Local and global chromatin interactions are altered by large genomic deletions associated with human brain development

Ying Zhang[1*], Xianglong Zhang[2,3*], Xiaowei Zhu[2,3*], Carolin Purmann[2,3], Michael S Haney[3], Thomas Ward[2,3], Jie Yao[4], Sherman M Weissman[5#], Alexander E Urban[2,3#]

1 Department of Genetics and Genomic Sciences, Icahn School of medicine at Mount Sinai, 2 Department of Psychiatry and Behavioral Sciences, 3 Department of Genetics, Stanford University School of Medicine, 4 Department of Cell Biology, Yale University School of Medicine, 5 Department of Genetics, Yale University

Email addresses:

Ying Zhang, zzhang006@gmail.com;

Xianglong Zhang, xzhang8@stanford.edu;

Xiaowei Zhu, xwzhu@stanford.edu;

Carolin Purmann, purmann@stanford.edu;

Michael S Haney, haney@stanford.edu;

Thomas Ward, wardtr@stanford.edu;

Jie Yao, yao.j@yale.edu;

Sherman M Weissman, sherman.weissman@yale.edu;

Alexander E Urban, aeurban@stanford.edu.

*These authors contributed equally to this work.

#To whom correspondence should be addressed.

Abstract

Background: Large copy number variants (CNVs) in the human genome are strongly associated with common neurodevelopmental, neuropsychiatric disorders such as schizophrenia and autism. Using Hi-C analysis of long-range chromosome interactions and ChIP-Seq analysis of regulatory histone marks we studied the epigenomic effects of the prominent large deletion CNV on chromosome 22q11.2 and also replicated a subset of the findings for the large deletion CNV on chromosome 1q21.1.

Results: We found that, in addition to local and global gene expression changes, there are pronounced and multilayered effects on chromatin states, chromosome folding and topological domains of the chromatin, that emanate from the large CNV locus. Regulatory histone marks are altered in the deletion proximal regions, and in opposing directions for activating and repressing marks. There are also significant changes of histone marks elsewhere along chromosome 22q and genome wide. Chromosome interaction patterns are weakened within the deletion boundaries and strengthened between the deletion proximal regions. We detected a change in the manner in which chromosome 22q folds onto itself, namely by increasing the long-range contacts between the telomeric end and the deletion proximal region. Further, the large CNV affects the topological domain that is spanning its genomic region. Finally, there is a widespread and complex effect on chromosome interactions genome-wide, i.e. involving all other autosomes, with some of the effect directly tied to the deletion region on 22q11.2.

Conclusions: These findings suggest novel principles of how such large genomic deletions can alter nuclear organization and affect genomic molecular activity.

Keywords: Germline CNVs, Chromatin States, Chromosome Interactions, Topological Domains

Introduction

Two of the most exciting discoveries in human genetics of the past decade are that small- to medium-sized Copy Number Variants (CNVs) are very common in the human genome, and that there is a group of large CNVs that are strongly associated with brain development and neuropsychiatric disorders such as schizophrenia and the autism spectrum disorders (ASDs) [1,2]. These large CNVs are understood to be providing enticing points of entry to the analysis of the strong but complex genetic and molecular (and possibly even cellular) basis of these common disorders.

We now know that small to medium-sized CNVs, i.e. deletions or duplications of genomic DNA sequence ranging in size from hundreds to tens of thousands of basepairs, are present in any human genome with their numbers in the thousands [3-7]. Not very much is known as of yet about such smaller CNVs' effects on the normal phenotype, but some examples already exist that show that such effects could be considerable, such as the copy number variation of the amylase gene associated with a given human population's ability to digest high-starch food [8].

Large CNVs, typically sized from hundreds of thousands to millions of basepairs of genomic DNA sequence, were previously known to be in strong association with often-severe but rare congenital malformations, or found in cancer genomes. It was a striking discovery when a series of studies [1,2] showed that there is a group of more than ten large CNVs that are strongly associated with aberrant brain development and a resulting neuropsychiatric phenotype such as schizophrenia, ASD or Williams Syndrome. These large

68 neuropsychiatric CNVs can encompass many genes and their effects across the various
69 molecular levels of gene activity and regulation and the clinical phenotype are complex and
70 only poorly understood.

71 On the molecular level these large neuropsychiatric CNVs have been mostly studied by
72 applying the research paradigm of trying to determine which single gene from within the
73 CNV boundaries is at the root of the large CNV's effects on brain development. Many very
74 interesting insights have been gained using this approach into a considerable number of
75 genes with these large CNVs.

76 However, these insights about individual genes fall far short of explaining the full
77 effects of the large CNVs. Also there already have been a number of transcriptome wide
78 studies that at least hint at certain network effects emanating from the large CNVs. Which
79 mechanisms mediate such transcription network effects is then the question. Furthermore
80 there are an increasing number of studies that show a potentially very important role of
81 chromatin regulation in the molecular etiology of neuropsychiatric disorder [9-15].

82 Against this backdrop we reasoned that it was worthwhile testing whether large CNVs
83 with association to brain development might cause a disruption or at least alteration of one
84 or several aspects of chromatin conformation, such as the distribution of regulatory
85 chromatin marks, the long-range direct physical interactions between distant regions on
86 one chromosome or between different chromosomes or the higher-order chromatin domain
87 structures that are defined by such marks or interactions. Such effects on these important
88 layers of molecular regulation of gene activity would then constitute a novel principle by
89 which large CNVs could transmit their presence to the machinery of cellular physiology.

90 Here we show in a cohort of cell lines derived from patients with 22q11 Deletion
91 Syndrome that indeed chromatin marks, chromatin domains and long-range chromosome
92 interactions are affected in several distinct ways by a large, common and disease-associated
93 CNV on chromosome 22q11.2. We then go on to show in a smaller set of cell lines from
94 different patients that at least some of the same observations can also be made with another
95 neuropsychiatric large CNV, on chromosome 1q21.1.

96 22q11 Deletion Syndrome (22q11DS) is a disorder caused by a heterozygous deletion
97 of 3 million basepairs containing more than 60 known genes on chromosome 22q11.2. It
98 occurs in at least 1 in 4,000 live births. The common phenotypes of 22q11DS include a large
99 spectrum of congenital anomalies, for example of the facial structures and the
100 cardiovascular system – and notably there is a strong association with several
101 neurodevelopmental psychiatric disorders, in particular schizophrenia and ASDs [2,16-19]. In
102 this study we are using the large CNV on 22q11.2 as a model to determine the generalizable
103 principles along which large CNVs of this important category can lead to changes to the
104 various ways in which chromatin is ordered, using unbiased, genome-wide, sequencing-
105 based assays for discovery.

106 **Results**

107 **High-volume sequencing data generation for Hi-C chromosomal contact maps, ChIP-**
108 **Seq chromatin marks and RNA-Seq gene expression data**

109 To determine possible effects of the 22q11.2 deletion on chromosomal interactions, we
110 generated Hi-C contact maps for 11 human lymphoblastoid cell lines (5 patient cell lines
111 with 22q11.2 deletion and 6 control cell lines without), with a total of 3.1 billion Hi-C
112 contact-reads of which 680 million read-pairs were of high quality and used for the

113 downstream analyses (**Supplementary Table 1**). The existence of the 3 Mbp deletion in the
114 patient cell lines was validated by whole genome sequencing **(Supplementary Fig. 1)**.

115 As a first means of quality control we included in our data generation and analysis the
116 cell line GM06990, which was the cell line used in the original Hi-C paper [20].
117 Interchromosomal contacts of GM06990 as determined by our own Hi-C data for this line
118 (**Supplementary Fig. 2a**) show the same patterns of chromosomal interactions across the
119 nucleus as in [20]; i.e. small chromosomes generally have more interactions with each other
120 than larger chromosomes with each other, and many more than chromosomes in the
121 medium size range. Specific interaction pairs, e.g. between chromosomes 17 and 19 were
122 also replicated in our GM06990 data. Global interchromosomal contact maps resulting from
123 combining all our controls and cases, respectively, again replicated these global interaction
124 patterns (**Supplementary Fig. 2b,c**).

125 To determine the effects of 22q11.2 deletion on gene expression patterns and
126 chromatin marks we performed RNA-Seq on 14 cell lines (**Supplementary Table 1**) and
127 ChIP-seq of H3K27ac and H3K27me3 for 6 cell lines and CTCF for 5 of the same cell lines,
128 respectively (**Supplementary Table 2**).

129 **Normalization of Hi-C data**

130 Many factors, such as mappability of sequencing reads, GC content, length of the restriction
131 enzyme fragment etc., can lead to biases in Hi-C data. Several alternative computational
132 approaches have been developed for the normalization of Hi-C data with the aim to remove
133 these biases [21-30]. While these approaches use different algorithmic principles, it has been
134 demonstrated that when using them on Hi-C data generated for control genomes without
135 large CNVs, such as GM12878, the normalization metrics will be highly correlated across the

136 various normalization schemes [31]. For our Hi-C data however, we needed to be certain to
137 use a normalization method that is not thrown off from the outset by the presence of the
138 heterozygous 3 Mbp deletion in 22q11.2.

139 We tested three commonly used different normalization algorithms that were
140 developed for Hi-C data, using the rationale that the normalization methods should not
141 change the general patterns of interaction we can see in the raw data.

142 We found that not all of the available normalization methods are robust for use with
143 Hi-C data coming from genomes with large CNVs but that the hicpipe algorithm [21] is quite
144 suitable for this purpose (details of the comparison between normalization methods in the
145 **Supplementary Information** and **Supplementary Fig. 3**).

146 **Chromosomal contacts decrease within the 22q11.2 deletion boundaries and**
147 **between the deletion region and the remainder of chromosome 22q**

148 We observed that the chromosomal contacts within the 22q11.2 deletion regions in
149 22q11DS lines decreased significantly compared to control cell lines (**Fig. 1a**). Also strongly
150 reduced in the deletion cell lines were the chromosomal contacts between the 22q11.2
151 deletion region with the entire remainder of chromosome 22 (**Fig. 1a; Supplementary Fig.**
152 **4a**). This dosage effect is consistent with the copy number of the 22q11.2 deletion region in
153 the patient cell lines, as all of the 22q11DS cell lines are heterozygously deleted for this
154 region. No such dosage effect on the chromosomal contacts was observed that did not
155 involve the 22q11.2 region. Furthermore, there was no such strong dosage effect on
156 intrachromosomal contacts on any of the other autosomes (for example chromosome 19 in
157 **Fig. 1b; Supplementary Fig. 4b**).

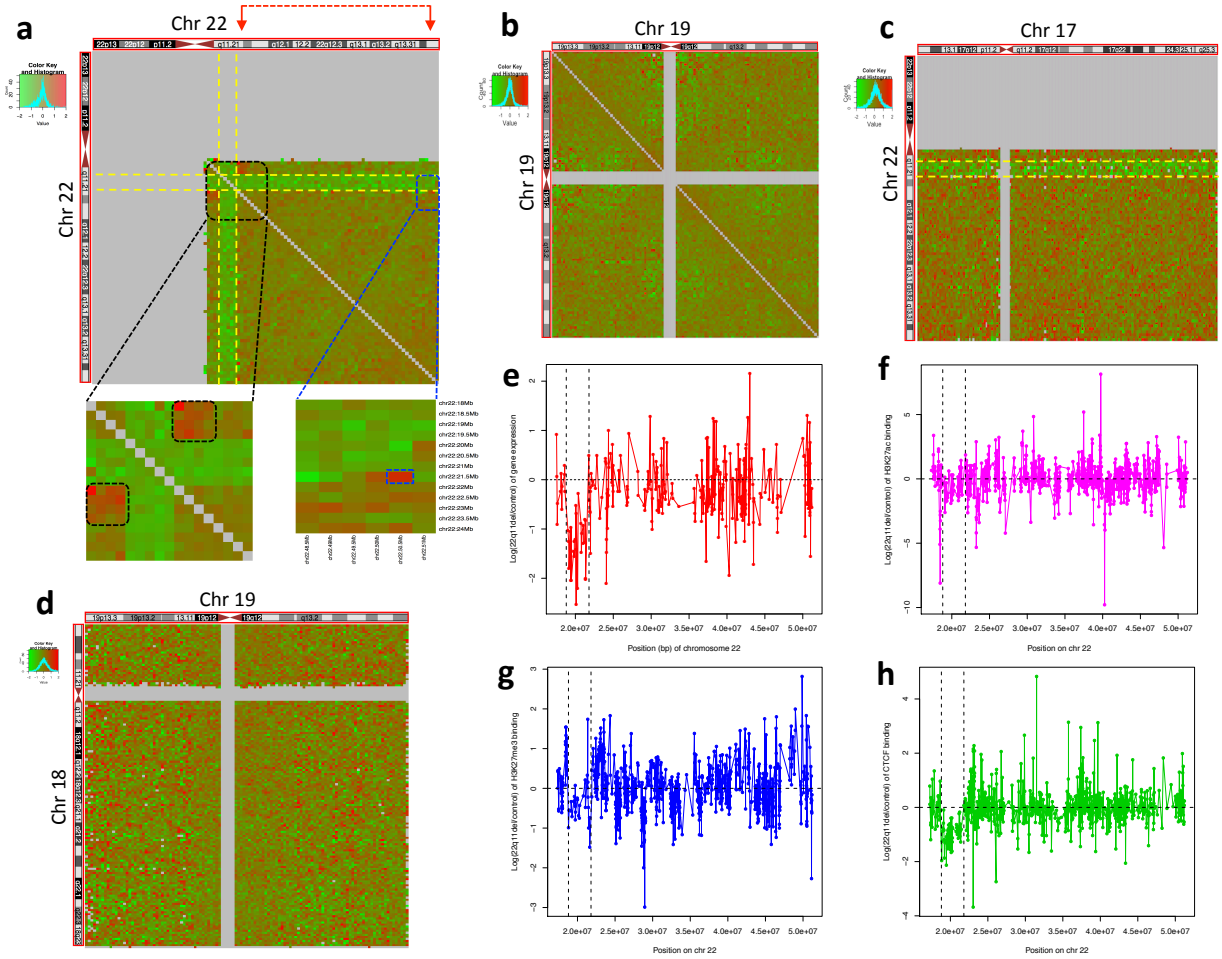


Figure 1. Effects of the 22q11.2 deletion on chromosome contacts, gene expression and epigenetic profiles. *Panels a-d: each pixel in the heat-maps represents the intra- or inter-chromosomal contact frequency in Hi-C data from 22q11.2del cell lines versus control cell lines for a 500 kbp region. Yellow dashed lines indicate the 3 Mbp deletion on chromosome 22q. (a) Fold change of cis-contacts along chromosome 22 in 22q11.2del versus control cell lines. Black boxes indicate increased contacts between the deletion-flanking regions in 22q11.2del cell lines. Blue box: the signal for increased contacts between the downstream deletion-flanking region and the telomeric end of chromosome 22q (red arrows and dashed red line indicate the corresponding chromosome folding event). (b) Lack of intra-chromosomal fold change of contacts for chromosome 19. (c) Fold change of inter-chromosomal contacts between*

169 *chromosome 22 and chromosome 17. (d) Lack of intra-chromosomal fold change of contacts*
170 *between chromosome 18 and chromosome 19. (e) Log2 transformed fold change of gene*
171 *expression for genes on chromosome 22q in RNA-Seq data from 22q11.2del versus cell lines.*
172 *Each point represents a gene. Panels f-h: Log2 transformed fold change in ChIP-Seq signals in*
173 *22q11.2del versus control cell lines. (f) H3K27ac histone modifications. (g) H3K27me3 histone*
174 *modifications. (h) CTCF binding sites. Black dashed lines indicate the 3 Mbp deletion on*
175 *chromosome 22q in e-h.*

176 Then we investigated whether this dosage effect of the 22q11.2 deletion on *cis*-contacts
177 (i.e. within the deletion boundaries and between the deletion region and elsewhere on
178 chromosome 22q) holds for *trans*-contacts as well (i.e. for contacts between the region
179 within the deletion boundaries and the rest of the genome). We found the *trans*-contacts
180 involving the 22q11.2 deletion in the patient cell lines and any other chromosome also
181 decreased compared to control cell lines (**Fig. 1c; Supplementary Fig. 4c**). No other
182 regions of chromosome 22q showed such an effect, and neither did any other pair of
183 autosomes that did not include 22q (**Fig. 1d; Supplementary Fig. 4d**).

184 **The 22q11.2 deletion has an effect on gene expression and epigenetic profiles in the**
185 **22q11.2 region**

186 Next we set out to identify the effects of the 22q11.2 deletion on gene expression and
187 epigenetic profiles. RNA-Seq analysis showed that all of the genes that were expressed
188 within the deletion boundaries in the 22q11.2 region showed decreased expression in
189 22q11.2DS patient lines relative to control cell lines (**Fig. 1e**). Our findings are consistent
190 with a previous study [32].

191 Differential pattern analysis of H3K27ac, H3K27me3 and CTCF by ChIP-Seq showed the

192 majority of binding sites for these essential chromatin marks within the 22q11.2 deletion
193 boundaries with decreased signals in 22q11.2DS patient lines compared to control cell lines
194 (**Fig. 1f,g,h**). These decreased binding signals for such a large region are specific to
195 chromosome 22q11.

196 **Chromosome contacts increase across the 22q11DS breakpoint junction and**
197 **chromatin marks in its flanking regions are affected in a concerted manner**

198 The flanking regions upstream and downstream of the 22q11.2 deletion are brought into
199 close proximity to each other by the formation of the deletion breakpoint junction. We
200 hypothesized that since Hi-C contacts between two given regions will increase with
201 decreasing genomic distance, the contacts between the upstream and downstream deletion-
202 flanking regions in 22q11DS patient cells would be markedly enhanced. We indeed found
203 such stronger contacts between the deletion-flanking regions in 22q11.2DS cell lines (**Fig.**
204 **1a; Supplementary Fig. 3a**). These regions of increased contact extended out to two
205 million base pairs both upstream and downstream of the 22q11.2 deletion boundaries.

206 Following this observation we wanted to examine whether there is an effect on the
207 chromatin marks concurrent to these increased chromosomal contact patterns. To do so we
208 performed enrichment analysis in our ChIP-Seq data for significantly differential signal
209 patterns for H3K27ac, H3K27me3 and CTCF, respectively. We found that both the upstream
210 and downstream flanking regions of the 22q11.2 deletion were enriched with significantly
211 differentially bound sites of both H3K27ac and H3K27me3 (**Fig. 2a,b**). For CTCF the
212 downstream deletion-flanking region was enriched with significantly differentially bound
213 sites (Fisher's exact test p = 2.53E-06) (**Fig. 2c**).

214

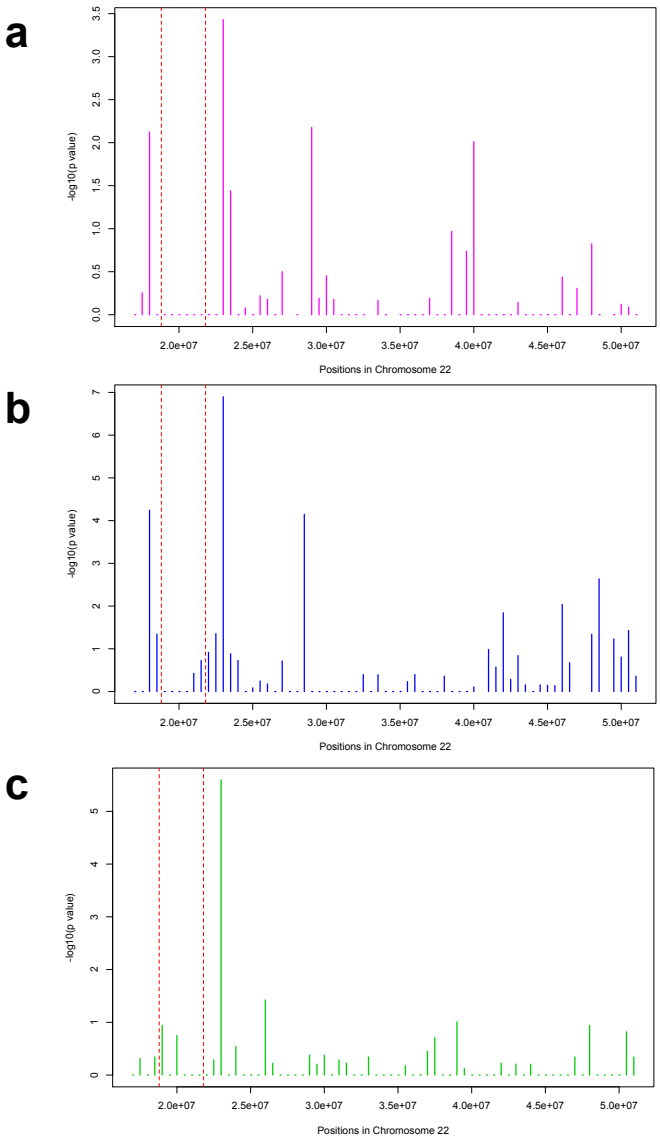


Figure 2. Distribution of significantly differentially enriched H3K27ac histone marks (a), H3K27me3 histone marks (b) and CTCF binding sites (c) along chromosome 22q. *Each point represents a 500 kbp bin. Red dashed lines mark the boundaries of the 3 Mbp deletion in 22q11.2.*

More specifically, within the upstream deletion-flanking region, from 18 to 18.5 Mbp, we found that 5 out of 24 sites with binding by H3K27ac and 6 out of 15 sites with binding

222 by H3K27me3 showed significantly differential binding (Fisher's exact test p = 0.0075 for
223 H3K27ac and p = 5.67E-05 for H3K27me3 respectively). Intriguingly, for H3K27ac all of the
224 5 sites with significant differential binding were bound less strongly while for the same
225 region for H3K27me3 all of the 6 sites with significant differential binding were bound
226 more strongly in the 22q11DS patient lines. For the downstream region, from 23Mb to
227 23.5Mb, 5 out of 13 sites with H3K27ac binding and 11 out of 30 sites with H3K27me3
228 binding showed significantly differential binding (Fisher's exact test p = 0.0004 for
229 H3K27ac and p = 1.26E-07 for H3K27me3 respectively). Again we observed the reciprocity
230 between significantly differential changes for the two different histone marks: 4 out of 5
231 sites for H3K27ac gave a less strong signal while all of the 11 of such sites for H3K27me3
232 showed a stronger signal in the 22q11.2DS patient cell lines.

233 Such reciprocity in signal strengths between these two histone marks is a general
234 feature of their principle of action and makes it much more likely that the observed changes
235 are of physiological relevance.

236 Next we investigated whether there are gene expression changes in these same
237 deletion-flanking regions. Based on the observed changes in histone marks we reasoned
238 that gene expression in these regions could be downregulated in the 22q11DS patient lines,
239 as H3K27ac is associated with active genes and H3K27me3 with repressed genes. We found
240 that among the 4 genes with detectable expression in the deletion-upstream region from
241 18Mbp to 18.5 Mbp, the gene *BID* showed significantly decreased expression in 22q11.2DS
242 patient lines while the other three remained unchanged. In the deletion-downstream
243 region, from 23 Mbp to 23.5 Mbp, only one gene, *IGLL5,* had detectable levels of expression
244 and we found it to exhibit significantly decreased expression in 22q11.2DS patient lines as
245 well.

246 Taken together this part of our study shows that the effect of the 22q11.2 deletion is
247 not limited to the genes within the deletion boundaries but also exerts extensive influence
248 on its flanking regions on the level of intrachromosomal contacts, histone modifications and
249 gene expression.

250 **A/B compartments of the chromatin and the topological domain structure of the**
251 **nuclear genome are affected by the 22q11.2 deletion**

252 Previous studies of cell lines without large CNVs, have revealed that there is one level of
253 organization where the chromatin is partitioned into two compartments, A and B, which
254 correspond to active and inactive regions, respectively [20], and which can be derived from Hi-
255 C data. We sought to determine whether the 22q11.2 deletion might lead to changes in
256 these A/B compartments. Indeed when computing A/B compartments from our Hi-C data
257 we observed that the eigenvectors changed signs in the 22q11.2 del region while all the
258 other regions remained the same between 22q11.2DS patient and control cell lines (**Fig.**
259 **3a**). The original compartment that is spanning the deletion region in control cells is being
260 partitioned into two shorter compartments with different eigenvectors signs, signifying a
261 change in the A/B compartment structure in the deletion region.

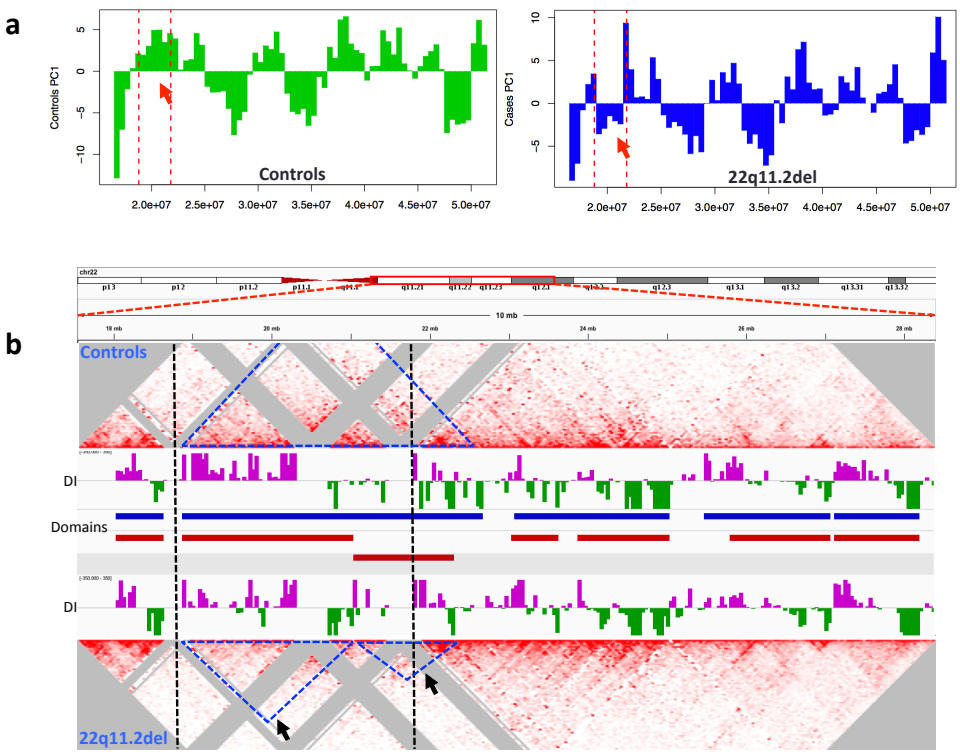


Figure 3. Change in A/B compartments and topological domain signal on chromosome 22q11. *(a) The A/B compartment spanning the 22q11 deletion in controls is partitioned into two compartments in the patient cell lines (red arrows). Red dashed lines: boundaries of 22q11.2del region. Shown is the first eigenvector for the principal component analysis of the normalized contact matrix. X-axis: position on chromosome 22q. Y-axis: value of the first eigenvector. (b) Normalized Hi-C contacts for 22q11.2del and control cell lines, displayed as a triangular heat-map that is overlayed on directionality indices (DI) and chromatin domains. The signal for the topological domain spanning the 22q11.2 deletion and its downstream region (blue triangle in controls) was broken into two shorter signals (blue triangles and black arrows in 22q11.2del). Black dashed lines: boundaries of 22q11.2del region. Regions with low mappability of reads were removed from the analysis and are shown in grey.*

274 Topological domains are a megabase-sized structural feature of the genome
275 organization that is constituted of highly self-interacting chromosome regions [33]. Using our
276 Hi-C and ChIP-Seq data we identified clear changes of the topological domain structure in
277 the deletion-downstream flanking region of 22q11.2DS cell lines while the deletion-
278 upstream topological domain structure was preserved (**Fig. 3b**). The topological domain
279 spanning the 22q11.2 deletion and its downstream region was broken into two shorter
280 domains across the downstream boundary of the deletion. Of note, this effect of the 22q11.2
281 deletion on the topological domain structure extended to 2 million base pairs downstream
282 of the deletion region, where we observed another topological domain breaking into two
283 shorter ones. Our results demonstrate that the 22q11.2 deletion can result in changing the
284 order of topological domains both in the deletion region proper and its downstream
285 regions.

286 **Change of chromosomal *cis*-contacts involving the downstream flanking region of**
287 **22q11.2del**

288 Based on the above findings that the genomic region downstream of the 22q11.2 deletion
289 was affected on different levels of molecular regulation in 22q11.2DS patient cell lines, we
290 next sought to investigate whether the presence of the 22q11.2 deletion also affected the
291 chromosomal *cis*-contacts of the downstream deletion-flanking region with any other
292 region on chromosome 22q. To do so, we analyzed our Hi-C data by calculating the fold
293 change of contacts between the 21.5-22 Mbp window, which is situated right downstream
294 of the 22q11.2 deletion, and all the other 1 Mbp-sized regions on chromosome 22q. As
295 expected, we observed that region 17-18 Mbp, i.e. the breakpoint-proximal region right
296 *upstream* of the deletion, showed the largest fold change (2.04) of contact with region 21.5-
297 22 Mbp between 22q11DS and control cell lines, given that these regions were brought into

298 close proximity to each other by the 22q11.2 deletion (**Fig. 1a**).

299 The second largest fold change (1.96) of *cis*-contacts involving region 21.5-22 Mbp was
300 for contacts with region 50-51 Mbp, i.e. towards the very telomeric end of chromosome 22q
301 (**Fig. 1a**). Intriguingly, we also observed strong positive correlation between region 22-22.5
302 Mbp and region 50-50.5 Mbp for CTCF binding (Pearson's $r = 0.933$, $p = 0.02$)
303 (**Supplementary Fig. 5a**) and between region 22-22.5 Mbp and region 50.5-51 Mbp for
304 H3K27ac enrichment (Pearson's $r = 0.811$, $p < 0.05$) (**Supplementary Fig. 5b**).
305 Furthermore there was weak correlation between region 22-22.5 Mbp and 50.5-51 Mbp for
306 H3K27me3 enrichment (Pearson's $r = 0.74$, $p = 0.090$) (**Supplementary Fig. 5c**). Given that
307 both region 21.5-22 Mbp and region 22-22.5 Mbp are in the downstream flanking region of
308 the 22q11.2 deletion, our results strongly imply that increased chromosomal contacts
309 between the downstream deletion-flanking region and the telomeric region 50-51 Mbp may
310 be associated with the differential changes of histone modifications and CTCF binding that
311 we found to be in correlation between these two regions.

312 To validate the increased chromosomal contact between the downstream deletion-
313 proximal region and the telomeric end of 22q we performed DNA FISH on 5 cell lines with
314 the 22q11.2 deletion and on 6 control cell lines, using FISH probes to loci in region 21.8-
315 22.5 Mbp (RP11-47L18: chr22: 21,931,796 – 22,118,344) and in region 50-51Mbp (RP11-
316 125K3: chr22: 50,149,996- 50287311). FISH showed that the downstream flanking region
317 of the 22q11.2 deletion is significantly closer to region 50-51 Mbp in nuclear space in
318 22q11DS cell lines than in control cell lines (p value 0.008) (**Fig. 4; Supplementary Fig. 6**).
319 Taken together, our results strongly indicate that the 22q11.2 deletion causes
320 conformational changes on several levels on chromosome 22q.

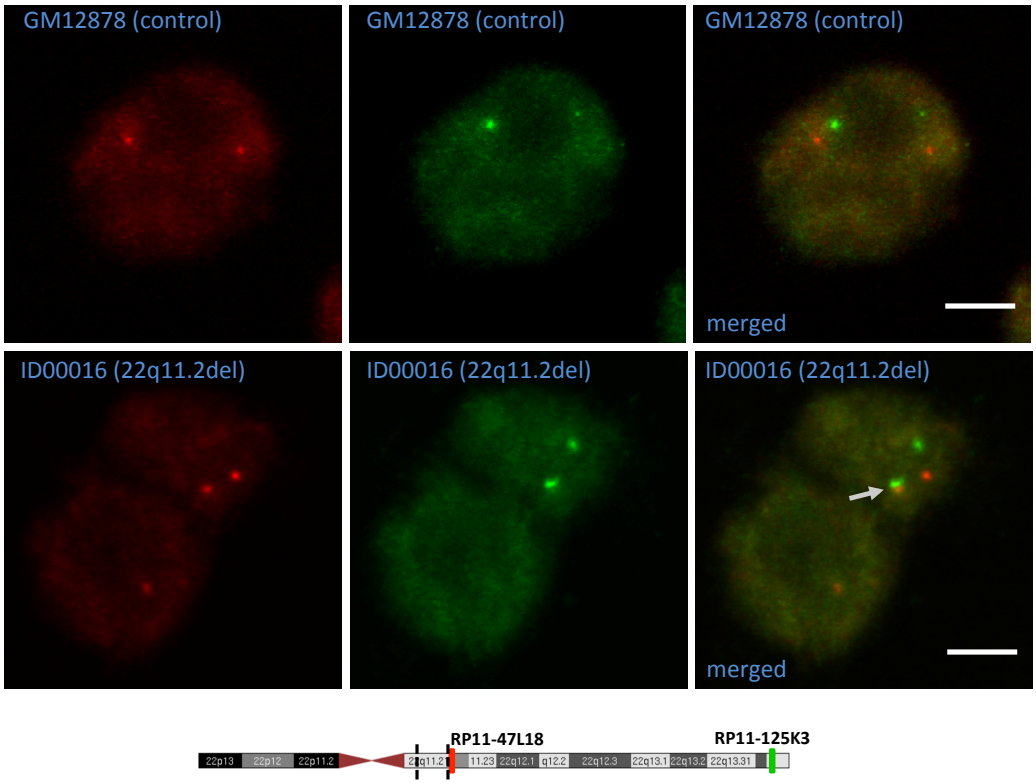
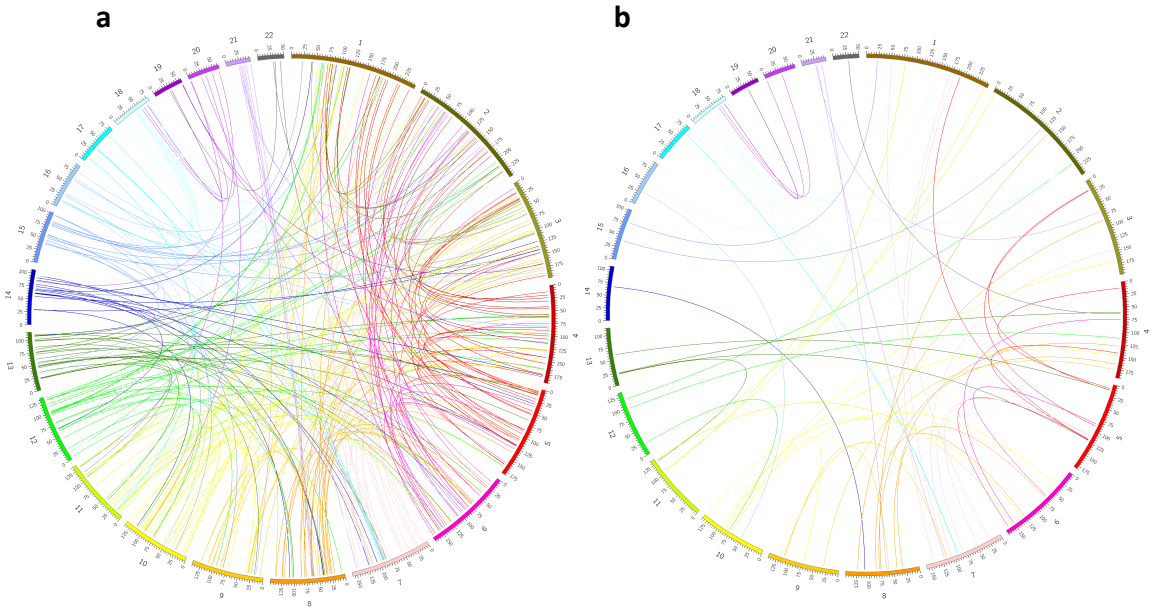


Figure 4. Examples of 3D FISH visualization of intra-chromosomal interaction changes. *The Hi-C predicted interaction regions of 21.8-22.5 Mbp and 50-51 Mbp on chromosome 22q were visualized by 3D DNA FISH using BAC probes RP11-47L18 and RP11-125K3, labeled with biotin (red) or digoxigenin (green). Arrow: an example for the red and green FISH probes in close proximity in a cell carrying the 22q11.2 deletion. Scale bars are 5 µm. Red and green bars on chromosome 22q indicate the locations of the biotin and digoxigenin FISH probes, respectively. Black dashed lines indicate the position of the 3 Mbp deletion.*

Global changes of interchromosomal contact patterns

To explore the effect of the 22q11.2 deletion on *trans*-contacts, i.e. between any non-homologous autosomes in the nucleus of 22q11.2DS patient lines, we analyzed our Hi-C data for significantly different *trans*-contacts between 22q11DS and control cell lines. On

333 the genome-wide level, we found 272 *trans*-contacts with a p-value of less than 0.0001 (**Fig.**
334 **5a**). Interestingly the majority of these chromosomal *trans*-contacts did not involve
335 chromosome 22q as one of the interacting partners. Notably, 56 of these interchromosomal
336 contact signals are among the top 5% of the strongest *trans*-contacts (**Fig. 5b**). This
337 enrichment of significantly different genome-wide chromosomal *trans*-contacts within the
338 strongest *trans*-contacts is again highly statistically significant (Fisher's exact test $p < 2.2e-$
339 16). We found many fewer significantly different genome-wide chromosomal *trans*-contacts
340 by randomly swapping the assignment of 22q11.2 deletion and control status across our Hi-
341 C data sets, by comparing within control cell lines and by comparing within 22q11.2del cell
342 lines (**Supplementary Table 3**). None of these swapping analyses achieved the same
343 enrichment. This indicates that a relevant amount of the significantly different genome-
344 wide chromosomal *trans*-contacts is not due to random chromosomal motion or to as-of-yet
345 unknown factors such as cell-culture variations across the lymphoblastoid cell lines. Rather,
346 our analysis points to a genome-wide disturbance of the network of chromosomal *trans*-
347 contacts that is at least in part attributable to the presence of the 22q11.2 deletion on
348 chromosome 22q.



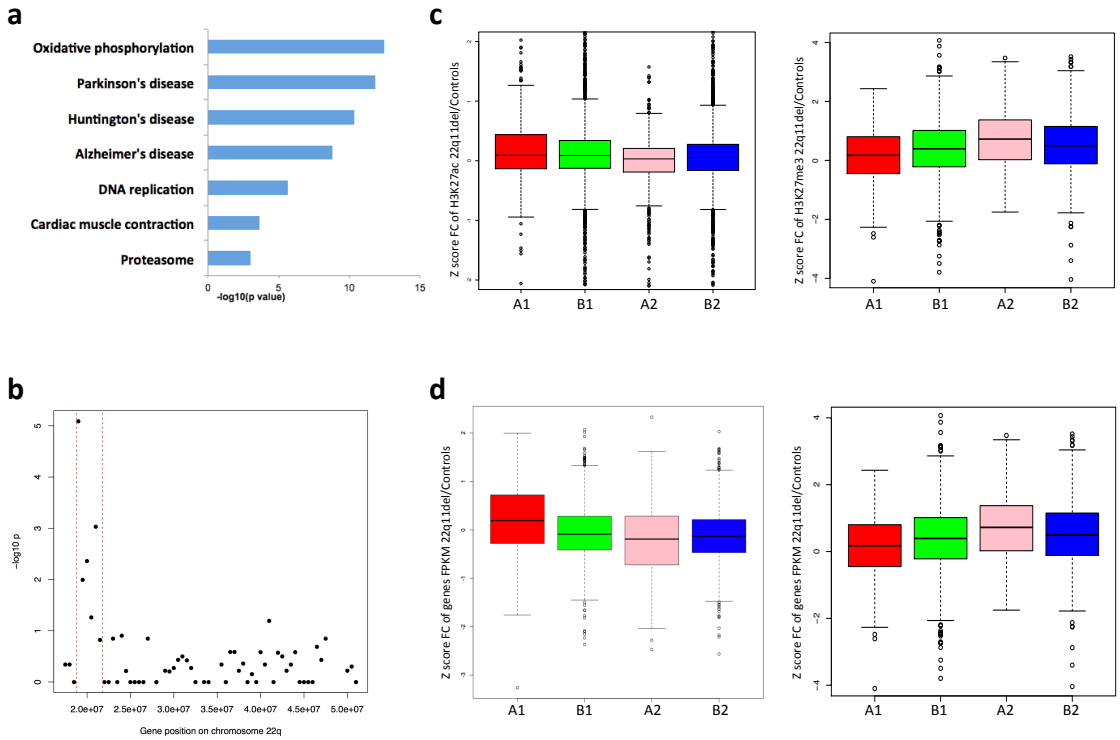
349

350 ***Figure 5. Genome-wide inter-chromosomal contact changes (either more or less***
351 ***frequent) determined by Hi-C analysis in 22q11.2del versus control cell lines.*** *(a) Circos*
352 *plot of the inter-chromosomal contacts exhibiting differential interaction in 22q11.2del versus*
353 *control cell lines at significance level of 0.0001. (b) Circos plot of the inter-chromosomal*
354 *contacts exhibiting differential interaction in 22q11.2del versus control cell lines at*
355 *significance level of 0.0001 and showing only the top 5% strongest inter-chromosomal*
356 *contacts.*

357 **Global changes of gene expression patterns**

358 To investigate the global effect of the 22q11.2 deletion on gene expression, we performed
359 differential expression analysis between the 22q11.2DS and control cell lines. Of the 11,374
360 genes with detectable levels of expression (FPKM > 0.5), 1,610 genes showed significantly

differential expression (FDR < 0.05). Gene ontology analysis indicated that these differentially expressed genes are enriched for genes involved in mitochondrial pathways such as the respiratory chain (n = 32, p value = 2.96E-11). KEGG pathway analysis demonstrated enrichment of genes involved in oxidative phosphorylation neurodegenerative diseases such as Parkinson's disease (**Fig. 6a**). Earlier studies have noted that there are several genes related to mitochondrial function that are located within the 22q11 deletion boundaries [34], therefore the pathways associated with this cellular function would have a high likelihood to be affected by the change in copy number. The enrichment for pathways related to neurodegenerative disorder is encouraging, as an association between 22q11DS and Parkinson's Disorder has been reported previously [35-37]. This reinforces the notion that the LCL cell culture model can be of relatively more relevance for the molecular study of 22q11DS in general.



374 ***Figure 6. Differential expression analysis and correlation between gene expression and***
375 ***epigenetic profiles.*** *(a) KEGG pathway analysis for genes differentially expressed between*
376 *22q11.2del and control cell lines. (b) Enrichment analysis for chromosome 22q for*
377 *differentially expressed genes. Red dashed lines mark the boundaries of 22q11.2del region. (c)*
378 *Differentially expressed genes correlated with different H3K27ac and H3K27me3 signal*
379 *strengths in 22q11.2del versus control cell lines. A1 represents genes significantly up-regulated*
380 *in 22q11.2del cell lines while B1 represents non-significantly up-regulated genes; A2*
381 *represents genes significantly down-regulated in 22q11.2del cell lines while B2 represents*
382 *non-significantly down-regulated genes. The y-axis shows the z-score transformed fold change*
383 *of H3K27ac signals (left) or H3K27me3 signals (right) between 22q11.2del and control cell*
384 *lines. (d) Differential H3K27ac and H3K27me3 signal strengths exhibited correlation with*
385 *differential gene expression between 22q11.2del and control cell lines. A1 represents genes*
386 *whose TSSs were significantly more marked by H3K27ac (left) or by H3K27me3 (right) in*
387 *22q11.2del cell lines. B1 represents genes with TSSs non-significantly more marked by*
388 *H3K27ac (left) or by H3K27me3 (right). A2 represents genes whose TSSs were significantly*
389 *less marked by H3K27ac (left) or by H3K27me3 (right) in 22q11.2del cell lines. B2 represents*
390 *genes with TSSs non-significantly less marked by H3K27ac (left) or by H3K27me3 (right). The*
391 *y-axis shows the z-score transformed fold change of genes' FPKM in 22q11.2del versus control*
392 *cell lines.*

393 We next sought to identify whether genome-wide there were entire genomic regions
394 that are enriched for differentially expressed genes. As expected, the most significant
395 signals for this analysis were located in the 22q11.2 region (**Fig. 6b; Supplementary Fig.**
396 **7**). No other regions genome-wide achieved FDR corrected significance (**Supplementary**
397 **Fig. 7**).

398 **Correlation between histone modification and gene expression**

399 To examine whether the gene expression changes are consistent with the histone
400 modification changes, we assigned the H3K27ac and H3K27me3 peaks to their nearest
401 genes based on the distance to their transcription start sites and for each gene, we only
402 retained the closest peak for both histone marks. We observed that significantly
403 upregulated genes in 22q11.2DS cell lines exhibited significantly higher fold change of
404 H3K27ac enrichment (permutation test p = 0.0140) and significantly lower fold change of
405 H3K27me3 enrichment (permutation test p = 0.0018) than genes non-significantly
406 upregulated in 22q11.2DS cell lines (**Fig. 6c**). Consistently, significantly downregulated
407 genes in 22q11.2DS cell lines exhibited significantly lower fold change of H3K27ac
408 enrichment (permutation test p = 0.0002) but significantly higher fold change of H3K27me3
409 enrichment (permutation test p = 0.0227) than genes non-significantly down regulated in
410 22q11.2DS cell lines (**Fig. 6c**).

411 Moreover, genes whose transcription start sites (TSSs) showed significantly
412 upregulated binding by H3K27ac in 22q11DS cell lines exhibited significantly higher fold
413 change of expression (permutation test p < 2.2e-16) than those with non-significantly
414 upregulated binding, while those TSSs showing significantly downregulated binding
415 exhibited a significantly lower fold change of expression (permutation test p = 0.0202) than
416 those with non-significantly downregulated binding (**Fig. 6d**). Consistently, genes whose
417 TSSs showed significantly upregulated binding by H3K27me3 in 22q11DS showed
418 significantly lower fold change of expression (permutation test p = 0.0037) while those with
419 significantly downregulated binding exhibited significantly higher fold change of expression
420 (permutation test p = 0.0027) between 22q11DS and control cell lines (**Fig. 6d**).

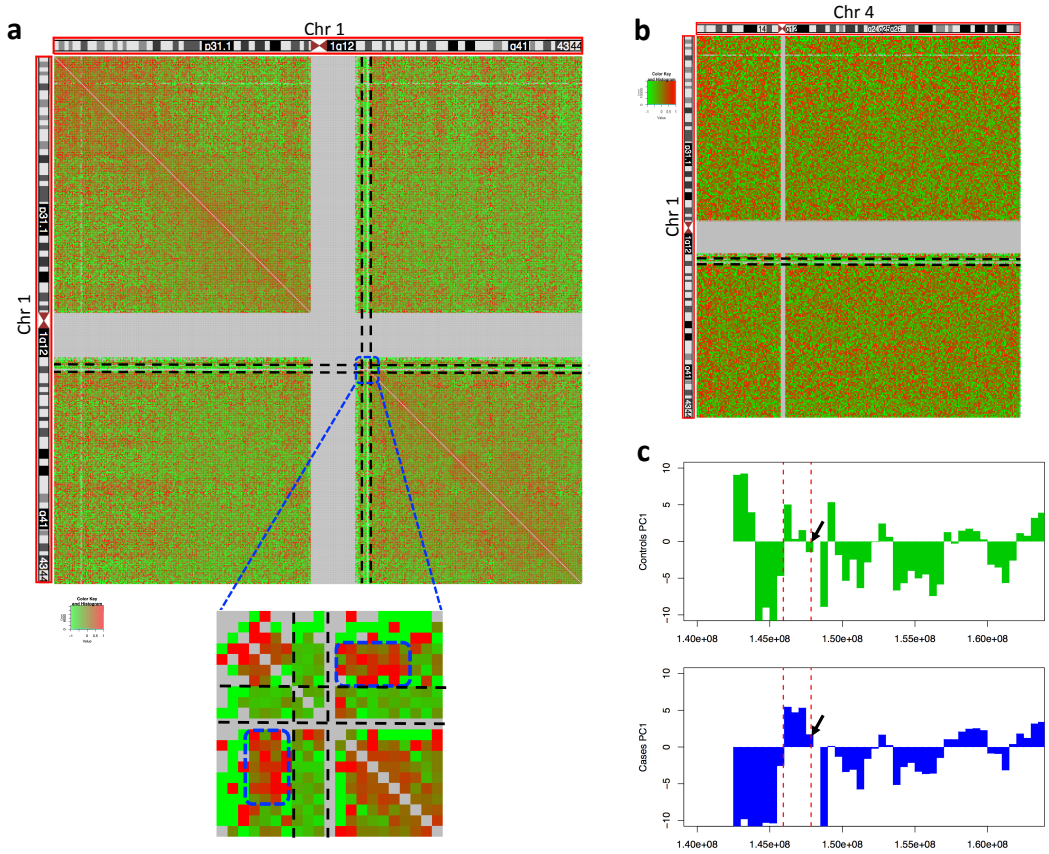
421 We also calculated the correlation between gene expression and H3K27ac binding

422 affinity across the individuals for all the genes with TSSs bound by H3K27ac. In line with the
423 above results, we found significantly higher Pearson's correlation coefficients than the
424 coefficients obtained from permutations (Wilcoxon rank sum test $p < 2.2e-16$)
425 (**Supplementary Fig. 8a**). This difference was even more significant when we included the
426 differential expressed genes (absolute fold change > 2) only (Wilcoxon rank sum test $p <$
427 $2.2e-16$) or the differential H3K27ac bound genes only into our analysis (Wilcoxon rank
428 sum test $p < 2.2e-16$) (**Supplementary Fig. 8a**). Unsurprisingly, we also observed
429 significantly higher Pearson's correlation coefficients between gene expression and
430 H3K27me3 binding than the coefficients obtained from permutations for the differential
431 expressed genes (absolute fold change > 2) only (Wilcoxon rank sum test $p = 9.72e-04$) and
432 the differential H3K27me3 bound genes only (Wilcoxon rank sum test $p = 3.71e-06$)
433 (**Supplementary Fig. 8b**). Together, our results demonstrated that gene expression
434 changes are associated with histone modification changes in 22q11DS cell lines compared
435 with control cell lines.

436 **Effects of a large CNV on chromosome 1q21.1 on chromosome folding**

437 To explore whether large CNVs other than the one on chromosome 22q11.2 can lead to
438 changes in the patterns of chromosome folding, we performed Hi-C on two lymphoblastoid
439 cell lines with a heterozygous deletion of approximately 1.35 Mbp in size on chromosome
440 1q21.1 (1q21.1del). This deletion is strongly associated with the development of
441 schizophrenia [38-40]. As for the 22q11.2 deletion, we observed that both *cis*-contacts and
442 *trans*-contacts between the 1q21.1 deletion regions and other regions were decreased in
443 1q21.1del cell lines relative to control cell lines (**Fig. 7a,b**). As observed in the 22q11DS cell
444 lines, in the 1q21.2del cell lines there was an increase of intrachromosomal contacts
445 between the regions directly flanking the main CNV of 1q21.1 (**Fig. 7a**). Analysis of A/B

446 compartments showed that the eigenvector sign of the downstream boundary region of the
447 1q21.1 deletion was inverted (**Fig. 7c**), showing that also on the level of chromatin
448 compartmentalization the large CNV on 1q21.2 can have effects, similar to the large CNV on
449 22q11.2. Taken together the large deletion CNV on chromosome 1q21.1 exhibited similar
450 effects on chromosomal *cis*-contacts and *trans*-contacts as the large deletion CNV on
451 chromosome 22q11.2, which points towards our findings in 22q11DS being generalizable
452 across large neuropsychiatric CNVs.



453

454 ***Figure 7. Effect of the large deletion CNV on 1q21.1 on chromosome conformation.*** *(a)*
455 *Fold change of cis-contacts of chromosome 1 in 1q21.1del versus control cell lines. Blue boxes*
456 *mark the regions of increased contacts between the upstream and downstream flanking*
457 *regions of the 1q21.1 deletion. (b) Fold change of trans-contacts between chromosome 1 and 4*

458 *in 1q21.1del cell lines versus control cell lines. Black dashed lines mark the boundaries of the*
459 *1q21.1 deletion. Each cell in the heatmap represents the inter-chromosomal contact level*
460 *between two 500 kbp regions. (c) Change in A/B compartments on chromosome 1q21. Upper*
461 *panel: control cell lines. Lower panel: 1q21.1del cell lines. Shown is the first eigenvector for the*
462 *principal component analysis on the normalized contact matrix of chromosome 1, for*
463 *1q21.1del and control cell lines. X-axis: position on chromosome 1q. Y-axis: value of the first*
464 *eigenvector. Black arrows indicate the location of change in the eigenvector sign in the*
465 *downstream boundary region of 1q21.1, in 1q21.1del cell lines.*

466 **Discussion**

467 Large CNVs are an important feature of the genetic architecture of several major
468 neurodevelopmental psychiatric disorders as well as of conditions involving aberrant
469 morphology of many organ systems. Their effects on the level of phenotype are complex and
470 the molecular mechanisms mediating these effects are very incompletely understood. While
471 it is still a good assumption that a considerable portion of these mechanisms are a direct
472 consequence of the copy number change of the genes within a given CNV's boundaries and
473 the resulting changes in expression levels for these genes, it also seems plausible to
474 investigate whether additional levels of complexity exist regarding the effects of a large CNV
475 across multiple layers of the control of gene activity. This plausibility stems from several
476 observations and lines of reasoning. Namely there are large numbers of genes affected
477 genome wide and far distal from the main CNV, and expression changes within CNV
478 boundaries do not always follow exactly the change in copy number. This leads us to
479 consider the basic principles that govern organizational features of the nucleus and the
480 chromatin such as regulatory domains being bounded by protein factors that recognize
481 binding sites in the DNA sequence which in turn could be affected (i.e. deleted or

482 duplicated) by a CNV, or long-range chromosome contacts being also influenced by the
483 length of the involved chromosomes (and that length being changed by large CNVs). Finally
484 there have been several recent reports about mutations in chromatin remodeling genes in
485 the context of neuropsychiatric disorders pointing to the importance of proper molecular
486 management on the epigenomic level in these conditions.

487 Here we studied the effects of the important large CNV on chromosome 22q11.2, and in
488 a more limited fashion the effects of the large CNV on chromosome 1q21.2, on chromatin
489 conformation including long-range chromosome contacts and domain formation, epigenetic
490 profiles and gene expression. To our knowledge, this is the first study of changes in
491 chromatin conformation and epigenetic profile changes caused by deletions on 22q11.2,
492 and 1q21.2, or any such large CNVs with strong relevance to organ development and in
493 particular brain development and neuropsychiatric disorders. We observed dosage effects
494 of the large CNVs on long-range chromosome contacts, chromatin domains and epigenetic
495 profiles as well as on gene expression. More specifically we found increased contacts
496 between upstream and downstream flanking regions of the 22q11.2 deletion in 22q11DS
497 cell lines in contrast with control cell lines. Interestingly, both upstream and downstream
498 regions of the 22q11.2 deletion were enriched with differentially binding sites of H3K27ac
499 and H3K27me3 while downstream regions were also enriched with differentially binding
500 sites of CTCF. Importantly, the gene expression changes in these regions were consistent
501 with the histone modification changes as less H3K27ac binding and more H3K27me3
502 binding reflect decreased gene expression in 22q11Ds cell lines. Based on the above
503 findings, it is tempting to speculate that increased contacts between upstream and
504 downstream flanking regions of the 22q11.2 deletion contribute to gene expression changes
505 via the coordinated changes of epigenetic profiles in these flanking regions.

506 Although we found chromosome contacts to be correlated with both the epigenetic
507 profiles and the gene expression changes in the flanking regions of the 22q11.2 deletion, we
508 did not observe significant correlation either between chromosome contact changes and
509 gene expression changes or between chromosome contact changes and epigenetic profile
510 changes (data not shown) on the genome-wide level. Moreover, we did not observe the end
511 of chromosome 22 (50-51Mb) to be enriched with differentially expressed genes (2 out of
512 24 genes) despite the Hi-C finding, validated by FISH, of significantly increased contacts
513 with the downstream flanking region of the 22q11.2 deletion in 22q11.2DS cell lines. Our
514 results indicate that chromosome conformation changes caused by the 22q11.2 deletion
515 contribute to gene expression changes and epigenetic profile changes between 22q11.2DS
516 and control cell lines but not always in a clearly and linearly deterministic way.

517 However, we did observe significant positive correlation between H3K27ac changes
518 and gene expression changes but negative correlation between H3K27me3 and gene
519 expression changes on the genomic level. These findings demonstrated that the epigenetic
520 profiles were reshaped genome-wide to cause extensive gene expression changes in
521 22q11.2DS cell lines.

522 The dosage effects of the deletion on *cis*- and *trans*-contacts and the increased
523 chromosomal contacts between upstream and downstream deletion-flanking regions in
524 22q11Ds cell lines were also observed in cell lines with the 1q21.1 deletion. The similar
525 results that were obtained on that level of observation for the 22q11.2 deletion and the
526 1q21.1 deletion would suggest that these changes caused by a large deletion CNV are
527 universal for such chromosomal aberrations instead of specific to the 22q11.2 deletion.

528 As a cautionary note on the technical level, we demonstrated here that for genomes
529 with a large deletion CNV the appropriate normalization methods for Hi-C data have to be

530 chosen with great care to avoid false findings. For instance we would have reached the
531 conclusion that the chromosomal contacts within the deletion regions are not decreased in
532 cell lines with deletion compared with control cell lines if the hiclib software package [22] had
533 been used for normalization of Hi-C data, instead of the hicpipe package [21].

534 Lastly, we also observed A/B chromatin compartment changes and changes in the
535 signal from topological domain analyses across the 22q11.2 deletion region. This is the first
536 time that such a phenomenon has been reported in this context. Earlier, Lupiáñez et al. [41],
537 using the 4C assay, demonstrated in a mouse model as well as in human lines from patients
538 with rare malformations of the limbs, that large structural changes in the genomic sequence
539 can affect the topological domain architecture directly on top of the sequence change, giving
540 further support to our Hi-C based findings in 22q11.2 and 1q21.1 deletion lines. It is
541 important to remember that, just as for earlier studies that were reporting on A/B
542 compartments and topological domains (in non-CNV genomes), our analysis resulted from
543 combining the sequencing data of two homologous chromosomes. Only once haplotype
544 phasing of the genomes in question will be available will one be able to determine whether
545 the chromosome 22q with large deletion or the one without large deletion, or some
546 combination of the two, is driving such changes in the domain signals (the deletion-carrying
547 domain being the main cause for the change in signal certainly being the initial hypothesis).
548 Whether the observed chromosomal contact changes on chromosome 22q are existent only
549 on one chromosome 22q (again probably the one carrying the large CNV) or emanate from
550 both of the two homologous chromosomes is currently unknown and will be worth
551 investigating when haplotype-phasing data is available.

552 In summary, we found dosage effects of the large deletion CNV on chromosome
553 22q11.2 on long-range chromosome contacts, chromatin organization, epigenetic profiles

554 and gene expression. Extensive changes on these levels caused by the 22q11.2 deletion are
555 global and seem to be rippling along the entire chromosome carrying the deletion as well as
556 across the entire nucleus rather than being confined to the deletion region only. Such effects
557 had never been shown before outside of cancer cell lines [42]. Furthermore, in contrast to the
558 findings in cancer cells we used a larger cohort of individual patient cell lines, with all lines
559 carrying only one main large CNV that is clearly and strongly associated in a causative
560 manner with a neurodevelopmental phenotype. This, in contrast to the two cancer cell lines
561 that each carried multiple large CNVs which could also have been a consequence rather
562 than a cause of the disease phenotype, makes it much more likely that the higher-order
563 effects of the large CNVs that we observed may be contributing to the molecular etiology of
564 the disorders in question. This point is further strengthened by another recent paper [43],
565 where the authors describe studying the effects of large CNVs on chromosome 16p11 on
566 chromosome interactions. These large CNVs are almost as strongly associated with
567 neurodevelopmental disorders than the large CNV on 22q11.2. While the study on the
568 16p11 CNVs used a somewhat smaller number of cell lines than our study and also used 4C
569 as a method of discovery, which is, unlike the Hi-C method used by us, not able to detect
570 changes in a global and unbiased fashion, it is one more independent piece of supporting
571 evidence for the biological validity and general relevance of the findings which we describe
572 here.

573 While we were able to show possible correlations across several pairs of the molecular
574 levels that we assayed in this study there are other combinations of molecular levels that
575 show no obvious connection in our data. We believe that this could be a function of either
576 the developmental time point or the cell type, or both, being removed from those where the
577 22q11.2 and 1q21.1 deletions most likely exert some of their strongest effects (e.g. during
578 embryonal development and in cells of the developing central nervous system). The clear

579 effects that we were able to observe in lymphoblastoid cell lines could represent the
580 afterglow of a molecular tragedy that played out earlier in the development of various
581 organs in the patients carrying these deletions. At the same time the strength of this
582 afterglow would hint at the strength of the effects that impacted across various levels of
583 molecular control and gene regulation by CNVs of this size.

584 **Methods**

585 **Cell lines and data generation for this study.** All cell lines were either acquired from the
586 Coriell cell repository (cell lines IDs starting with GM or ID) or were taken from the
587 Molecular Genetics of Schizophrenia (MGS) cohort (dbGaP Study Accession:
588 phs000167.v1.p, cell lines 52425 and 82699), and were appropriately consented. The Hi-C
589 assay was carried out according to the original protocol [20], with several modifications. ChIP-
590 Seq was performed as described in previous studies [44,45]. The RNA-seq libraries were
591 generated according as in [46]. All experimental procedures are described in the
592 **Supplementary Information**.

593 **Hi-C Data analysis.** All Hi-C data were produced using Illumina paired-end sequencing with
594 a read length 2 x 101 bp. As there might be ligation junctions in the reads, we performed
595 iterative mapping using bowtie2 as in Imakaev et al. 2012 [22]. Briefly, we computationally cut
596 all the reads to 25 bp first and mapped them to human genome (hg19). Then we extended
597 the non-uniquely mapped reads by 5 bp to 30 bp and mapped them again. This process was
598 repeated until the read length was extended to 101 bp. This iterative mapping did improve
599 the mapping rate (**Supplementary Table 4**). Each read end was mapped separately using
600 the single ends mode. Only uniquely mapping reads were used and PCR duplicate read pairs
601 were removed. We only included autosomes in our study. The filtered contact number is
602 listed in **Supplementary Table 5**.

603 We compared three different data normalization methods: hiclib [22], hicpipe [21] and
604 HiCNorm [23]. All of the three tools were run using the default parameters except for the
605 segment length threshold being set to 600 bp. We chose normalized metrics on hicpipe for
606 the following analyses, using a bin size of 40 kb for topological domain analysis and of 500
607 kb for the other analyses. The total number of contacts was normalized for each sample
608 before combining cell lines in each category (control, 22q11del and 1q21del, respectively).
609 Fold changes of log2 transformed mean contacts between deletion cell lines and control cell
610 lines were calculated by (deletion – control)/control.

611 To identify the differential interchromosomal contacts, we only included contacts with
612 at least 1 supporting read pair in each of the cell lines. Differential contacts analysis was
613 conducted by Student's t-test using the normalized metrics. Fisher's exact test was used to
614 assess the enrichment of differential contacts within the top 5% strongest contacts. We also
615 performed the same analysis by permuting the control and 22q11del status of the cell lines
616 ten times. Comparison within control cell lines and within 22q11del cell lines were
617 performed by randomly dividing the cell lines into two groups three times.

618 Identification of A and B compartments was performed as in Lieberman et al. 2009 [20]
619 and topological domains were identified as in Dixon et al. 2012 [33]. We modified the
620 algorithm for identifying topological domains such that when calculating the direction index
621 for a given bin we will exclude neighboring bins that are spanned by segmental duplications
622 as well as the bins on the opposite side of the given bin and equidistant to the given bin than
623 the bins spanned by segmental duplications (even if such symmetrical bins are not spanned
624 by segmental duplications themselves). For topological domain analysis, we pooled the raw
625 data of all the samples for each category to obtain sufficient Hi-C sequencing coverage
626 before normalization.

627 **3D FISH.** Two human DNA BACs (clones RP11-47L18 and RP11-125K3) covering two
628 distinct regions of chromosome 22 were labeled with biotin or digoxigenin (DIG) by nick
629 translation kit (Roche Applied Sciences) to make FISH probes (Roche). In-situ hybridization
630 was performed according to the method published by [47], with several modifications.
631 Experimental details and data analysis could be found in the **Supplementary Information**.

632 **Differential expression analysis.** All RNA-seq data were generated using Illumina paired-
633 end sequencing with read length 101 bp. Reads were mapped to hg19 and transcriptome
634 reference with TopHat 2 [48]. TopHat 2 was run using default parameters but with the
635 coverage search being turned off. The mapped reads were analyzed by Cufflinks [49].
636 Differential expression was estimated with Cuffdiff 2 [50]. We excluded the genes with low
637 expression (FPKM < 0.5) from downstream analysis.

638 Pathway analysis of significantly differential expressed genes was conducted with
639 DAVID [51] using all the expressed genes as background.

640 **ChIP-Seq data analysis.** All ChIP-Seq data were generated using Illumina paired-end
641 sequencing with read length 101 bp. Reads were aligned to hg19 with BWA-MEM using
642 default parameters [52]. Reads with low mapping quality (< 30) were removed. PCR duplicate
643 reads were removed using Picard (http://broadinstitute.github.io/picard). As a quality
644 control, we calculated the normalized strand cross-correlation coefficient (NSC) and relative
645 strand correlation (RSC) [53] to assess the signal-to-noise ratios. All the data showed higher
646 NSC than RSC (**Supplementary Fig. 9**). Replicates for the same cell line on average showed
647 higher correlation than datasets from different cell lines (**Supplementary Fig. 10**).

648 For CTCF and H3K27ac, we used MACS2 [54] to call narrow peaks with default
649 parameters. For H3K27me3, we used the broad peak calling in MACS2. For all peak calling,

650 we used the corresponding whole-cell extract input library as background. For differential
651 bound analysis, we used the R package DiffBind [55] with the effective library size for read
652 counts normalization. Then DBA_DESEQ2 method was employed to conduct the differential
653 bound analysis. Signal artifact blacklist regions were excluded from our analysis
654 (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapabil
655 ity).

656 **Enrichment analysis.** For the enrichment analysis of differentially expressed genes we
657 divided each chromosome into 500 kbp bins. Within each bin, we calculated the total
658 number of expressed genes and the number of genes with significantly differential
659 expression between 22q11del and control cell lines. Then we conducted Fisher's exact test
660 to identify bins enriched with significantly differentially expressed genes against the
661 background of the whole genome.

662 For the analysis of differentially enriched sites for CTCF, H3K27ac and H3K27me3, we
663 also used 500 kbp bins. Log2 transformed fold changes of normalized read numbers in
664 binding sites between 22q11del and control cell lines were further transformed to Z-scores.
665 We considered binding sites with Z-score >2 or < -2 as significantly bound sites. Then
666 within each 500 kbp bin, we calculated the total number of binding sites and the number of
667 significantly differentially enriched sites between 22q11del and control cell lines. Then we
668 conducted Fisher's exact test to identify bins with significantly differentially enriched sites
669 against the background of the whole genome.

670 **Correlation analysis of gene expression and histone modification.** To estimate the
671 correlation between gene expression and histone modification, we assigned each binding
672 site of H3K27ac and H3K27me3 to its nearest ENSEMBL TSS using the R package
673 ChIPpeakAnno [56]. If a TSS was associated with multiple binding sites, only the nearest

674 binding site was retained. We then used this assignment for downstream correlation
675 analysis. To determine the cutoff for the distance in which binding sites are associated with
676 TSSs, we plotted the distribution of distances between binding sites and their assigned TSSs
677 (**Supplementary Fig. 11**). Based on the distribution, we set the cutoff to distance to TSS to
678 ± 1 kbp for H3K27ac and to ± 5 kbp for H3K27me3.

679 To investigate the effects of histone modifications on gene expression, we divided the
680 genes into two categories based on the differential expression analysis: differentially
681 expressed genes (FDR < 0.05) (category A) and not differentially expressed genes (FDR >
682 0.05) (category B). Within each category, we further categorized the genes into two groups:
683 up-regulated expressed genes (A1, B1) and down-regulated expressed genes (A2, B2) in
684 22q11del cell lines relative to control cell lines. Then for the genes within each of the four
685 groups (A1, B1, A2, B2), we calculated the Z-score transformed fold changes of the
686 normalized read counts in the TSS binding sites of histone marks between 22q11del and
687 control cell lines. To obtain the statistical significance of the fold change differences
688 between A1 and B1 and between A2 and B2, we performed permutation tests with 9,999
689 permutations.

690 We also carried out the reverse analysis. TSS binding sites of histone marks were
691 partitioned into two categories based on the differential bound analysis: differentially
692 bound sites (|Z-score| > 1 for H3K27ac, |Z-score| > 2 for H3K27me3) (category A) and non-
693 differentially bound sites (|Z-score| < 1 for H3K27ac, |Z-score| < 2 for H3K27me3) (category
694 B). Within each category, we further categorized the binding sites into two groups: up-
695 regulated bound sites (A1, B1) and down-regulated bound sites (A2, B2) in 22q11del cell
696 lines. Then we calculated the fold changes of the genes' FPKM between 22q11del and
697 control cell lines within each group. Permutation tests were performed with 9,999

698 permutations to obtain statistical significance.

699 To estimate the correlation between gene expression and histone modification in a
700 direct way, we calculated Pearson's correlation coefficient between gene's FPKM and
701 normalized read counts in the corresponding TSS binding site for each gene across all of the
702 cell lines. To obtain statistical significance, we first permuted genes' FPKM across the cell
703 lines for each TSS 10 times to assess the background correlation levels, and then performed
704 the Wilcoxon rank sum test between the observed correlation coefficients and the
705 background correlation coefficients. We also performed the same analysis using the
706 differentially expressed genes only and differentially enriched binding sites only.

707 **Correlation analysis for epigenetic marks.** To assess the correlation of epigenetic marks
708 binding between different regions on the same chromosomes, we divided the chromosomes
709 into 500 kbp bins. Within each bin, we calculated the mean value of the normalized read
710 counts for all the binding sites of each epigenetic mark. Then we calculated the Pearson's
711 correlation coefficients of obtained mean values across the cell lines between any two bins
712 on the same chromosomes.

713 **Data availability**. Hi-C, ChIP-Seq and RNA-Seq data from this study have been submitted to
714 the NCBI Gene Expression Omnibus (GEO; http://www.ncbi. nlm.nih.gov/geo/) under
715 accession number GSE76922.

716 **Competing interests**

717 The authors declare that no competing interests exist.

718 **Authors' contributions (in order of position in Authors List)**

719 YZ contributed to study design and produced experimental data, XZ contributed to study
720 design, carried out data analysis and contributed to writing the paper, XZ contributed to
721 study design and carried out data analysis, CP produced experimental data and contributed
722 to data interpretation, MSH produced experimental data, TW produced experimental data,
723 JY produced experimental data, SMW and AEU conceived of and designed the study,
724 directed data production and data analysis and wrote the paper.

725 **Acknowledgments**

726 This research was supported by external funding from the National Institute of Mental
727 Health (NIMH) (grant number: MH100010), NARSAD and the March of Dimes, as well as
728 funds from Stanford Pediatric Research Fund (Grant number: UL1 RR025744) and
729 departmental funds from the Stanford Department of Psychiatry and Behavioral Sciences.

730 We thank Dr. Eitan Yaffe (Stanford University) for his advice on the use of hicpipe.

731 **References**

732 1 Kirov, G. CNVs in neuropsychiatric disorders. *Hum Mol Genet* **24**, R45-49,
733 doi:10.1093/hmg/ddv253 (2015).
734 2 Torres, F., Barbosa, M. & Maciel, P. Recurrent copy number variations as risk
735 factors for neurodevelopmental disorders: critical overview and analysis of
736 clinical implications. *J Med Genet* **53**, 73-90, doi:10.1136/jmedgenet-2015-
737 103366 (2016).
738 3 Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**,
739 68-74, doi:10.1038/nature15393 (2015).

740 4 Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504
741 human genomes. *Nature* **526**, 75-81, doi:10.1038/nature15394 (2015).
742 5 Korbel, J. O. *et al.* Paired-end mapping reveals extensive structural variation
743 in the human genome. *Science* **318**, 420-426, doi:10.1126/science.1149504
744 (2007).
745 6 Abecasis, G. R. *et al.* A map of human genome variation from population-scale
746 sequencing. *Nature* **467**, 1061-1073, doi:10.1038/nature09534 (2010).
747 7 Mills, R. E. *et al.* Mapping copy number variation by population-scale genome
748 sequencing. *Nature* **470**, 59-65, doi:10.1038/nature09708 (2011).
749 8 Santos, J. L. *et al.* Copy number polymorphism of the salivary amylase gene:
750 implications in human nutrition research. *J Nutrigenet Nutrigenomics* **5**, 117-
751 131, doi:10.1159/000339951 (2012).
752 9 Gavin, D. P. & Floreani, C. Epigenetics of schizophrenia: an open and shut
753 case. *Int Rev Neurobiol* **115**, 155-201, doi:10.1016/B978-0-12-801311-
754 3.00005-6 (2014).
755 10 Fullard, J. F. *et al.* Understanding the genetic liability to schizophrenia
756 through the neuroepigenome. *Schizophr Res*,
757 doi:10.1016/j.schres.2016.01.039 (2016).
758 11 Nestler, E. J., Peña, C. J., Kundakovic, M., Mitchell, A. & Akbarian, S. Epigenetic
759 Basis of Mental Illness. *Neuroscientist*, doi:10.1177/1073858415608147
760 (2015).

761 12 Mitchell, A. C. *et al.* The genome in three dimensions: a new frontier in human
762 brain research. *Biol Psychiatry* **75**, 961-969,
763 doi:10.1016/j.biopsych.2013.07.015 (2014).
764 13 Barnard, R. A., Pomaville, M. B. & O'Roak, B. J. Mutations and Modeling of the
765 Chromatin Remodeler CHD8 Define an Emerging Autism Etiology. *Front*
766 *Neurosci* **9**, 477, doi:10.3389/fnins.2015.00477 (2015).
767 14 Vallianatos, C. N. & Iwase, S. Disrupted intricacy of histone H3K4 methylation
768 in neurodevelopmental disorders. *Epigenomics* **7**, 503-519,
769 doi:10.2217/epi.15.1 (2015).
770 15 Zhubi, A., Cook, E. H., Guidotti, A. & Grayson, D. R. Epigenetic mechanisms in
771 autism spectrum disorder. *Int Rev Neurobiol* **115**, 203-244,
772 doi:10.1016/B978-0-12-801311-3.00006-8 (2014).
773 16 Biswas, A. B. & Furniss, F. Cognitive phenotype and psychiatric disorder in
774 22q11.2 deletion syndrome: A review. *Res Dev Disabil* **53-54**, 242-257,
775 doi:10.1016/j.ridd.2016.02.010 (2016).
776 17 Swillen, A. & McDonald-McGinn, D. Developmental trajectories in 22q11.2
777 deletion. *Am J Med Genet C Semin Med Genet* **169**, 172-181,
778 doi:10.1002/ajmg.c.31435 (2015).
779 18 Schneider, M. *et al.* Psychiatric disorders from childhood to adulthood in
780 22q11.2 deletion syndrome: results from the International Consortium on
781 Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J Psychiatry* **171**, 627-
782 639, doi:10.1176/appi.ajp.2013.13070864 (2014).

783 19 Urban, A. E. *et al.* High-resolution mapping of DNA copy alterations in human
784 chromosome 22 using high-density tiling oligonucleotide arrays. *Proc Natl*
785 *Acad Sci U S A* **103**, 4534-4539, doi:10.1073/pnas.0511340103 (2006).
786 20 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range
787 interactions reveals folding principles of the human genome. *Science* **326**,
788 289-293, doi:10.1126/science.1181369 (2009).
789 21 Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates
790 systematic biases to characterize global chromosomal architecture. *Nat Genet*
791 **43**, 1059-1065, doi:10.1038/ng.947 (2011).
792 22 Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of
793 chromosome organization. *Nat Methods* **9**, 999-1003,
794 doi:10.1038/nmeth.2148 (2012).
795 23 Hu, M. *et al.* HiCNorm: removing biases in Hi-C data via Poisson regression.
796 *Bioinformatics* **28**, 3131-3133, doi:10.1093/bioinformatics/bts570 (2012).
797 24 Schmid, M. W., Grob, S. & Grossniklaus, U. HiCdat: a fast and easy-to-use Hi-C
798 data analysis tool. *BMC Bioinformatics* **16**, 277, doi:10.1186/s12859-015-
799 0678-x (2015).
800 25 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data
801 processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).
802 26 Heinz, S. *et al.* Simple combinations of lineage-determining transcription
803 factors prime cis-regulatory elements required for macrophage and B cell
804 identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

805 27 Li, W., Gong, K., Li, Q., Alber, F. & Zhou, X. J. Hi-Corrector: a fast, scalable and
806 memory-efficient package for normalizing large-scale Hi-C data.
807 *Bioinformatics* **31**, 960-962, doi:10.1093/bioinformatics/btu747 (2015).
808 28 Servant, N. *et al.* HiTC: exploration of high-throughput 'C' experiments.
809 *Bioinformatics* **28**, 2843-2844, doi:10.1093/bioinformatics/bts521 (2012).
810 29 Shavit, Y. & Lio', P. Combining a wavelet change point and the Bayes factor for
811 analysing chromosomal interaction data. *Mol Biosyst* **10**, 1576-1585,
812 doi:10.1039/c4mb00142g (2014).
813 30 Sauria, M. E., Phillips-Cremins, J. E., Corces, V. G. & Taylor, J. HiFive: a tool
814 suite for easy and efficient HiC and 5C data analysis. *Genome Biol* **16**, 237,
815 doi:10.1186/s13059-015-0806-y (2015).
816 31 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals
817 principles of chromatin looping. *Cell* **159**, 1665-1680,
818 doi:10.1016/j.cell.2014.11.021 (2014).
819 32 Rees, E. *et al.* Evidence that duplications of 22q11.2 protect against
820 schizophrenia. *Mol Psychiatry* **19**, 37-40, doi:10.1038/mp.2013.156 (2014).
821 33 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by
822 analysis of chromatin interactions. *Nature* **485**, 376-380,
823 doi:10.1038/nature11082 (2012).
824 34 Maynard, T. M. *et al.* Mitochondrial localization and function of a subset of
825 22q11 deletion syndrome candidate genes. *Mol Cell Neurosci* **39**, 439-451,
826 doi:10.1016/j.mcn.2008.07.027 (2008).

827 35 Butcher, N. J. *et al.* Association between early-onset Parkinson disease and
828 22q11.2 deletion syndrome: identification of a novel genetic form of
829 Parkinson disease and its clinical implications. *JAMA Neurol* **70**, 1359-1366,
830 doi:10.1001/jamaneurol.2013.3646 (2013).
831 36 Boot, E. *et al.* Movement disorders and other motor abnormalities in adults
832 with 22q11.2 deletion syndrome. *Am J Med Genet A* **167A**, 639-645,
833 doi:10.1002/ajmg.a.36928 (2015).
834 37 Mok, K. Y. *et al.* Deletions at 22q11.2 in idiopathic Parkinson's disease: a
835 combined analysis of genome-wide association data. *Lancet Neurol* **15**, 585-
836 596, doi:10.1016/S1474-4422(16)00071-5 (2016).
837 38 Stefansson, H. *et al.* Large recurrent microdeletions associated with
838 schizophrenia. *Nature* **455**, 232-236, doi:10.1038/nature07229 (2008).
839 39 Consortium, I. S. Rare chromosomal deletions and duplications increase risk
840 of schizophrenia. *Nature* **455**, 237-241, doi:10.1038/nature07239 (2008).
841 40 Levinson, D. F. *et al.* Copy number variants in schizophrenia: confirmation of
842 five previous findings and new evidence for 3q29 microdeletions and VIPR2
843 duplications. *Am J Psychiatry* **168**, 302-316,
844 doi:10.1176/appi.ajp.2010.10060876 (2011).
845 41 Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause
846 pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025,
847 doi:10.1016/j.cell.2015.04.004 (2015).

848 42 Taberlay, P. C. *et al.* Three-dimensional disorganization of the cancer genome
849 occurs coincident with long-range genetic and epigenetic alterations. *Genome*
850 *Res* **26**, 719-731, doi:10.1101/gr.201517.115 (2016).
851 43 Loviglio, M. N. *et al.* Chromosomal contacts connect loci associated with
852 autism, BMI and head circumference phenotypes. *Mol Psychiatry*,
853 doi:10.1038/mp.2016.84 (2016).
854 44 Kasowski, M. *et al.* Extensive variation in chromatin states across humans.
855 *Science* **342**, 750-752, doi:10.1126/science.1242510 (2013).
856 45 Raha, D., Hong, M. & Snyder, M. ChIP-Seq: a method for global identification
857 of regulatory elements in the genome. *Curr Protoc Mol Biol* **Chapter 21**, Unit
858 21.19.21-14, doi:10.1002/0471142727.mb2119s91 (2010).
859 46 Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by
860 induced pluripotent stem cells. *Nature* **492**, 438-442,
861 doi:10.1038/nature11629 (2012).
862 47 Yao, J., Fetter, R. D., Hu, P., Betzig, E. & Tjian, R. Subnuclear segregation of
863 genes and core promoter factors in myogenesis. *Genes Dev* **25**, 569-580,
864 doi:10.1101/gad.2021411 (2011).
865 48 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence
866 of insertions, deletions and gene fusions. *Genome Biol* **14**, R36,
867 doi:10.1186/gb-2013-14-4-r36 (2013).
868 49 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals
869 unannotated transcripts and isoform switching during cell differentiation.
870 *Nat Biotechnol* **28**, 511-515, doi:10.1038/nbt.1621 (2010).

871 50 Trapnell, C. *et al.* Differential analysis of gene regulation at transcript
872 resolution with RNA-seq. *Nat Biotechnol* **31**, 46-53, doi:10.1038/nbt.2450
873 (2013).
874 51 Huang, d. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment
875 tools: paths toward the comprehensive functional analysis of large gene lists.
876 *Nucleic Acids Res* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).
877 52 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-
878 Wheeler transform. *Bioinformatics* **25**, 1754-1760,
879 doi:10.1093/bioinformatics/btp324 (2009).
880 53 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and
881 modENCODE consortia. *Genome Res* **22**, 1813-1831,
882 doi:10.1101/gr.136184.111 (2012).
883 54 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**,
884 R137, doi:10.1186/gb-2008-9-9-r137 (2008).
885 55 Stark, R. & Brown, G.
886 http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/do
887 c/DiffBind.pdf, 2011.
888 56 Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq
889 and ChIP-chip data. *BMC Bioinformatics* **11**, 237, doi:10.1186/1471-2105-11-
890 237 (2010).
891