat Rev Genet* 2009; **10**: 551–564.
12 Migliavacca E, Golzio C, Mannik K, Blumenthal I, Oh EC, Harewood L *et al.* A potential contributory role for ciliary dysfunction in the 16p11.2 600 kb BP4-BP5 pathology. *Am J Hum Genet* 2015; **96**: 784–796.
13 Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009; **459**: 569–573.
14 Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008; **82**: 477–488.
15 Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L *et al.* A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J Med Genet* 2012; **49**: 660–668.
16 Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* 2010; **463**: 666–670.
17 Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 2010; **463**: 671–675.
18 Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* 2011; **478**: 97–102.
19 Shinawi M, Liu P, Kang SH, Shen J, Belmont JW, Scott DA *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J Med Genet* 2010; **47**: 332–341.
20 Reinthaler EM, Lal D, Lebon S, Hildebrand MS, Dahl HH, Regan BM *et al.* 16p11.2 600 kb duplications confer risk for typical and atypical Rolandic epilepsy. *Hum Mol Genet* 2014; **23**: 6069–6080.
21 Dimassi S, Labalme A, Lesca G, Rudolf G, Bruneau N, Hirsch E *et al.* A subset of genomic alterations detected in rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. *Epilepsia* 2014; **55**: 370–378.
22 McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 2009; **41**: 1223–1227.
23 D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L *et al.* Defining the effect of the 16p11.2 duplication on cognition, behavior, and medical comorbidities. *JAMA Psychiatry* 2016; **73**: 20–30.
24 Mannik K, Magi R, Mace A, Cole B, Guyatt AL, Shihab HA *et al.* Copy number variations and cognitive phenotypes in unselected populations. *JAMA* 2015; **313**: 2044–2054.
25 Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S *et al.* The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Mol Psychiatry* 2015; **20**: 140–147.
26 Blumenthal I, Ragavendran A, Erdin S, Klei L, Sugathan A, Guide JR *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.
27 Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S *et al.* KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* 2012; **485**: 363–367.
28 Walters RG, Coin LJ, Ruokonen A, de Smith AJ, El-Sayed Moustafa JS, Jacquemont S *et al.* Rare genomic structural variants in complex disease: lessons from the replication of associations with obesity. *PLoS One* 2013; **8**: e58048.
29 Barge-Schaapveld DQ, Maas SM, Polstra A, Knegt LC, Hennekam RC. The atypical 16p11.2 deletion: a not so atypical microdeletion syndrome? *Am J Med Genet A* 2011; **155A**: 1066–1072.
30 Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S *et al.* Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genet Med* 2010; **12**: 641–647.
31 Sampson MG, Coughlin CR II, Kaplan P, Conlin LK, Meyers KE, Zackai EH *et al.* Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. *Am J Med Genet A* 2010; **152A**: 2618–2622.
32 Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med* 2012; **367**: 1321–1331.
33 Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C *et al.* A copy number variation morbidity map of developmental delay. *Nat Genet* 2011; **43**: 838–846.
34 Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, Anderlid BM *et al.* Extending the phenotype of recurrent rearrangements of 16p11.2: deletions in mentally retarded patients without autism and in normal individuals. *Eur J Med Genet* 2009; **52**: 77–87.
35 Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA Psychiatry* 2013; **70**: 253–260.
36 Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 2009; **41**: 35–46.
37 Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, Toncheva D *et al.* Genetic structure of Europeans: a view from the North-East. *PLoS One* 2009; **4**: e5472.
38 John U, Greiner B, Hensel E, Ludemann J, Piek M, Sauer S *et al.* Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz Praventivmed* 2001; **46**: 186–194.
39 Firmann M, Mayor V, Vidal PM, Bochud M, Pecoud A, Hayoz D *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord* 2008; **8**: 6.
40 Chabchoub E, Willekens D, Vermeesch JR, Fryns JP. Holoprosencephaly and ZIC2 microdeletions: novel clinical and epidemiological specificities delineated. *Clin Genet* 2012; **81**: 584–589.
41 de Leeuw N, Pfundt R, Koolen DA, Neefs I, Scheltinga I, Mieloo H *et al.* A newly recognised microdeletion syndrome involving 2p15p16.1: narrowing down the critical region by adding another patient detected by genome wide tiling path array comparative genomic hybridisation analysis. *J Med Genet* 2008; **45**: 122–124.
42 Felix TM, Petrin AL, Sanseverino MT, Murray JC. Further characterization of microdeletion syndrome involving 2p15-p16.1. *Am J Med Genet A* 2010; **152A**: 2604–2608.
43 Florisson JM, Mathijssen IM, Dumee B, Hoogeboom JA, Poddighe PJ, Oostra BA *et al.* Complex craniosynostosis is associated with the 2p15p16.1 microdeletion syndrome. *Am J Med Genet A* 2013; **161A**: 244–253.
44 Hancarova M, Vejvalkova S, Trkova M, Drabova J, Dleskova A, Vlckova M *et al.* Identification of a patient with intellectual disability and de novo 3.7 Mb deletion supports the existence of a novel microdeletion syndrome in 2p14-p15. *Gene* 2013; **516**: 158–161.
45 Hucthagowder V, Liu TC, Paciorkowski AR, Thio LL, Keller MS, Anderson CD *et al.* Chromosome 2p15p16.1 microdeletion syndrome: 2.5 Mb deletion in a patient with renal anomalies, intractable seizures and a choledochal cyst. *Eur J Med Genet* 2012; **55**: 485–489.
46 Liang JS, Shimojima K, Ohno K, Sugiura C, Une Y, Ohno K *et al.* A newly recognised microdeletion syndrome of 2p15-16.1 manifesting moderate developmental delay, autistic behaviour, short stature, microcephaly, and dysmorphic features: a new patient with 3.2 Mb deletion. *J Med Genet* 2009; **46**: 645–647.
47 Piccione M, Piro E, Serraino F, Cavani S, Ciccone R, Malacarne M *et al.* Interstitial deletion of chromosome 2p15-16.1: report of two patients and critical review of current genotype-phenotype correlation. *Eur J Med Genet* 2012; **55**: 238–244.
48 Prontera P, Bernardini L, Stangoni G, Capalbo A, Rogaia D, Romani R *et al.* Deletion 2p15-16.1 syndrome: case report and review. *Am J Med Genet A* 2011; **155A**: 2473–2478.

49 Rajcan-Separovic E, Harvard C, Liu X, McGillivray B, Hall JG, Qiao Y *et al.* Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. *J Med Genet* 2007; **44**: 269–276.
50 Jorgez CJ, Rosenfeld JA, Wilken NR, Vangapandu HV, Sahin A, Pham D *et al.* Genitourinary defects associated with genomic deletions in 2p15 encompassing OTX1. *PLoS One* 2014; **9**: e107028.
51 Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 2006; **38**: 1348–1354.
52 Simonis M, Kooren J, de Laat W. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* 2007; **4**: 895–901.
53 Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. The dynamic architecture of Hox gene clusters. *Science* 2011; **334**: 222–225.
54 Gheldof N, Leleu M, Noordermeer D, Rougemont J, Reymond A. Detecting long-range chromatin interactions using the chromosome conformation capture sequencing (4C-seq) method. *Methods Mol Biol* 2012; **786**: 211–225.
55 Blaker-Lee A, Gupta S, McCammon JM, De Rienzo G, Sive H. Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. *Dis Models Mech* 2012; **5**: 834–851.
56 Kishi H, Mukai T, Hirono A, Fujii H, Miwa S, Hori K. Human aldolase A deficiency associated with a hemolytic anemia: thermolabile aldolase due to a single base mutation. *Proc Natl Acad Sci USA* 1987; **84**: 8623–8627.
57 Watabe-Rudolph M, Schlautmann N, Papaioannou VE, Gossler A. The mouse rib-vertebrae mutation is a hypomorphic Tbx6 allele. *Mech Dev* 2002; **119**: 251–256.
58 Fei Q, Wu Z, Wang H, Zhou X, Wang N, Ding Y *et al.* The association analysis of TBX6 polymorphism with susceptibility to congenital scoliosis in a Chinese Han population. *Spine* 2010; **35**: 983–988.
59 Sparrow DB, McInerney-Leo A, Gucev ZS, Gardiner B, Marshall M, Leo PJ *et al.* Autosomal dominant spondylocostal dysostosis is caused by mutation in TBX6. *Hum Mol Genet* 2013; **22**: 1625–1631.
60 Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M *et al.* TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med* 2015; **372**: 341–350.
61 Duan C, Li M, Rui L. SH2-B promotes insulin receptor substrate 1 (IRS1)- and IRS2-mediated activation of the phosphatidylinositol 3-kinase pathway in response to leptin. *J Biol Chem* 2004; **279**: 43684–43691.
62 Ren D, Zhou Y, Morris D, Li M, Li Z, Rui L. Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *J Clin Invest* 2007; **117**: 397–406.
63 Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41**: 25–34.
64 Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009; **41**: 18–24.
65 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937–948.
66 Jamshidi Y, Snieder H, Ge D, Spector TD, O'Dell SD. The SH2B gene is associated with serum leptin and body fat in normal female twins. *Obesity (Silver Spring)* 2007; **15**: 5–9.
67 Doche ME, Bochukova EG, Su HW, Pearce LR, Keogh JM, Henning E *et al.* Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest* 2012; **122**: 4732–4736.
68 Fuller DM, Zhu M, Ou-Yang CW, Sullivan SA, Zhang W. A tale of two TRAPs: LAT and LAB in the regulation of lymphocyte development, activation, and autoimmunity. *Immunol Res* 2011; **49**: 97–108.
69 Shim EK, Jung SH, Lee JR. Role of two adaptor molecules SLP-76 and LAT in the PI3K signaling pathway in activated T cells. *J Immunol* 2011; **186**: 2926–2935.
70 Fuller DM, Zhu M, Koonpaew S, Nelson MI, Zhang W. The importance of the Erk pathway in the development of linker for activation of T cells-mediated autoimmunity. *J Immunol* 2012; **189**: 4005–4013.
71 David FP, Delafontaine J, Carat S, Ross FJ, Lefebvre G, Jarosz Y *et al.* HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PLoS One* 2014; **9**: e85879.
72 Tolhuis B, Blom M, Kerkhoven RM, Pagie L, Teunissen H, Nieuwland M *et al.* Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet* 2011; **7**: e1001343.
73 Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; **326**: 289–293.
74 de Wit E, Braunschweig U, Greil F, Bussemaker HJ, van Steensel B. Global chromatin domain organization of the Drosophila genome. *PLoS Genet* 2008; **4**: e1000045.
75 Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014; **159**: 1665–1680.
76 Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2013; **41**: D808–D815.
77 Alexa A, Rahnenfuhrer J. topGO: Enrichment analysis for Gene Ontology. BioConductor, 2010.
78 da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**: 44–57.
79 Reimand J, Arak T, Vilo J. g:Profiler—a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res* 2011; **39**: W307–W315.
80 Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013; **14**: 128.
81 Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat* 2000; **15**: 57–61.
82 Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L *et al.* Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 2011; **43**: 860–863.
83 Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y *et al.* De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat Genet* 2012; **44**: 1365–1369.
84 Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S *et al.* Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 2011; **43**: 864–868.
85 Vimaleswaran KS, Tachmazidou I, Zhao JH, Hirschhorn JN, Dudbridge F, Loos RJ. Candidate genes for obesity-susceptibility show enriched association within a large genome-wide association study for BMI. *Hum Mol Genet* 2012; **21**: 4537–4542.
86 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
87 Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
88 van Dam TJ, Wheway G, Slaats GG, Group SS, Huynen MA, Giles RH. The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* 2013; **2**: 7.
89 Stehbens S, Wittmann T. Targeting and transport: how microtubules control focal adhesion dynamics. *J Cell Biol* 2012; **198**: 481–489.
90 Bukoreshtliev NV, Haase K, Pelling AE. Mechanical cues in cellular signalling and communication. *Cell Tissue Res* 2013; **352**: 77–94.
91 Seeger-Nukpezah T, Golemis EA. The extracellular matrix and ciliary signaling. *Curr Opin Cell Biol* 2012; **24**: 652–661.
92 He G, Liu X, Qin W, Chen Q, Wang X, Yang Y *et al.* MPZL1/PZR, a novel candidate predisposing schizophrenia in Han Chinese. *Mol Psychiatry* 2006; **11**: 748–751.
93 Eminaga S, Bennett AM. Noonan syndrome-associated SHP-2/Ptpn11 mutants enhance SIRPalpha and PZR tyrosyl phosphorylation and promote adhesion-mediated ERK activation. *J Biol Chem* 2008; **283**: 15328–15338.
94 Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D *et al.* High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* 1990; **247**: 64–69.
95 Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM *et al.* Population variation and genetic control of modular chromatin architecture in humans. *Cell* 2015; **162**: 1039–1050.
96 Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM *et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.
97 Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A *et al.* Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* 2013; **342**: 744–747.
98 Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013; **501**: 506–511.
99 CDC. Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill Summ* 2014; **63**: 1–21.
100 Gheldof N, Witwicki RM, Migliavacca E, Leleu M, Didelot G, Harewood L *et al.* Structural variation-associated expression changes are paralleled by chromatin architecture modifications. *PLoS One* 2013; **8**: e79973.
101 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; **485**: 376–380.

102 Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY *et al.* Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* 2013; **504**: 306–310.
103 Butler MG, Dasouki MJ, Zhou XP, Talebizadeh Z, Brown M, Takahashi TN *et al.* Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J Med Genet* 2005; **42**: 318–321.
104 Herman GE, Butter E, Enrile B, Pastore M, Prior TW, Sommer A. Increasing knowledge of PTEN germline mutations: two additional patients with autism and macrocephaly. *Am J Med Genet A* 2007; **143**: 589–593.
105 O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012; **338**: 1619–1622.
106 Pal A, Barber TM, Van de Bunt M, Rudge SA, Zhang Q, Lachlan KL *et al.* PTEN mutations as a cause of constitutive insulin sensitivity and obesity. *N Engl J Med* 2012; **367**: 1002–1011.
107 Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD *et al.* Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am J Hum Genet* 2013; **92**: 221–237.
108 Harvard C, Strong E, Mercier E, Colnaghi R, Alcantara D, Chow E *et al.* Understanding the impact of 1q21.1 copy number variant. *Orphanet J Rare Dis* 2011; **6**: 54.
109 Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* 2008; **40**: 1466–1471.
110 Griswold AJ, Ma D, Cukier HN, Nations LD, Schmidt MA, Chung RH *et al.* Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. *Hum Mol Genet* 2012; **21**: 3513–3523.
111 Liu X, Malenfant P, Reesor C, Lee A, Hudson ML, Harvard C *et al.* 2p15-p16.1 microdeletion syndrome: molecular characterization and association of the OTX1 and XPO1 genes with autism spectrum disorders. *Eur J Hum Genet* 2011; **19**: 1264–1270.
112 Fannemel M, Baroy T, Holmgren A, Rodningen OK, Haugsand TM, Hansen B *et al.* Haploinsufficiency of XPO1 and USP34 by a de novo 230 kb deletion in 2p15, in a patient with mild intellectual disability and cranio-facial dysmorphisms. *Eur J Med Genet* 2014; **57**: 513–519.
113 Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S *et al.* Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am J Hum Genet* 2012; **90**: 133–141.
114 Seltzer LE, Paciorkowski AR. Genetic disorders associated with postnatal microcephaly. *Am J Med Genet C Semin Med Genet* 2014; **166**: 140–155.
115 Bjursell MK, Blom HJ, Cayuela JA, Engvall ML, Lesko N, Balasubramaniam S *et al.* Adenosine kinase deficiency disrupts the methionine cycle and causes hypermethioninemia, encephalopathy, and abnormal liver function. *Am J Hum Genet* 2011; **89**: 507–515.
116 Doornbos M, Sikkema-Raddatz B, Ruijvenkamp CA, Dijkhuizen T, Bijlsma EK, Gijsbers AC *et al.* Nine patients with a microdeletion 15q11.2 between breakpoints 1 and 2 of the Prader-Willi critical region, possibly associated with behavioural disturbances. *Eur J Med Genet* 2009; **52**: 108–115.
117 Kolehmainen J, Wilkinson R, Lehesjoki AE, Chandler K, Kivitie-Kallio S, Clayton-Smith J *et al.* Delineation of Cohen syndrome following a large-scale genotype-phenotype screen. *Am J Hum Genet* 2004; **75**: 122–127.
118 Chung JH, Eng C. Nuclear-cytoplasmic partitioning of phosphatase and tensin homologue deleted on chromosome 10 (PTEN) differentially regulates the cell cycle and apoptosis. *Cancer Res* 2005; **65**: 8096–8100.
119 Kwon CH, Luikart BW, Powell CM, Zhou J, Matheny SA, Zhang W *et al.* Pten regulates neuronal arborization and social interaction in mice. *Neuron* 2006; **50**: 377–388.
120 Adviento B, Corbin IL, Widjaja F, Desachy G, Enrique N, Rosser T *et al.* Autism traits in the RASopathies. *J Med Genet* 2014; **51**: 10–20.
121 Alfieri P, Piccini G, Caciolo C, Perrino F, Gambardella ML, Mallardi M *et al.* Behavioral profile in RASopathies. *Am J Med Genet A* 2014; **164A**: 934–942.
122 Du M, Yuan T, Schilter KF, Dittmar RL, Mackinnon A, Huang X *et al.* Prostate cancer risk locus at 8q24 as a regulatory hub by physical interactions with multiple genomic loci across the genome. *Hum Mol Genet* 2014; **24**: 154–166.
123 Balikova I, Lehesjoki AE, de Ravel TJ, Thienpont B, Chandler KE, Clayton-Smith J *et al.* Deletions in the VPS13B (COH1) gene as a cause of Cohen syndrome. *Hum Mutat* 2009; **30**: E845–E854.
124 Mahjoub MR, Stearns T. Supernumerary centrosomes nucleate extra cilia and compromise primary cilium signaling. *Curr Biol* 2012; **22**: 1628–1634.
125 Bartsch O, Labonte J, Albrecht B, Wieczorek D, Lechno S, Zechner U *et al.* Two patients with EP300 mutations and facial dysmorphism different from the classic Rubinstein-Taybi syndrome. *Am J Med Genet A* 2010; **152A**: 181–184.
126 Gavrilov AA, Gushchanskaya ES, Strelkova O, Zhironkina O, Kireev II, Iarovaia OV *et al.* Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res* 2013; **41**: 3563–3575.
127 Edelman LB, Fraser P. Transcription factories: genetic programming in three dimensions. *Curr Opin Genet Dev* 2012; **22**: 110–114.
128 Park SK, Xiang Y, Feng X, Garrard WT. Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. *Genes Dev* 2014; **28**: 1159–1164.
129 Mitchell AC, Bharadwaj R, Whittle C, Krueger W, Mirnics K, Hurd Y *et al.* The genome in three dimensions: a new frontier in human brain research. *Biol Psychiatry* 2014; **75**: 961–969.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)