*Year :* 2012

# COPY NUMBER VARIATION AND CHROMATIN STRUCTURE

## Robert WITWICKI

**UNIL | Université de Lausanne**

Faculté de biologie
et de médecine

**Centre intégratif de génomique**

**COPY NUMBER VARIATION AND CHROMATIN STRUCTURE**

**Thèse de doctorat ès sciences de la vie (PhD)**

présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne

par

# Robert WITWICKI

Master de l'Université de Wroclaw, Pologne

**Jury**

Prof. Jacques Beckmann, Président
Prof. Alexandre Reymond, Directeur de thèse
Dr Pierre Colas, expert
Dr Phillip Shaw, expert

Lausanne  2012

# Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

| | | |
|---|---|---|
| *Président* | Monsieur | Prof. Jacques **Beckmann** |
| *Directeur de thèse* | Monsieur | Prof. Alexandre **Reymond** |
| *Experts* | Monsieur | Dr Pierre **Colas** |
| | Monsieur | Dr Phillip **Shaw** |

le Conseil de Faculté autorise l'impression de la thèse de

# Monsieur Robert Witwicki

Master en biologie de l'Université de Wroclaw, Pologne

intitulée

# COPY NUMBER VARIATION
# AND CHROMATIN STRUCTURE

Lausanne, le 6 juillet 2012

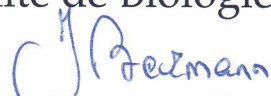pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Jacques Beckmann

# Table of content

# Summary

The functional consequences of structural variation in the human genome range from adaptation, to phenotypic variation, to predisposition to diseases. Copy number variation (CNV) was shown to influence the phenotype by modifying, in a somewhat dose-dependent manner, the expression of genes that map within them, as well as that of genes located on their flanks. To assess the possible mechanism(s) behind this neighboring effect, we compared histone modification status of cell lines from patients affected by Williams-Beuren, Williams-Beuren region duplication, Smith-Magenis or DiGeorge Syndrome and control individuals using a high-throughput version of chromatin immuno-precipitation method (ChIP), called ChIP-seq. We monitored monomethylation of lysine K20 on histone H4 and trimethylation of lysine K27 on histone H3, as proxies for open and condensed chromatin, respectively.

Consistent with the changes in expression levels observed for multiple genes mapping on the entire length of chromosomes affected by structural variants, we also detected regions with modified histone status between samples, up- and downstream from the critical regions, up to the end of the rearranged chromosome.

We also gauged the intrachromosomal interactions of these cell lines utilizing chromosome conformation capture (4C-seq) technique. We observed that a set of genes flanking the Williams-Beuren Syndrome critical region (WBSCR) were often looping together, possibly forming an interacting cluster with each other and the WBSCR. Deletion of the WBSCR disrupts the expression of this group of flanking genes, as well as long-range interactions between them and the rearranged interval.

We conclude, that large genomic rearrangements can lead to changes in the state of the chromatin spreading far away from the critical region, thus

possibly affecting expression globally and as a result modifying the phenotype of the patients.

# Résumé

Les conséquences fonctionnelles des variations structurelles dans le génome humain sont vastes, allant de l'adaptation, en passant par les variations phénotypiques, aux prédispositions à certaines maladies. Il a été démontré que les variations du nombre de copies (CNV) influencent le phénotype en modifiant, d'une manière plus ou moins dose-dépendante, l'expression des gènes se situant à l'intérieur de ces régions, mais également celle des gènes se trouvant dans les régions flanquantes. Afin d'étudier les mécanismes possibles sous-jacents à cet effet de voisinage, nous avons comparé les états de modification des histones dans des lignées cellulaires dérivées de patients atteints du syndrome de Williams-Beuren, de la duplication de la région Williams-Beuren, du syndrome de Smith-Magenis ou du syndrome de Di-George et d'individus contrôles en utilisant une version haut-débit de la méthode d'immunoprécipitation de la chromatine (ChIP), appelée ChIP-seq. Nous avons suivi la mono-méthylation de la lysine K20 sur l'histone H4 et la tri-méthylation de la lysine K27 sur l'histone H3, marqueurs respectifs de la chromatine ouverte et fermée.

En accord avec les changements de niveaux d'expression observés pour de multiples gènes tout le long des chromosomes affectés par les CNVs, nous avons aussi détecté des régions présentant des modifications d'histones entre les échantillons, situées de part et d'autre des régions critiques, jusqu'aux extrémités du chromosome réarrangé.

Nous avons aussi évalué les interactions intra-chromosomiques ayant lieu dans ces cellules par l'utilisation de la technique de capture de conformation des chromosomes (4C-seq). Nous avons observé qu'un groupe de gènes flanquants la région critique du syndrome de Williams-Beuren (WBSCR) forment souvent une boucle, constituant un groupe d'interactions privilégiées entre ces gènes et la WBSCR. La délétion de la WBSCR perturbe l'expression de ce groupe de gènes flanquants, mais également les interactions à grande échelle entre eux et la région réarrangée.

Nous en concluons que les larges réarrangements génomiques peuvent aboutir à des changements de l'état de la chromatine pouvant s'étendre bien plus loin que la région critique, affectant donc potentiellement l'expression de manière globale et ainsi modifiant le phénotype des patients.

# Background

Genomic variation is a prevalent phenomenon in the human genome[1]. These variations can be of a different scale and range from single nucleotides (Single Nucleotide Polymorphisms – SNPs) to megabases in length. Single nucleotide polymorphisms are the most frequent and the frequency drops with the size of the polymorphic region[2]. Copy number variants (CNVs), so deletions or duplications of a larger genomic region, are responsible for as much as 12% genomic differences between individuals[3]. CNVs are relatively big – 65%-80% of individuals in human population carry a variant larger than 100kb and 5%-10% a variant larger than 500kb[4].

Such polymorphisms can contribute to differences in phenotypes between individuals, susceptibility to disease or be pathogenic themselves[5]. It is estimated that 14% of mental retardation cases on genetic background is caused by a CNV[6].

In this study we focus on four recurrent CNVs, namely Williams-Beuren Syndrome (WBS; OMIM#194050) and it's reciprocal duplication (WBRdupS; OMIM#609757), Smith-Magenis Syndrome (SMS; OMIM#182290) and DiGeorge Syndrome (DGS; OMIM#188400).

Williams-Beuren Syndrome is caused by a deletion on the long arm of chromosome 7 (7q11.23). Occurrence rate of this syndrome is estimated at 1:10000[7]. The deletion is around 1.5-1.8Mb long and encompasses 26-28 genes[7]. The phenotype of WBS patient is characterized i.a. by elfin face, cardiovascular abnormalities, mental retardation[7]. Surprisingly, patients are highly social and develop good language skills when compared to IQ-matched individuals with Down Syndrome[8].

Because of its lower occurrence (1:13000-1:20000)[9], Williams-Beuren region duplication syndrome (WBRdupS) clinical characteristic is not well defined. Some phenotypes seem to mirror the ones observed in WBS patients (like autism-like behavior or poor language development) while other are

similar in both syndromes (e.g. mental retardation or cardiovascular abnormalities)[9].

Smith-Magenis Syndrome is caused by a deletion on the short arm of chromosome 17 (17p11.2)[10]. The typical size of a deleted fragment is 3.7Mb and estimated prevalence is 1:15000-1:25000. The clinical characteristic includes e.g. craniofacial anomalies, mental retardation, sleep disturbance, anxiety and aggression[10, 11].

DiGeorge Syndrome is the most frequent CNV syndrome in human, with occurrence rate estimated at 1:4000[5, 12]. The deletion involves 35 genes on chromosome 22 (22q11.2) and typically is 3Mb long. The phenotypic features involve, among others, cardiac anomalies, immunodeficiency and mental retardation[12].

It was previously shown, that CNVs could cause changes in the expression level of the genes located not only inside, but also outside the rearrangement[13-17], what could explain the complex phenotypes of the patients. In this study we wanted to investigate one hypothetical mechanism that could be responsible for that phenomenon, that is changes in chromatin structure caused by a deletion or duplication of a fragment of a chromosome.

Chromatin is an organized structure of DNA and nucleosomes (Figure 1). Nucleosome cores are protein complexes build from histone proteins. Each nucleosome core contains 4 homo-dimers, with dimers consisting of each of the main histones – H2A, H2B, H3 and H4. Additionally, nucleosome cores are bound together by linker histones – H1 and H5 – and that way can form higher order structures, allowing even further compaction[18].
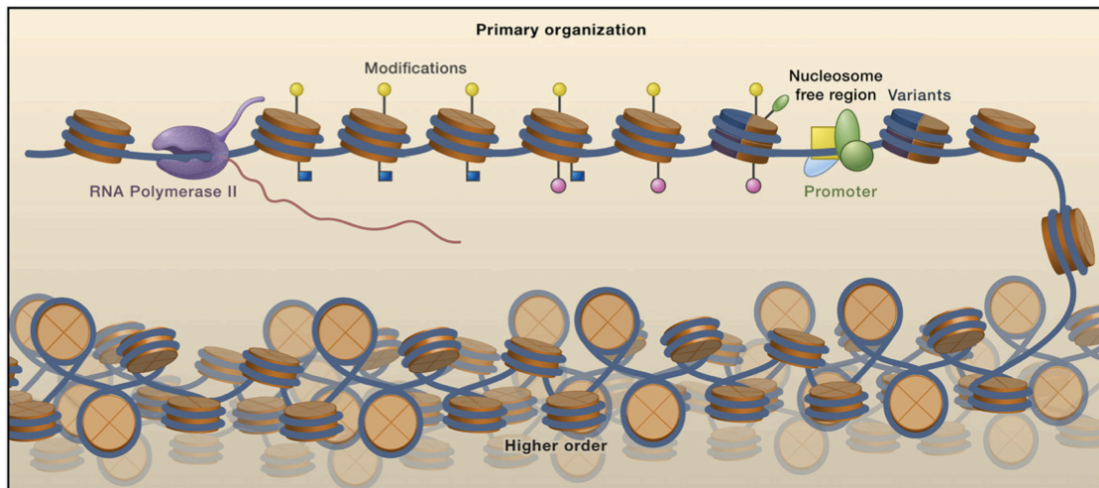
**Figure 1.** Structure of the chromatin. DNA is wrapped around nucleosomes, which then can be further condensed forming higher order chromatin organization[19].

In addition to structural function, histones can also play a role in chromatin regulation. Posttranslational modifications of long tails of histones H3 and H4 were seen to change properties of the chromatin. These modifications can be of many different kinds and include, among others, methylation, acetylation, phosphorylation, ADP-ribosylation and ubiquitination (Figure 2). The combination of these marks is sometimes referred to as a histone code[20].
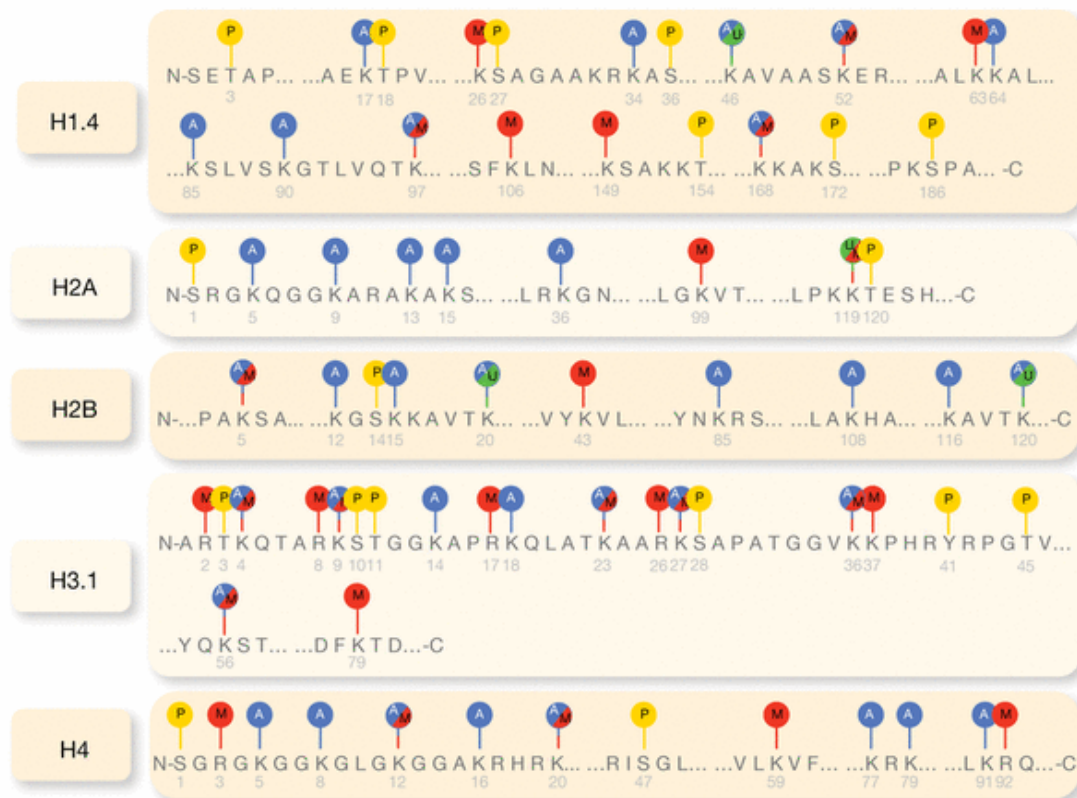
**Figure 2.** The most frequently studied histone modifications include phosphorylation (depicted as a circle with the letter P), acetylation (A), methylation (M) and ubiquitination (U)[21].

The histone code is a phenomenon that is not yet fully understood. But with recent studies screening the whole panels of histone modifications, we can draw some conclusions based on patterns common for many cell or tissue types[22, 23]. Some of the marks seem to be persistently linked to the regions of chromatin containing actively transcribed genes (for example trimethylation of 3rd lysine on histone H3 – H3K4me3 or monomethylation of 20th lysine on histone H4 – H4K20me1), while others are associated with silenced or condensed chromatin (like trimethylation of 27th lysine on histone H3 – H3K27me3)[24].

For the purpose of this study we have chosen H4K20me1 as a mark of open, and H3K27me3 as a mark of condensed chromatin and used chromatin immunoprecipitation technique coupled with high-throughput sequencing (ChIP-seq). The goal of chromatin immunoprecipitation is to enrich the pool of chromatin for the DNA bound with the protein of interest, which is precipitated

using specific antibody. Briefly, proteins are cross-linked with the DNA region, to which they are binding *in vivo,* by adding cross-linking agent directly to the tissue culture medium. Following steps are cell lysis using detergents and mechanical shearing of the DNA. DNA-protein complexes are then immunoprecipitated. Next, the complexes are dissociated by reverse cross-linking and DNA is precipitated from the solution. Purified DNA can be then used for creating a library for high-throughput sequencing and sequenced (Figure 3).
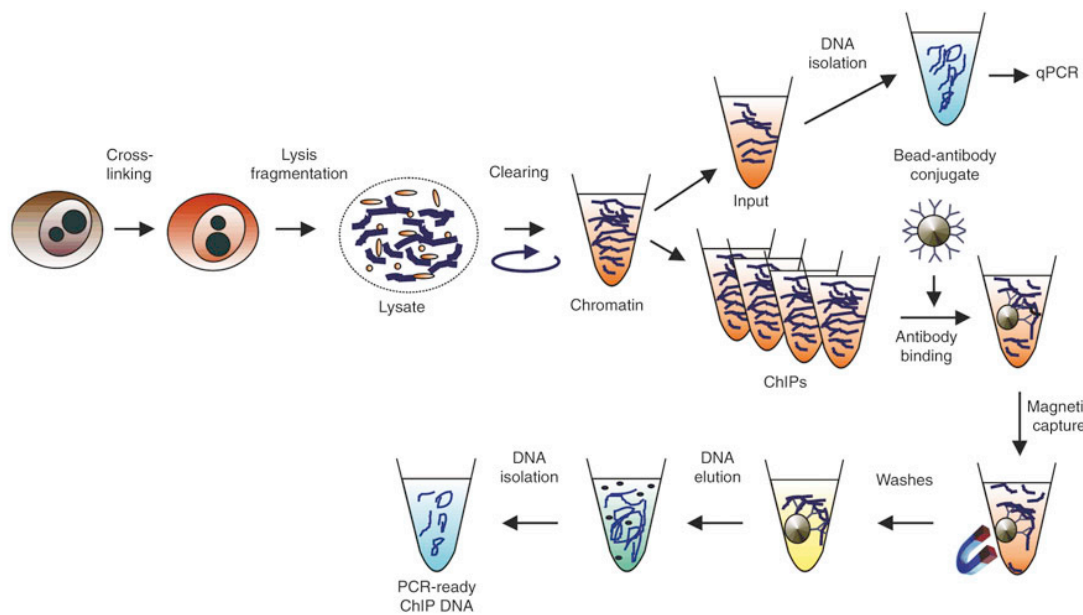


**Figure 3.** The workflow of chromatin immunoprecipitation (ChIP) procedure [25].

The results of ChIP-seq experiments that I performed are summarized in the publication cited in the next chapter of this thesis.

# Structural variation-induced expression changes are paralleled by chromatin architecture modifications

Nele Gheldof[1,*,$], Robert M. Witwicki[1,*], Eugenia Migliavacca[1,2,*], Marion Leleu[2,3], Gérard Didelot[1], Louise Harewood[1,4], Jacques Rougemont[2,3], Alexandre Reymond[1,$]

[1]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
[2]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland
[3]School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[4]present address: Laboratory of Chromatin and Gene Expression, Babraham Research Campus, Cambridge, UK.

*These authors contributed equally to this work

$Correspondence should be addressed to
Nele Gheldof, nele.gheldof@unil.ch
Or
Alexandre Reymond, alexandre.reymond@unil.ch
Center for Integrative Genomics, University of Lausanne, Genopode building, 1015 Lausanne, Switzerland, +41 21 692 3960 (phone), +41 21 692 3965 (fax).

*Abstract*

Copy number variants (CNVs) influence the expression of genes that map not only within the rearrangement, but also to its flanks. To assess the possible mechanism(s) underlying this "neighboring effect", we compared intrachromosomal interactions and histone modifications in cell lines of patients affected by genomic disorders and control individuals. Using chromosome conformation capture (4C-seq), we observed that a set of genes flanking the Williams-Beuren Syndrome critical region (WBSCR) were often looping together, possibly forming an interacting cluster with each other and the WBSCR. The newly identified interacting genes include *AUTS2*, mutations of which are associated with autism and intellectual disabilities. Deletion of the WBSCR disrupts the expression of this group of flanking genes, as well as long-range interactions between them and the rearranged interval. We also pinpointed concomitant changes in histone modifications between samples.

We conclude that large genomic rearrangements can lead to chromatin conformation changes that extend far away from the structural variant, thereby possibly modulating expression globally and modifying the phenotype.

*Introduction*

Copy number variation (CNV) of genomic segments is frequent in human [3] and model organisms (e.g. mouse [15, 16, 26-28]). More than 66,000 human CNVs mapping to 16,000 regions have so far been identified (http://projects.tcag.ca/variation/). They significantly contribute to genetic variation, covering more nucleotide content per genome than single nucleotide polymorphisms (e.g. approximately 0.8% of the length of the human genome differs between two human individuals [29]). Multiple associations between these structural changes and susceptibility to disease have been uncovered (reviewed in [30-34]). One of these is Williams-Beuren syndrome, a multi-system disorder caused by a recurrent megabase-scale segmental deletion (WBS, MIM ID #194050, [7]).

CNVs impact tissue transcriptomes on a global scale by modifying the level and timing of expression of genes that localize within the CNV [35, 36] and on its flanks [13-17], an effect that can extend over the entire length of the affected chromosome [37]. Structural changes *per se*, i.e. without changes in gene dosage were shown to profoundly impact the phenotypic outcome, as some phenotypic traits present in Smith-Magenis (deletion) and Potocki-Lupski syndromes (reciprocal duplication) mouse models were not rescued by restoration of the copy number in a strain carrying both the deletion and duplication on different alleles [37]. The mechanism(s) behind this chromosome-wide effect is(are) currently unknown. One hypothesis is that some of the genes that map within a rearrangement, and thus vary in dosage, directly or indirectly affect the expression of normal dosage flanking genes. However, as in multiple instances we found the flanking genes to be altered independent of CNV dosage (i.e. both the deletion of a given CNV and its reciprocal duplication upregulate the expression of a flanking gene)[37, 38], it is unlikely that this hypothesis constitute the only mechanism behind this "neighboring effect". Other mechanisms may include position effect (i.e. physical dissociation of a transcription unit from its *cis*-acting regulators [39]),

17

alteration of chromatin structure locally or globally [40], and/or repositioning of a genomic region within the nucleus [41].

As chromatin structure plays an important role in gene regulation, we anticipate that CNVs will affect the chromatin structure on a large scale, and hence possibly modify the clinical phenotype. However, studies investigating the impact of a structural aberration on long-range chromatin structure have been lacking. Here, we explored the chromosome-wide effect of a set of structural rearrangements on chromatin structure. First, we studied, by chromosome conformation capture, whether non-hemizygous genes neighboring a rearrangement and known to be affected in their expression levels also show changes in chromatin structure. Second, we monitored local chromatin changes as determined by histone modifications in multiple cell lines with structural rearrangements.

## *Materials and Methods*

### *Cell lines*
SMS (GM18319), WBS (GM13472) and Control (Ctrl, GM07006) lymphoblastoid cell lines were obtained from Coriell Institute for Medical Research (http://www.coriell.org/). WBRdupS (AUBLA) and DGS (SE160) lymphoblastoid cell lines were established by transfecting peripheral blood mononuclear cells with EBV. These female patients were enrolled after obtaining appropriate informed consent by the physicians in charge and approval by the ethics committee of the University of Lausanne. Cells were grown in RPMI 1640 medium (Gibco) with addition of 10% fetal calf serum and 1% penicillin-streptomycin. Rearrangements were examined by array CGH using Human CGH 3x720K whole-genome tiling array (Nimblegen) following the manufacturer's protocol. Known changes in the expression levels of *GBAS, ASL, KCTD7, HIP1, POR* and *MDH2* in WBS patient cell lines were confirmed in GM13472 relative to the Ctrl cell line cells by Taqman real-time quantitative PCR using previously published primers pairs and probes [13].

*Circularized Chromosome Conformation Capture – sequencing (4C-seq)*

The 4C-seq assay was performed as described in [42] and based on 4C protocol developed by [43, 44]. Briefly, GM07006 (Ctrl) and GM13472 (WBS) lymphoblastoid cell lines were grown at 37°C. $5 \times 10^7$ exponentially growing cells were harvested and crosslinked with 1% formaldehyde, lysed and cut with the restriction enzyme *Bgl*II. 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 make the 4C-seq templates were designed to contain the Illumina adaptor tails, as well as the bait-specific sequences for each of the six loci we interrogated. The list of primers is described in **Supplementary Table S1**. The six viewpoints were selected at the *Bgl*II fragment containing the transcriptional start sites of three genes located upstream of the WBSCR (*GBAS* 16.7 Mb, *ASL* 7.6 Mb, and *KCTD7* 7 Mb upstream respectively), and three other genes located immediately downstream of the WBSCR (*HIP1* 0.7 Mb, *POR* 0.96 Mb and *MDH2* 1 Mb downstream respectively). For the nearby downstream viewpoints, we amplified at least 0.6 µg of 4C template, whereas for the further away upstream viewpoints, we amplified at least 1 µg of 4C template (using about 100 ng per inverse PCR reaction). We multiplexed the 4C-seq templates by pooling the samples in equimolar ratios in two sets, representing 3 viewpoints each (*POR*, *KCTD7* and *GBAS* in one set and *ASL*, *MDH2* and *HIP1* in the second set). Replicate 4C libraries were prepared for both the Ctrl and the WBS cell lines. We randomly selected three of the six viewpoints (*ASL*, *POR* and *MDH2*) and replicated the experiments. These assays were pooled and assessed by sequencing in multiplex. All 4C-seq multiplexed samples were analyzed on a Illumina GAIIx flow-cell (Illumina) using a 76-bp single-end sequencing run.

*4C-seq data analysis*

4C-seq data were analyzed as described in [42]. Briefly, the multiplexed samples were separated, undigested self-ligated reads removed, and the reads mapped to a virtual library of *Bgl*II fragments. Reads were then

normalized to the total number of reads. A running mean algorithm was applied to smooth the data (19 fragments per window). As the data from the three replicated viewpoints were strongly correlated (**Supplementary Figure S1**), we used the average of each data point for these experiments. To remove the strongly interacting local "background" region, we modeled the data to apply a profile correction similar to the one described in [45] using a fit with a slope -1 in a log-log scale [46]. We used a domainogram algorithm to detect significantly interacting regions without imposing a fixed window size as suggested [47]. The positive signals were ranked per chromosome and Bricks (Blocks of Regulators In Chromosomal Kontext) were called based on a FDR threshold of 0.1 for "short-range" interactions, defined as interactions within 2.5 Mb up- and downstream of *GBAS* and *MDH2,* the first and last viewpoint, respectively (HSA7 coordinates: 53,532,296-78,116,172; about 25 Mb around the WBSCR). Genomic space outside of these borders was called the "long-range" region for which we used a more stringent FDR threshold of 0.001. To determine differentially interacting regions between the WBS and Ctrl cells, we first computed the log2 ratio of WBS over Ctrl of the smoothed profile corrected data and selected ratio Bricks (as described above) that were specific to either WBS or Ctrl. 4C data are deposited under accession number GSE33867.

To estimate if the long-range interactions of all six viewpoints were significantly enriched in genes or histone modifications we performed permutation tests (n=10000) with all RefSeq genes or histone modified regions identified by SICER with a FDR = $1 \times 10^{-4}$. To permute the interacting regions we used shuffleBed from BEDtools version 2.10.1 [48].

*Chromatin Immunoprecipitation - sequencing (ChIP-seq)*
Crosslinking was performed by adding formaldehyde (Sigma Aldrich) to the cells in growth medium to a final concentration of 0.5%. After 5 minutes incubation at room temperature, cross-linking agent was quenched with 0.125M glycine. $1 \times 10^6$ cells were centrifuged and used directly in the ChIP assay. Cells were lysed by addition of cell lysis buffer (1% SDS, EDTA, Tris-HCl pH 8.1) and a 10 minute incubation on ice. Next, chromatin was sheared

with a Bioruptor (Diagenode) at medium power settings (30 seconds on – 30 seconds off cycles for 45'). Sonication efficiency was tested by reversing cross-links of a chromatin sample and running the obtained DNA on a 1.5% agarose gel. Fragmented chromatin was used directly in the ChIP assay or frozen at -80$^{o}$C for latter usage.

ChIP was performed as suggested in [49]. Briefly, chromatin was diluted 10 fold in ChIP dilution buffer (0.01% SDS, 1.1% Triton X100, 1.2mM EDTA, 16.7mM Tris-HCl pH 8.1, 167mM NaCl) and subsequently immunoprecipitated using antibodies raised against H3K27me3 (Millipore 07-449) and H4K20me1 (Abcam ab9051). The antibody-histone complex was collected using magnetic beads (Invitrogen). Beads were washed twice with dialysis buffer (2mM EDTA, 50mM Tris-HCl pH 8.0, 0.2% sarcosyl) and four times with wash buffer (100mM Tris-HCl pH 9.0, 500mM LiCl, 1% NP40, 1% sodium deoxycholate). The DNA was then eluted and the crosslinks reversed. Following RNase A and proteinase K treatments, samples were purified using the DNA purification kit (Qiagen). The concentration was measured by Qubit (Invitrogen) and 10 ng of each sample was used for library preparation. Enrichment of the precipitated DNA was assessed by comparing the levels of DNA corresponding to known open and closed chromatin regions by quantitative PCR. Primer pairs corresponding to exon 2 of *GAPDH* and intron 5 of the *GRM8* gene were used for the H4K20me1 and H3K27me3 ChIP, respectively. The same primer pairs were used reciprocally as negative controls.

Sequencing libraries of immunoprecipitated DNA samples were prepared as described by the manufacturer (Illumina). They were sequenced on two lanes of an Illumina GAIIx flow-cell each with 36mer tags. Sequencing reads were mapped to the human reference genome (hg19, GRCh37) using Bowtie algorithm allowing 2 mismatches and no seed [50]. Duplicates potentially arisen were removed, i.e. only a single tag was retained from identical sequences [51]. Note that in the remaining analyses, we only considered uniquely matching tags, i.e. between 21.7 and 32.1 x 10$^6$ and 3.3 and 15.3 x 10$^6$ for H4K20me1 and H3K27me3, respectively.

The identification of ChIP-enriched regions was performed using SICER [52] version 1.1 with two libraries (SICER-df-rb.sh) and the following parameters: window size 200 bp, gap size 400bp, for H4K20me1 and gap size 600 bp for H3K27me3 as suggested by the package authors, and E-value 100. We selected candidate islands with a FDR=$1 \times 10^{-4}$ defined by SICER for the Ctrl and the rearranged sample and further used these islands to assess statistical significance of differential modification of a given region using the DEseq package [53]. For the DEseq analyses, each sample with a chromosomal rearrangement (e.g. SMS with a deletion on HSA17) was compared to the Ctrl and the other samples with rearrangements on other chromosomes (e.g. DGS, WBS and WBRdupS), thus providing a set of three or four "control" samples and reducing differences due to individual variability. To identify genes that were significantly altered in their chromatin status - and thus possibly also in expression - we screened the chromatin changes of RefSeq genes defined according to the genomic coordinates [54]. ChIP-seq data are deposited under accession number GSE33784.

## Results

*Outlining the chromatin architecture of the WBS region*
We have previously shown that *GBAS, ASL, KCTD7, HIP1, POR* and *MDH2* (normal-copy number genes that map to the flank of the 7q11.23 deletion that causes WBS) are modified in their relative expression levels in lymphoblastoid and/or skin fibroblast cell lines of WBS patients [13]. We replicated these experiments in a new set of cell lines (**Table 1**). To assess if these changes are associated with changes in chromatin conformation, we first examined the chromatin interaction landscape of these six flanking genes in lymphoblastoid cells from a female control (Ctrl) using an adaptation of the 4C method (4C-seq: Circularized Chromosome Conformation Capture combined with multiplexed high-throughput sequencing; see methods). This technology allows identification of chromosomal regions that physically associate with a given locus, termed the bait or viewpoint.

**Figure 1A** shows the windowed interaction profiles for each viewpoint along the entire human chromosome 7 (HSA7). As anticipated, the strongest signals are observed close to the viewpoints. We computed domainograms [47, 55] and used both a stringent and a relaxed false discovery rate to detect "long-" and "short-range" interactions (within a 25 Mb region encompassing the WBS deletion), respectively (methods). Results are highly reproducible (0.83 ≤ Pearson's correlation ≤ 0.97; **Supplementary Figure S1**). We identified between 66 and 147 interacting regions called Bricks (Blocks of Regulators In Chromosomal Kontext)[47] on HSA7 for the six tested viewpoints. A large fraction of the interacting regions are shared between multiple viewpoints (**Figure 1B**). For example, 23% (28/121) of the regions found to interact with *POR* also interact with the *ASL, KCTD7, HIP1* and *MDH2* viewpoints (**Supplementary Figure S2**). The *GBAS* gene, which maps to the short arm of HSA7, interacts more frequently with regions on the same arm (**Figure 1A-B**), consistent with previous findings that showed that chromatin loops are at least partially guided by chromosome architecture [45]. Nevertheless, that short arm viewpoint shares an important fraction of interacting regions with the HSA7 long arm viewpoints (e.g. 58 and 49% (85/147 and 72/147) of regions shared with *POR* and *ASL*, respectively (**Supplementary Figure S2**). The long-range interactions of all six viewpoints are significantly enriched in gene-dense regions ($P = 9.5 \times 10^{-2}$ for *GBAS*, $P = 4.1 \times 10^{-3}$ for *KCTD7* and $P = 1 \times 10^{-4}$ for the other four long arm viewpoints, permutation test with number of permutations N=10000), consistent with previous studies that showed clustering of gene-rich regions [56, 57]. We actually found a positive correlation between the number of viewpoints with which a region interacts and the gene density of that particular region: regions interacting with all, five (excluding *GBAS*), two or only a single viewpoint(s) have a gene density of $4.8 \times 10^{-2}$, $4.1 \times 10^{-2}$, $1.7 \times 10^{-2}$, $0.3 \times 10^{-2}$ RefSeq genes/kilobase, respectively.
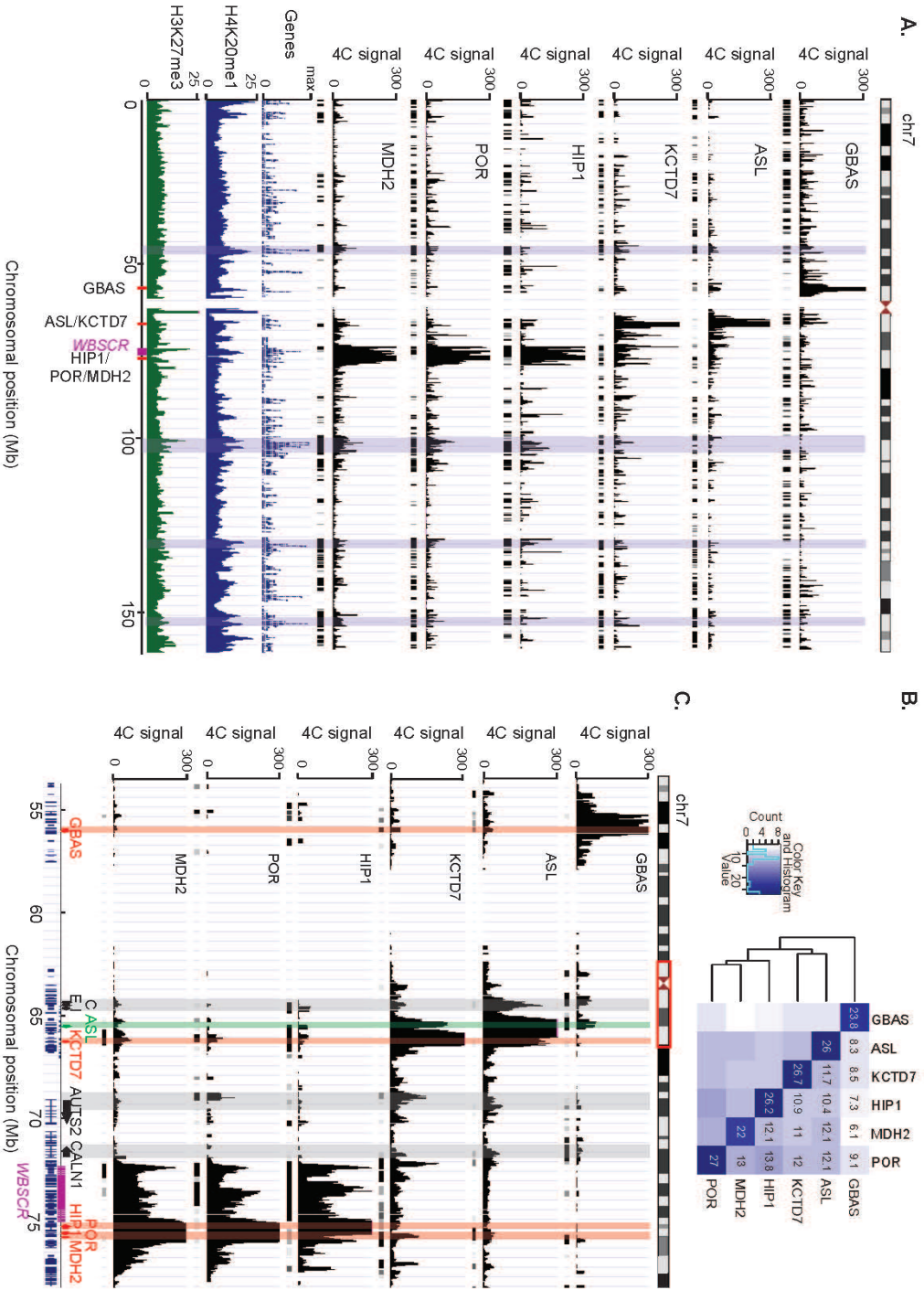
Figure 1

**Figure 1**. Extensive chromatin interactions of six genes on human chromosome 7 in cells from a healthy control individual. (**A**) Windowed 4C signal of each of the six viewpoints along the entire chromosome 7 (HSA7). The black ticks below each graph show the location of the Bricks (Blocks of Regulators In Chromosomal Kontext). The gene density across HSA7, as well as the windowed profiles of H4K20me1 and H3K27me3 marks in the same cell line are shown below. Some examples of strong correlation of gene-dense regions and high-density of H4K20me1 marks with highly interacting regions are highlighted in blue. The mapping of the assessed genes/viewpoints and of the WBSCR (WBS critical region) is indicated at the bottom. (**B**) Heatmap showing the percent coverage of HSA7 by Bricks of each viewpoint, as well as the percent of HSA7 that overlaps between Bricks of the different viewpoints, indicating that the viewpoint interactions cluster by their linear chromosomal position. (**C**) Close-up of the windowed 4C signal of the six viewpoints around the WBSCR for the region indicated with a red box on HSA7 (top panel). The position of all genes are displayed at the bottom, and the mapping of the assessed viewpoints is highlighted by red and green arrows indicating if the corresponding genes are down- or upregulated in cells from WBS patients, respectively. Black arrows underscore the mapping of newly identified interacting partners such as *AUTS2*, *CALN1, ERV3* (indicated by "E"), *CCT6P3* (C) and *INTS4L1* (I) genes (see text for details). The location of the WBSCR is indicated by a purple horizontal bar. A close-up of interactions within this WBSCR is provided in **Supplementary Figure S3**.

A close up of the interaction profiles of the six viewpoints around the WBS critical region (WBSCR) is depicted in **Figure 1C**. For the three genes immediately downstream of the WBS deletion (*HIP1*, *POR* and *MDH2*), we observed higher interactions with the entire WBS deletion region when compared to the region telomeric to these viewpoints. This could in part be due to spatial clustering of active gene-dense regions [43, 46] as the WBSCR contains more genes than the equidistant downstream flanking region. Even though extensive interactions were seen with the entire critical region, these three genes appear to interact primarily with the region that includes the elastin (*ELN*), *LIMK1*, *EIF4H* and *CLIP2* genes (**Supplementary Figure S3**). We also found interactions with the centromeric low-copy repeat (LCR) region, but we cannot exclude that this merely reflects its high similarity with the nearby telomeric LCR. Alternatively, as the *HIP1, POR* and *MDH2* viewpoints are immediately adjacent to the telomeric LCR, this interaction loop might be a chromatin loop caused by the mispairing of these two repetitive and highly homologous sequences. Existence of such loop was postulated to facilitate excision and thus deletion of the intervening sequence causing WBS [58]. The expression-modified centromeric genes, *ASL* and *KCTD7*, also interact with the WBSCR albeit not as preeminently as for the tested telomeric viewpoints (**Figure 1C**), maybe reflecting the fact that the centromeric genes are mapping at a greater distance from the WBS interval. The *GBAS* gene, on the other hand, located 17 Mb away from the WBSCR and on the other arm of chromosome 7, does not directly interact with the WBSCR.

We next examined whether the six viewpoints interact significantly with loci outside the WBSCR and LCRs. Interestingly, we found significant interactions between the expression-modified genes themselves (**Figure 1C**). Many of these interactions and their relative intensities are reciprocal (i.e. the same architecture with the same intensity is identified using two different starting viewpoints)(**Figure 2A)**, strengthening the legitimacy of the uncovered chromatin folding. Besides looping between the expression-modified genes, we also uncovered some other interacting partners shared between telomeric and centromeric viewpoints, such as the region around *ERV3*, *CCT6P3* and *INTS4L1* genes and genes *CALN1* and *AUTS2* (**Figure 1C**). Coherently, we

previously showed that the expression of the *AUTS2* gene is slightly modified – albeit not significantly – in lymphoblastoid cell lines of WBS patients (**Table 1**) [13].

*Structural changes concurrently modify gene expression, chromatin architecture and histones marks*

To analyze whether the changes in expression of flanking genes upon deletion of the WBSCR are congruent with modifications in chromatin loops, we replicated the 4C assays in a lymphoblastoid cells from a female WBS patient (**Figure 2C**, **Supplementary Figures S4,S5**). In most cases, interactions are not abrogated but only modified in their intensity in cells with the 7q11.23 microdeletion consistent with the maintenance of one normal allele. 89% (*MDH2* viewpoint), 74% (*HIP1*), 71% (*POR* and *KCTD7*), 69% (*ASL*) and 58% (*GBAS*) of the interacting regions are shared between the Ctrl and WBS cell lines. We next calculated changes in interaction frequency in both cell lines and determined positive and negative Bricks, corresponding to interactions that are significantly increased or decreased in WBS cells, respectively. We found that interactions within the WBSCR are on average decreased approximately two-fold in the WBS cells for the viewpoints mapping close to the WBSCR (*MDH2*, *POR*, *HIP1* and *KCTD7*), consistent with normal looping intensity in the remaining allele and absence of interaction in the deleted allele (**Figure 2D**). Interactions between the *KCTD7* and *POR* viewpoints and a region defined by the *CLIP2* and *GTF2IRD1* genes within the WBSCR were more than two fold diminished in WBS cells. As a result of the deletion, on the rearranged allele the viewpoints are positioned closer on the linear DNA molecule to the region mapping on the other side of the WBSCR. Interaction between these viewpoints and regions beyond the deletion may therefore be increased in WBS cells as previously found in the study of structural rearrangements with 4C [59]. We failed to identify such changes (**Figure 2C**), possibly because our viewpoints map too far away from the breakpoints (*HIP1* the closest viewpoint maps more than 1 Mb away). We hypothesized that only specific DNA/gene loops between regions on opposite sides of the WBSCR might be changed with the deletion, complicating the chromatin landscape. Corroboratively, in WBS cells the *GBAS* viewpoint is

closer in space to the *HIP1*, *POR* and *MDH2* genes, while the *POR* viewpoint and the *AUTS2* gene interact less (**Figure 2B, 2C**). Apart from these short-range changes, we observed multiple changes in long-range interactions over the entire chromosome length, about half of which are increased interactions (**Supplementary Figure S6**). In some instances, we identified interesting patterns of changes: around genes particularly, an increased interaction in WBS cells was concomitant with flanking reduction of looping intensity (**Supplementary Figure S7**). This observation suggests that chromatin reorganization is not dramatic, but rather that the intensity of long-range interactions is modified locally around certain loci. This is consistent with other work that showed that chromatin reorganization is mirrored in local changes in interactions (e.g. for example on the *Hox* gene clusters [55]) and that chromatin has constrained mobility [46, 60, 61] .

**Figure 2.** Modification of short-range interactions in WBS compared to control cells. Heatmap showing the interactions between the viewpoints, as well as the two newly identified interacting partners in control (**A**) and in (**B**) WBS cells. The darker the color in the heatmap, the stronger the interaction. Numbers indicate the corresponding average Brick signal. White color indicates that no Bricks or positive interaction was found. (**C**) Windowed 4C signal of each of the six viewpoints in both Ctrl and WBS cells around the WBSCR (see the legend of **Figure 1C** for details about the structures outlined). The log2-fold change of the windowed 4C data in WBS over control cells was calculated, and the resulting positive or negative Bricks are indicated below each viewpoint graphs, by blue or red bars, respectively. (**D**) Close-up of the log2-fold interaction changes within the WBSCR. The black line indicates the median of the changes within the WBSCR, which is also displayed at the right of each graph. The dashed lines show the 95% confidence interval. The positions of all genes are displayed at the bottom with purple arrows. The area highlighted in grey pinpoints the higher interactions in Ctrl cells between the *KCTD7* and *POR* viewpoints and the region around the *CLIP2* and *GTF2IRD1* genes.

Figure 2
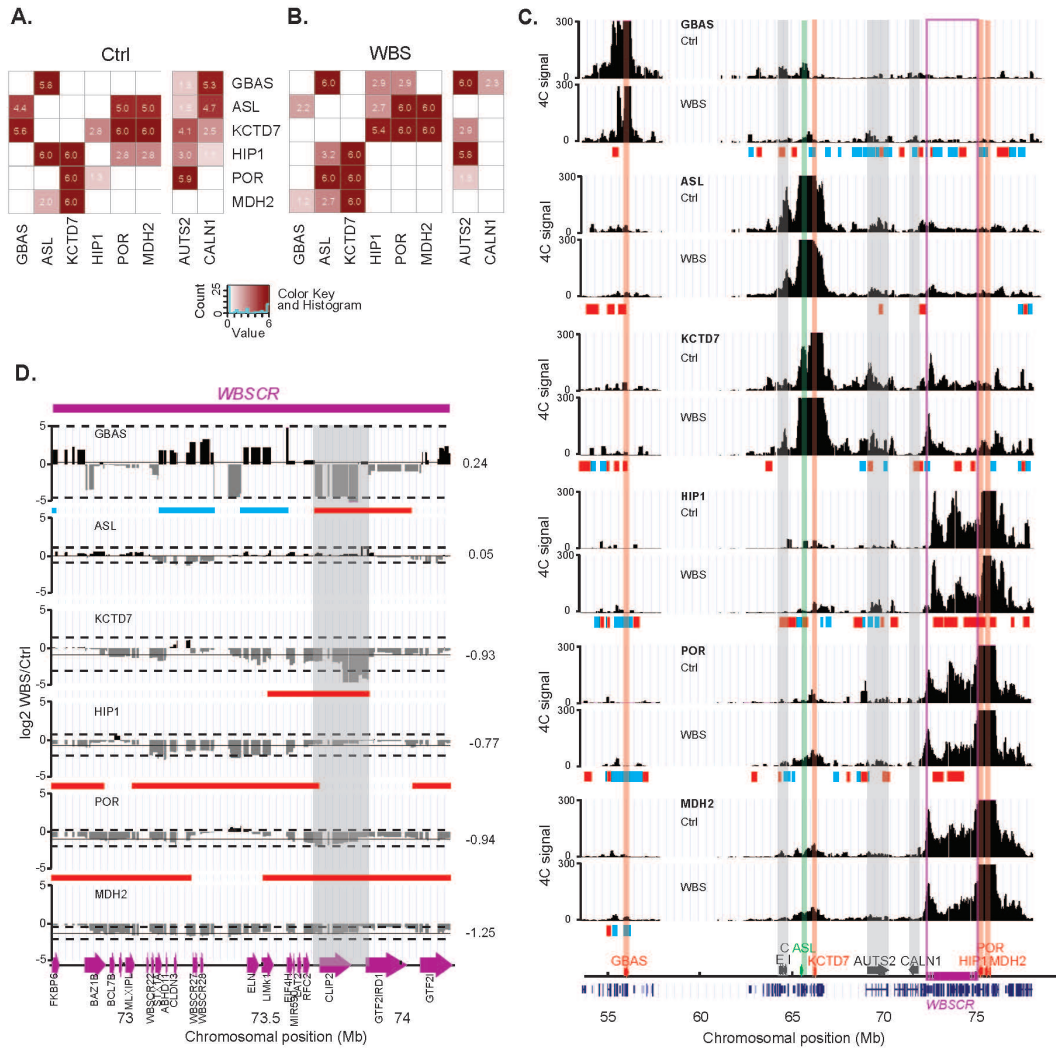
To gain insights into the effects of a structural rearrangement on the chromatin landscape at the nucleosome level, we monitored histone modifications on a genome-wide scale. We measured by ChIP-seq the status of H4K20me1 (monomethylation of Lysine 20 of histone H4) and H3K27me3 (trimethylation of Lysine 27 of histone H3), as proxies for open and condensed chromatin, respectively [22], in lymphoblastoid cell lines of female patients affected by WBS, Williams-Beuren region duplication (WBRdupS, MIM ID #609757 [62]), Smith-Magenis (SMS, MIM ID #182290 [10]) and DiGeorge Syndrome (DGS, MIM ID #188400) and compared them to the female Ctrl individual. WBRdupS is caused by a 7q11.23 microduplication, reciprocal to the WBS deletion, while SMS and DGS are triggered by 17p11.2 and 22q11.2 microdeletions, respectively. The results are detailed in **Supplementary text.** Briefly, the deleted/duplicated chromosomes often present the largest number of altered regions when normalized by their size. These chromatin-modified regions map along the entire length of the rearranged chromosomes (**Figure 3A**). In the vast majority of cases the decrease/increase of ChIP-tags mapping to chromatin-modified regions (or transcripts) within the rearranged intervals correlate with the number of copies of that genomic locus (e.g. three copies of the 7q11.23 band in WBRdupS and 1 copy in WBS cells, respectively)(**Figure 3B**). This observation suggests that the remaining allele in the deletion syndromes and the supernumerary copy in the duplication syndrome are not modified in their chromatin status. A few regions appear to escape this rule, possibly indicating that these are under a dosage compensation mechanism (**Figure 3B**). By comparing the chromatin status in the different cell lines (see methods for details), we identified a set of transcribed regions, which show modified chromatin on the rearranged chromosomes (**Figure 3C-D**). We suggest that some of these modified regions and their associated transcripts might be the origin of some of the phenotypes observed in WBS, WBRdupS, SMS and DGS patients (see **Supplementary text** and below).

We next compared the histone modification status with the interaction profiles of the Ctrl and WBS cells. We found that 4C interacting regions of the five long arm viewpoints are enriched in H4K20me1 marks compared to the rest of

chromosome 7 in Ctrl cells (P=1 x 10$^{-4}$ for ASL, HIP1, POR and MDH2 and P=6 x 10$^{-4}$ for KCTD7, permutation test N=10000), consistent with the clustering of open, actively transcribed regions (**Figure 1A**). H3K27me3 epigenetic marks are similarly enriched in regions interacting with the *POR* and *ASL* viewpoints (P=1x10$^{-3}$, permutation test N=10000), suggesting that chromatin clustering might be determined more by the presence of genes than accessibility of the chromatin (**Figure 1A**). Overlapping islands of both open and closed chromatin marks were observed in mammalian embryonic stem cells and differentiated cells, as well as in various ENCODE cell lines [11, 24, 63, 64]. These regions are defined as "bivalent domains", in which gene promoters are in a poised state with very low levels of transcription. Upon close examination of the histone modifications at the expression-modified genes, we found that four of the six expression-modified genes used as viewpoints (*GBAS*, *POR, ASL* and *HIP1*) show a statistically significant change in chromatin opening between Ctrl and WBS cells (**Table 1**, difference between histone modification peaks defined by SICER with a FDR<1%, see methods for details). *GBAS* and *POR* show a decrease in H4K20me1 marks that parallel their diminished relative expression level in WBS patient cells, whereas an increase in this mark of open chromatin is seen at the *ASL* locus concomitant to its higher expression (**Table 1**). Similarly, *AUTS2* and *CALN1*, which are interaction partners of several of the studied viewpoints showed significant chromatin changes in WBS cells (FDR < 1%). Only *HIP1* shows an increase in H4K20me1 that does not parallel its diminished expression in WBS cells. However, it also presents a significant increase in H3K27me3 marks, which its changed in expression (**Table 1**). In summary, we conclude that structural changes induce concurrent changes in gene expression, chromatin architecture and histones marks extending beyond the borders of the structural change.
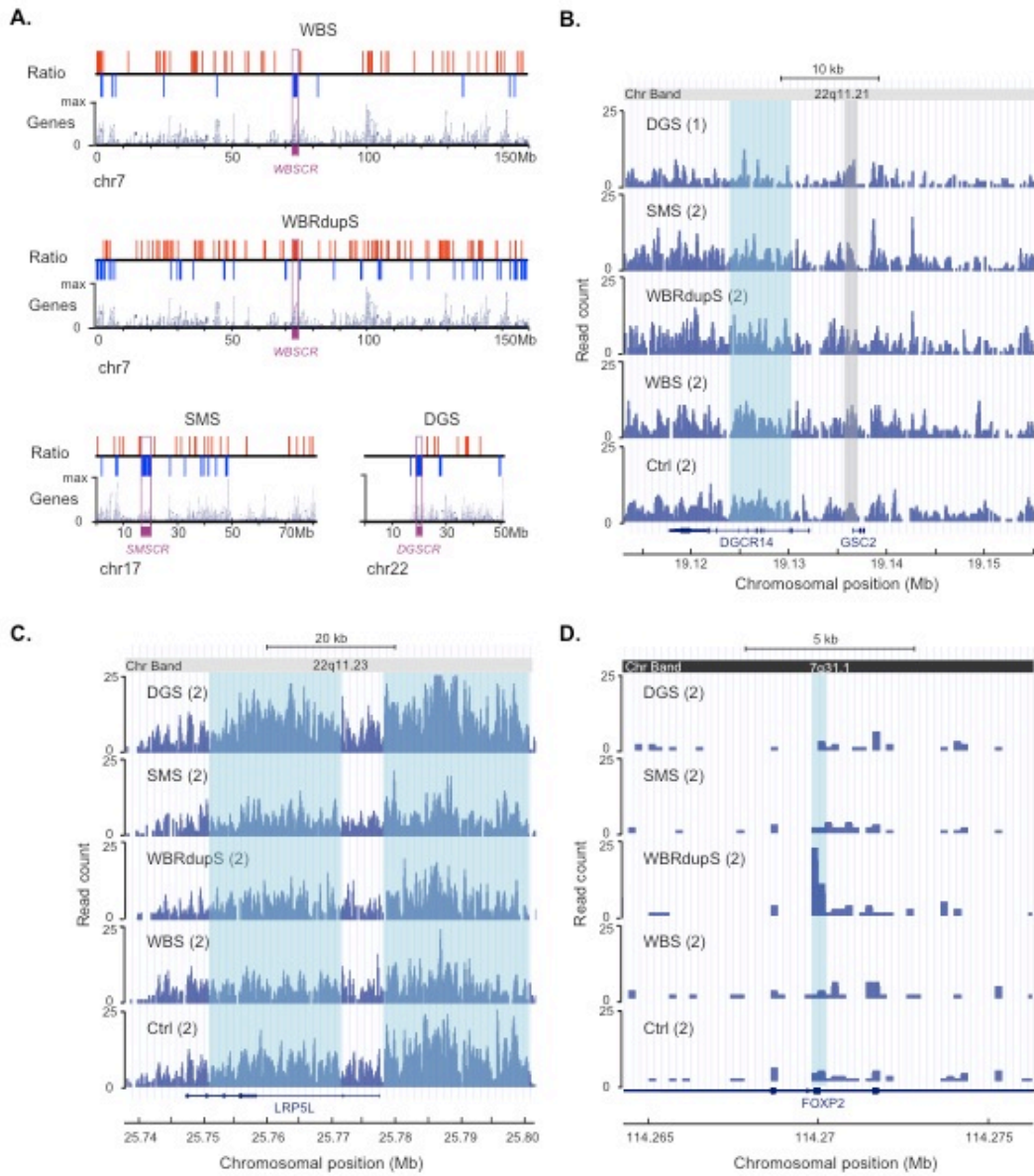
Figure 3

**Figure 3.** Histone modified regions in the WBS, WBRdupS, SMS and DGS patients' cell lines. (**A**) Significant H4K20me1 modifications in the rearranged sample versus controls. Red ticks pinpoint H4K20me1 enrichment in the rearranged sample, while blue ticks represent enrichment in the other samples (FDR=1x10$^{-2}$). Gene density along each of the chromosomes is plotted at the bottom of each graph. The purple horizontal bar below each figure indicates the rearranged regions for WBS, WBRdupS, SMS and DGS on chromosome 7, 7, 17 and 22, respectively. (**B**) Example of a region within the DGS critical region that is depleted in H4K20me1 histone marks in DGS patient cell lines. The introns and exons of the genes mapping within the selected genomic interval are schematically represented at the bottom. The different cell lines assessed are mentioned on the left with indication of the ploidy within the region shown. The *DGSCR14* gene (highlighted in light blue) shows a positive correlation between the number of DNA copies (one in DGS, two in the other samples used as controls; negative binomial test as implemented in DESeq, log2 fold change: -1.09, FDR: 0.00099) and the abundance of histone marks, whereas the *GSC2* gene (highlighted in grey) appears to be under a dosage compensation mechanism as no significant difference was observed between DGS sample and the cell lines used as controls.

Examples of domains with significant increase of H4K20me1 marks outside of the rearranged intervals in DGS (**C,** *LRP5L* gene; log2 fold change: 0.66, FDR: 0.0077) and WBRdupS cells (**D,** *FOXP2* gene; log2 fold change: 2.88, FDR: 0.0019) compared to the other cell lines. The different cell lines assessed are mentioned on the left with indication of the ploidy within the region shown.

## Discussion

Structural variants have been shown to capture 10% to 25% of the expression variance [14, 65]. They influence gene expression by modifying gene dosage and altering the expression of normal-copy number genes located in their vicinity [13, 15, 16, 36, 66]. This effect can be long range with changes in expression of genes positioned megabases away [37, 41]. We investigated the underlying mechanism of genome organization by combining high-throughput chromosome conformation capture and chromosome-wide profiling of histone modifications. Our results suggest that structural rearrangements can influence expression levels of flanking normal-copy genes by affecting large-scale chromatin conformation in various ways.

First, deletion or duplication of specific long-range regulatory elements within the rearrangement, such as enhancers and/or boundary elements, can cause changes in their finely tuned regulatory function and thus in the expression of their target genes. Concordantly, we detect alteration of intrachromosomal interactions (chromosomal looping) between expression-affected gene loci and the rearranged interval in WBS cells using chromosome conformation capture. Specifically, we observe that the interaction between the *POR* and *CLIP2/GTF2IRD1* loci is abolished in WBS cells rather than diminished by 50% as observed for most of the other interactions, suggesting allele-specific chromatin interaction, which was recently postulated by studying the inactive X chromosome [67]. We infer that chromosome looping can be allelically biased through allele-specific enhancer activity and/or gene expression [68-70].

Second, in addition to modifying specific *cis*-acting DNA regulatory elements, a structural rearrangement could also exert its effect on gene expression by changing the overall chromatin topology and DNA accessibility. Genes might be co-regulated by clustering into a "chromatin globule" independently of functional relationship [71]. A strong correlation between interaction frequency and the position of DNase I hypersensitivity sites shows that the organization of the chromatin is tightly linked to the accessibility to regulatory factors [46,

72, 73]. Dislocation of a spatially clustered set of genes might disrupt or modify specific position effect as well as chromatin accessibility, and thereby affect the expression of these genes. Consistent with these assumptions, we pinpointed a multi-looped structure that brings at least seven normal-copy genes with changed expression in presence of the WBS deletion into close proximity to the CR interval and to each other either directly or indirectly. The identified chromatin interactions are modified in cells from WBS patients, suggesting that changes in the genome structure cause concomitant modifications of gene expression, chromatin interactions and histone marks. The complexity of the observed changes prevent us, however, to distinguish whether the changes are a primary or secondary effect of the mechanisms described above.

Some of these modifications may be associated with specific phenotypic features observed in genomic disorders patients. A tantalizing example from our study is the *AUTS2* gene. Its looping architecture, chromatin structure changes and mild expression modification in WBS cells designate this gene as a potential candidate in some of the phenotypes shown by WBS or WBRdupS patients. *AUTS2* is mutated or translocated in autistic patients and individuals with intellectual disabilities [74-76], phenotypes shared by patients with Williams-Beuren region duplication syndrome. Even though the lymphoblastoid cell lines used in this study might not be the best target cell/tissue for many of the genes involved in these disease processes, experiments with these cells are still worth pursuing, simply because we cannot exclude a broad to ubiquitous expression pattern for these genes. Of note previous experiments have shown a high degree of correlation in gene expression levels between different tissues/cell lines for the genes mapping within the aneuploid segments [13, 77]. Further studies are warranted to confirm that AUTS2 expression is modified in other tissue.

**Authors Contribution**

NG, RMW and AR designed the study. NG performed the 4C experiments, while RMW completed the ChIPs. GD and LH helped in preparing the necessary materials and producing part of the data. EM, ML, NG, RMW and JR conducted the statistical analyses. AR wrote the manuscript with contributions from NG, RMW and EM. AR obtained the necessary financial support. All authors read and approved the final manuscript.

**Competing interest**

The authors declare no competing interests.

## Appendix I: Supplementary text

We identified between 5580 and 16388 regions with altered histones status when comparing each genomic syndrome patient cell line with the Ctrl. The deleted/duplicated chromosome often presents the largest number of altered regions when normalized by their size. If we postulate that each syndrome considered here (WBS, WBRdupS, SMS and DGS) exert their pathological effects through independent pathways, we can consider for a given syndrome all other samples with a rearrangement on a different chromosome as controls (see methods for details). For example, not only the control cell line, but also the results obtained with the WBS, WBRdupS and SMS cell line can be used to monitor chromatin changes in DGS cells on HSA22. We performed these comparisons using DESeq. This approach yields from 107 to 248 and from 0 to 97 regions showing chromatin status changes for H4K20me1 and H3K27me3, respectively (FDR < 10%). Chromatin-modified regions map not only within the microdeletion/microduplication, but also along the entire length of the rearranged chromosomes (**Figure 3A**). In the vast majority of cases the decrease/increase of ChIP-tags mapping to chromatin-modified regions (or transcripts) within the rearranged interval correlates with the number of copies of that genomic locus (e.g. three copies of the 7q11.23 band in WBRdupS and 1 copy in WBS cells, respectively)(**Figure 3B**). This observation suggests that the remaining allele in the deletion syndromes and the supernumerary copy in the duplication syndrome are not modified in their chromatin status. A few regions and transcripts appear to escape this rule, possibly indicating that these are under a dosage compensation mechanism. For example, the *GSC2* (goosecoid homeobox 2) gene region shows similar density of H4K20me1 marks in cell lines with one and two copies of that gene (**Figure 3B**). Coherently, compensation at the expression level of some genes was shown in multiple aneuploidies and was typically used to downgrade possible candidate genes [13, 77-79].

We next zoomed in on the chromatin status only of RefSeq genes (see methods for details). This second approach allows identifying a set of genes,

which show modified chromatin on the rearranged chromosomes. Some of the flanking genes modified in their expression levels in WBS cells have less H4K20me1 marks in these cells (see main text). Certain identified regions and transcripts suggest interesting hypotheses about the possible origin of some of the phenotypes shown by WBS, WBRdupS, SMS and DGS patients. For example, we found that a region corresponding to the *FOXP2* (forkhead-box DNA-binding domain) gene contained significantly more H4K20me in the WBRdupS cell line compared to the other cell lines. Because deleterious variants in *FOXP2* are associated with severe language and speech disorders in human and heterozygote mice knockout show reduced vocalization in pups, it is tantalizing to hypothesize that the severe delay in expressive language observed in patients with WBR duplication or triplication [80, 81] might be caused, at least in part, by changes in the regulation of this gene. The identified chromatin change maps to a region encoding 4 exons of *FOXP2* corresponding to the beginning of protein coding transcripts FOXP2-201 (ENST00000393491) and CCDS5761.1 (**Figure 4D**). The function of these isoforms is unclear, but we should mention however that both encode the Forkhead DNA-binding domain (Pfam PF00250). Consistent with the above hypothesis, we had found previously that *FOXP2* is slightly upregulated in skin fibroblasts of WBS patients (fold change = 1.2, BH-adjusted $P$ = 0.04)[82], who reciprocally harbor only one copy of the 7q11.23 region and often show enhanced quantity and quality of speech and vocabulary.

**Table 1: Expression changes and chromatin architecture modifications in WBS cells**

| Gene | Category | Expression changes (ref 16) | Expression changes (this work) | H4K20me1 changes [*] | H3K27me changes |
|------|----------|------|------|------|------|
| *GBAS* | viewpoint | -0.43 | -1.12 | -0.67 | NS |
| *ASL* | viewpoint | 0.67 | 0.41 | 1 | NS |
| *KCTD7* | viewpoint | -1.36 | -0.22 | *-0.1* | NS |
| *HIP1* | viewpoint | -1.09 | -1.2 | 0.81 | 1.38 |
| *POR* | viewpoint | -0.17 | *-0.44* | -0.73 | NS |
| *MDH2* | viewpoint | 0.3 | -0.56 | *-0.22* | NS |
| *AUTS2* | novel interactor | *-1.47* | ND | -1.55 | 2.58 |
| *CALN1* | novel interactor | BDL | BDL | -0.45 | 0.77 |
| *WBSCR22* | positive control | -1.22 | -1.41 | -1.67 | NS |

Changes in expression and chromatin structure of the six viewpoints, two interacting genes and one gene within the WBSCR as positive control in WBS versus Ctrl cells. Changes are presented as the log2-fold ratio between WBS and Ctrl cells (values in italics are not statistically different).

BDL = below detection level
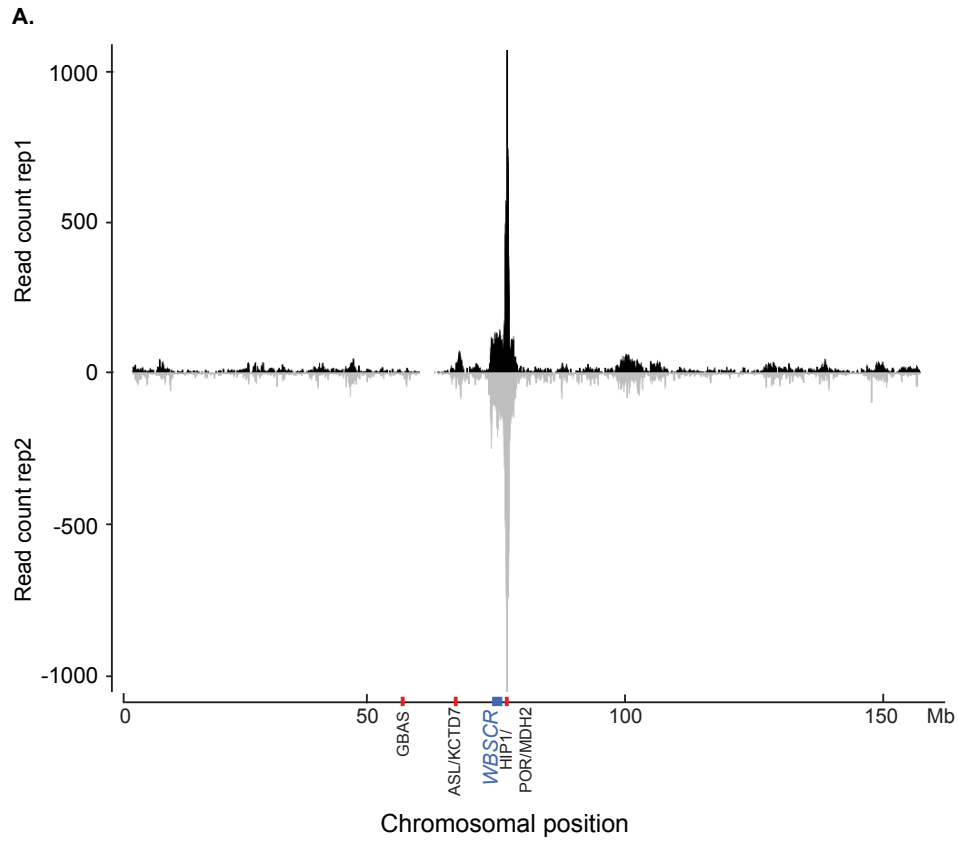AUTS2 expression change was not significant (P=0.06)
ND= not determined
NS= no regions within the gene were defined as significantly changed
[*] most significant block according to SICER within the gene (FDR<1%)

# Appendix II: Supplementary figures, figures legends and tables

**A.**



**B.**

| Viewpoint | Cells | Replicate | # mappable reads (million) | correlation (Pearson) between replicate scores of smoothed 4Cseq signal |
|---|---|---|---|---|
| GBAS | Ctrl | R1 | 2.65 | |
| | WBS | R1 | 2.05 | |
| ASL | Ctrl | R1 | 6.33 | 0.83 |
| | | R2 | 4.38 | |
| | WBS | R1 | 6.4 | 0.93 |
| | | R2 | 3.74 | |
| KCTD7 | Ctrl | R1 | 2.4 | |
| | WBS | R1 | 1.45 | |
| HIP1 | Ctrl | R1 | 2.03 | |
| | WBS | R1 | 1.26 | |
| POR | Ctrl | R1 | 3.17 | 0.96 |
| | | R2 | 2.52 | |
| | WBS | R1 | 5.1 | 0.93 |
| | | R2 | 6.05 | |
| MDH2 | Ctrl | R1 | 3.07 | 0.97 |
| | | R2 | 7.38 | |
| | WBS | R1 | 2.73 | 0.96 |
| | | R2 | 3.88 | |

WBS cells
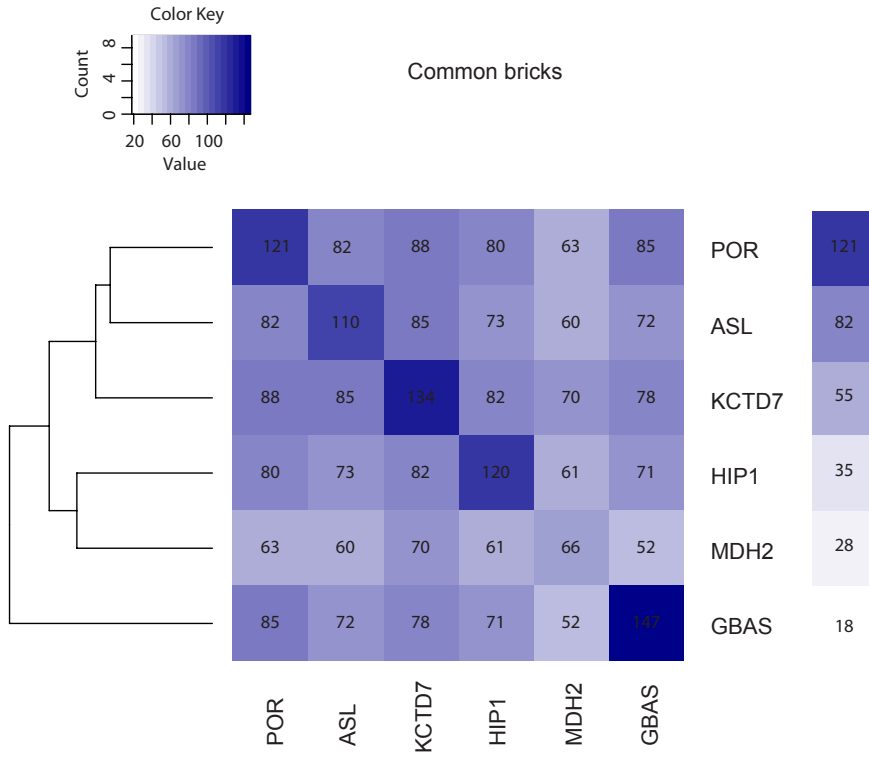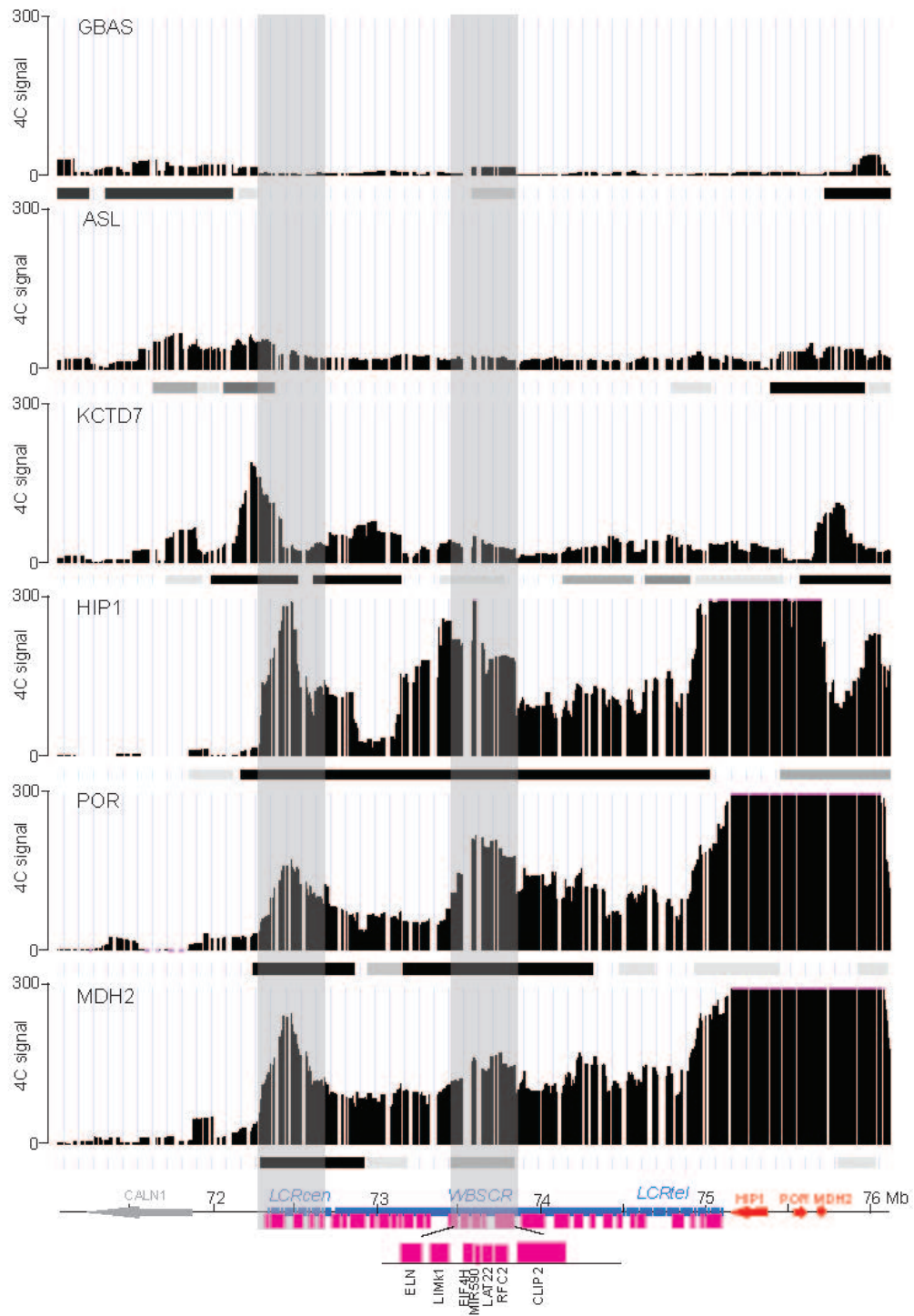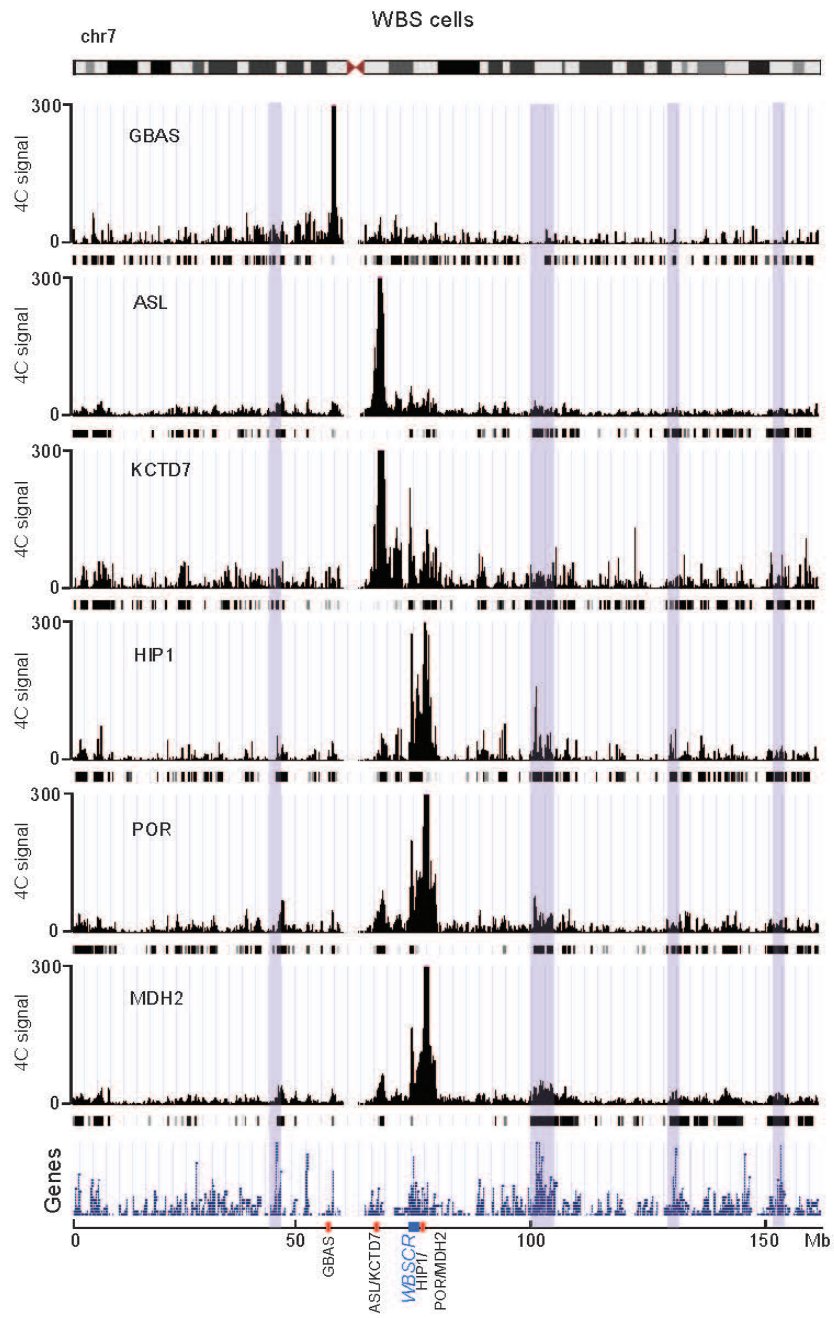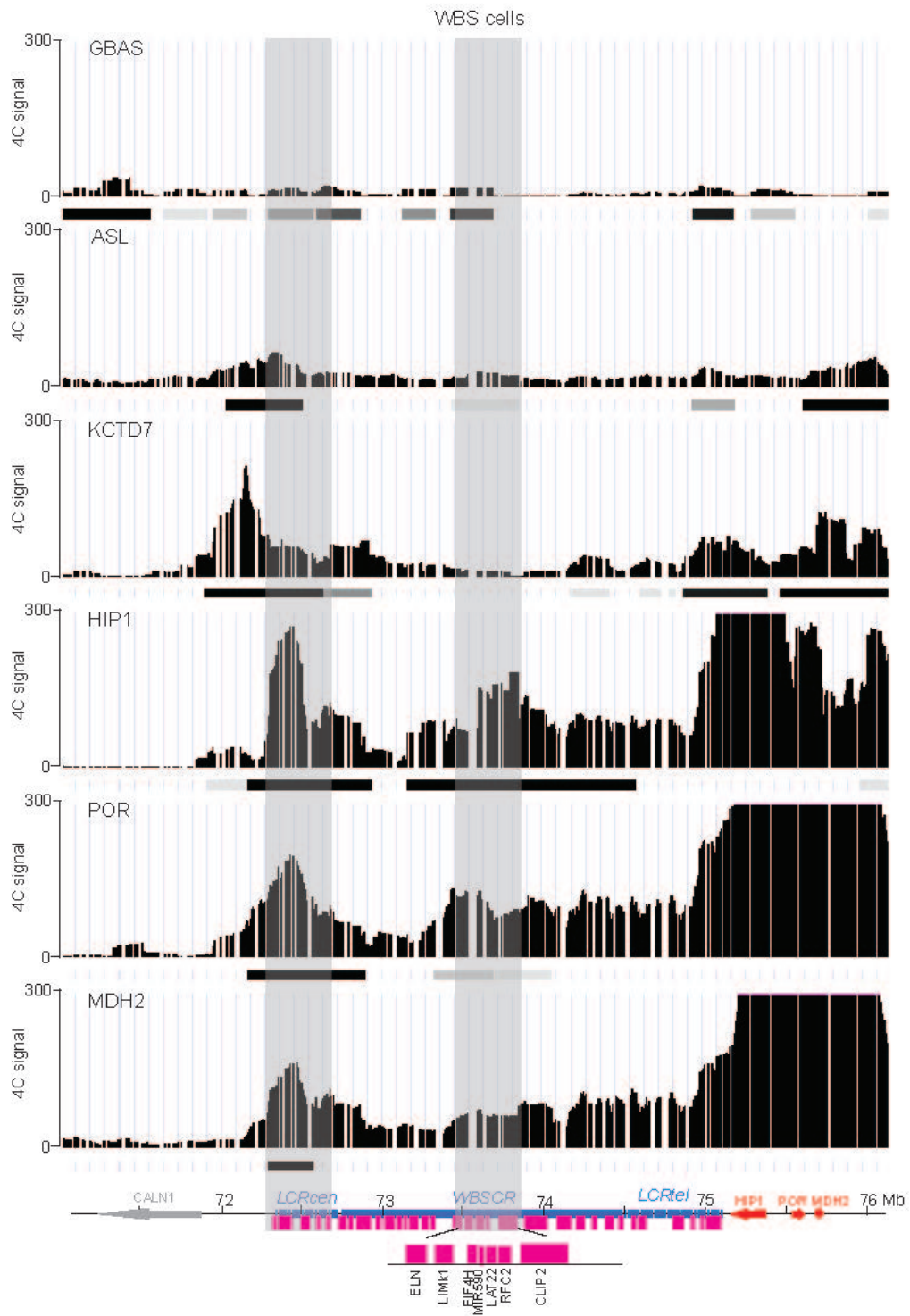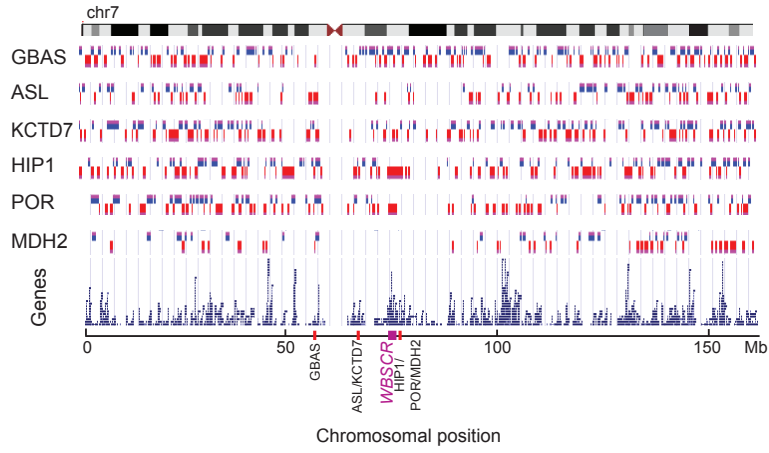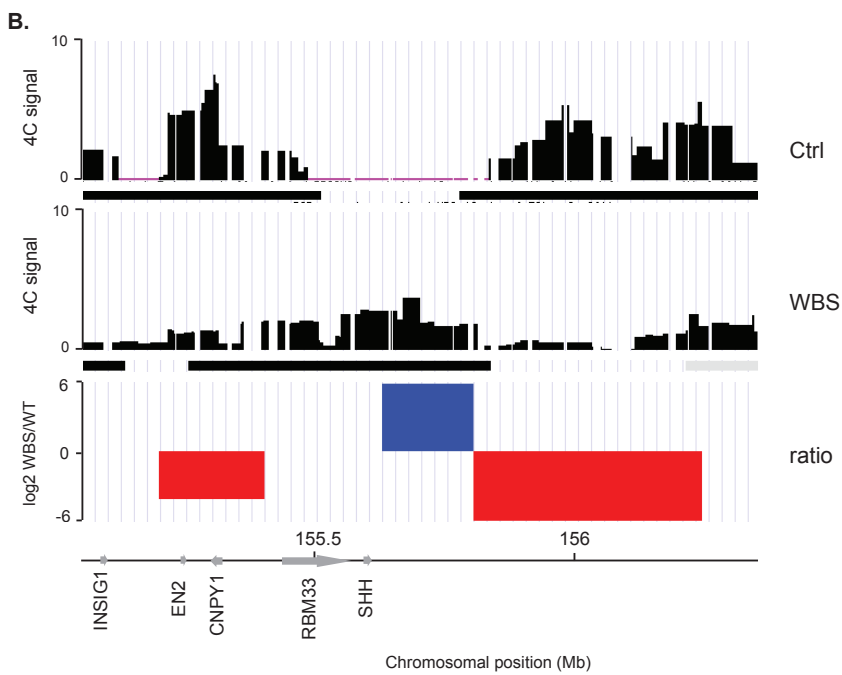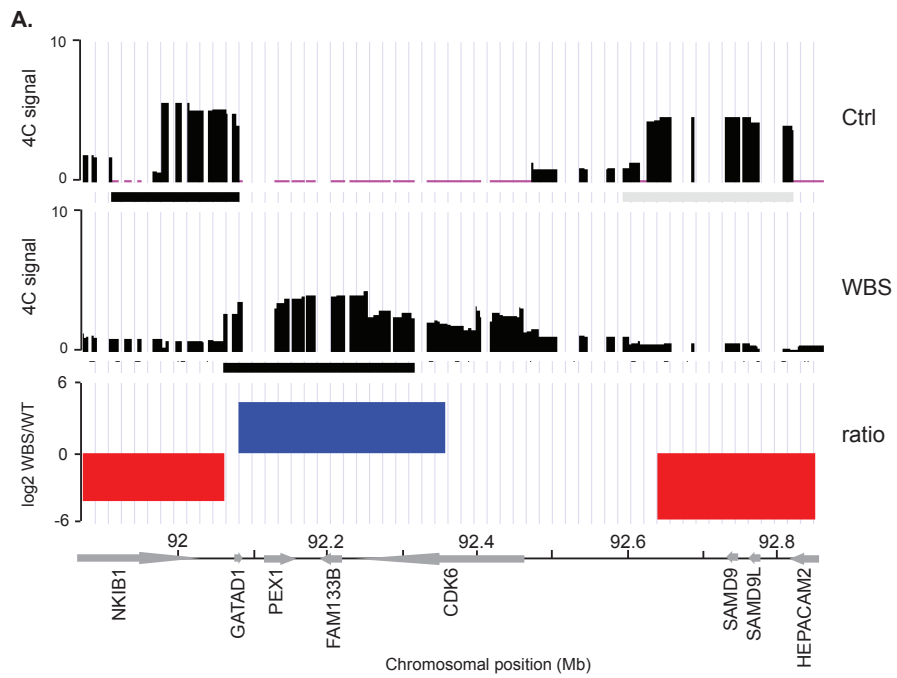
**Supplementary Figure S1.** Reproducibility of 4C-seq experiments. (**A**) Mirror plot of the windowed 4C scores of two biologically independent replicates using *MDH2* as viewpoint (Pearson correlation = 0.97). (**B**) Overview of the number of mappable reads per viewpoint and per cell line, as well as Pearson correlation coefficient between bioreplicates.

**Supplementary Figure S2.** Heatmap showing the number of Bricks in each viewpoint in control cells, as well as the pairwise overlap between Bricks of the different viewpoints. The column at the right shows the number of Bricks that cumulatively overlap between the different viewpoints.

**Supplementary Figure S3.** Close-up of the interactions of the six viewpoints with the WBSCR in cells from a healthy control individual. The two areas highlighted in grey show the strongly interacting regions at the LCRcen (centromeric LCR) and the region within WBSCR. Pink boxes indicate the mapping of genes within the WBSCR.

**Supplementary Figure S4.** Interactions of six genes on human chromosome 7 in cells from a WBS patient. Windowed 4C signal of each of the six viewpoints along the entire chromosome 7 (HSA7). The black ticks below each graph show the location of the Bricks. The density of genes is shown at the bottom. Areas highlighted in blue pinpoint some examples of strong correlation of gene-dense regions with H4K20me1 marks and highly interacting regions. The mapping of the viewpoints and the WBSCR is indicated at the bottom.

**Supplementary Figure S5.** Close-up of the interactions of the six viewpoints with the WBSCR in cells from a WBS patient. The two areas highlighted in grey show the strongly interacting regions at the LCRcen (centromeric LCR) and the region within WBSCR. Pink boxes indicate the mapping of genes within the WBSCR.

**Supplementary Figure S6.** Differentially interacting regions in WBS compared to control cells along the entire chromosome 7. The blue and red

boxes correspond to significantly increased and decreased interactions (Bricks: Blocks of Regulators In Chromosomal Kontext) in WBS cells, respectively. The position of the viewpoints and the WBSCR (purple horizontal bar) are indicated at the bottom. The percentage of increased interactions is 53% (*GBAS* viewpoint), 45% (*ASL*), 51% (*KCTD7*), 45% (*HIP1*), 49% (*POR*) and 44% (*MDH2*).

**Supplementary Figure S7.** Examples of regions with modified interactions with the *POR* viewpoint. The log2-fold change of the windowed 4C data in WBS over control cells is plotted. Positive or negative Bricks are indicated below each viewpoint graph, by blue or red bars, respectively. In WBS cells, the region around the *CDK6* gene (**A**) or sonic hedgehog (*SHH* gene) (**B**) interacts with the *POR* gene, whereas in control cells, the flanking regions interact more frequently, indicating local changes in interactions.

**Supplementary Table S1**. 4Cseq primer sequences

| Gene viewpoint | Primer name | Sequence |
|---|---|---|
| GBAS | GBAS_4C_seq_F | AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT TTCTAAGCGGACATTTTCCT |
|  | GBAS_4C_seq_R | CAAGCAGAAGACGGCATACGAGTGTGGTGATTCATCCCTGT |
| ASL | ASL_4C_seq_F3 | CAAGCAGAAGACGGCATACGA GCTCCAGTGATCAGGACCAG |
|  | ASL_4C_seq_R3 | AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT TGGGTTGAATGAGCAACAGT |
| KCTD7 | KCTD7_4C_seq_F | AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT CT TTCAGAGCTCACCAAGGTTTG |
|  | KCTD7_4C_seq_R | CAAGCAGAAGACGGCATACGACCACGTACTCTGAAAA |
| HIP1 | HIP1_4C_seq_R | AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCT GTAGTGAGCGGGGGCATT |
|  | HIP1_4C_seq_F | CAAGCAGAAGACGGCATACGAGTTGGGCACCTGCTTCAT |
| POR | POR_4C_seq_F | AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCTCTTCCCCTCCCTACCAC |
|  | POR_4C_seq_R | CAAGCAGAAGACGGCATACGGCTAAGGAACGCGTCCAA |
| MDH2 | MDH2_4C_seq_F2 | AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGATACTGGAGCCCGAGATGA |
|  | MDH2_4C_seq_R | CAAGCAGAAGACGGCATACGACCCAGTTCGTTTAGGCGTTC |

## Perspectives

The insides into the copy number variation pathology that were gained as a consequence of this study are pointing to the chromatin structure changes as one of the mechanisms influencing expression of the genes located beyond the boarders of a given structural rearrangement. We identified examples of structurally modified chromatin regions within the genes that could be directly linked to known phenotypes of the patients affected by CNV-related disorders. One of the most striking examples is an elevated level of an open chromatin mark mapping to the locus of FOXP2 gene. The altered H4K20me1 signal is strongly localized to the 4th exon of this gene, possibly enabling formation of the particular isoform of its transcript. As FOXP2 is a gene that has been implicated in language development [83], promoting expression of different isoforms could potentially be causative for language skill deficiency in WBRdupS patients. The attempts to investigate this discovery in more details using quantitative PCR to assess expression of different isoforms of FOXP2 in WBRdupS and control cell lines failed due to low expression levels of this gene in lymphoblastoid cell lines. The tissue that is implicated in language, so would most likely be affected by FOXP2 expression changes, is neuronal tissue, but availability of neuronal cells from WBRdupS patient is very limited. One of the possible ways of investigating expression changes in neurons would be the use of animal models. Numerous animal models of WBS have been reported (reviewed in [84]), but the models of WBRdupS are still lacking. Whilst creating our own mouse model would be possible, it would be very time consuming. The other limitation in using animal models to assess complex traits – like level of expressive language development – is the lack of ability to measure them in a standardized manner. Although there are reports describing association of FOXP2 levels with changes in vocalization in mouse pups, direct translation of this observation to delayed language development in human patients would be difficult.

The other possibility to validate the influence of increased levels of H4K20me1 on gene expression would be experiments in the cells with

reduced levels of this histone marks. That could be achieved by either knocking-out a specific H4K20 methylase or overexpressing H4K20 demethylase. Unfortunately, despite the fact that unspecific enzymes know to (un-) methylate 20th lysine on histone H4 have been described, the H4K20 specific enzymes remain unknown [85]. Over- or underexpression of the unspecific histone (de-) methylase would possibly trigger massive chromatin structure rearrangements, preventing us from inferring the impact on gene expression that is specific to H4K20me1.

Another way of validation of our findings would be performing additional ChIP assays using antibodies against histones with modifications that correlate with the marks used in our experiments or simply repeating the experiments adding more biological replicates. Both approaches would give us more statistical power to confirm the results obtained previously.

## Current work

To further characterize biology of the genome and its regulatory mechanisms we will continue to investigate the influence of structural variation and polymorphism in the genome on genome-wide epigenetic profiles and gene expression phenotypes.

We are currently studying the influence of genomic variation on gene expression and histone modification status in general human populations. We are characterizing, in the frame of a Swiss National Science Foundation Sinergia grant grouping the laboratories of Emmanouil Dermitzakis (University of Geneva), Nouria Hernandez (University of Lausanne), Bart Deplancke (EPFL, Lausanne) and Alexandre Reymond, genomes of lymphoblastoid cell lines from related individuals of European and African descent (two mother-father-sibling trios). These individuals are part of the 1000 Genome Project[86], giving us the opportunity to assess the genomic variation in their genomes and correlate it to expression, transcription factor binding and histone modification studies that we are performing. The goal of this project is to assess the correlation between these different assays and DNA sequence information in healthy individuals and to measure the influence of allele specificity on these assays. This would allow to better understand regulatory mechanisms in the human genome and the degree to which they are heritable. Simply stated, we want to understand the interplay between sequence, chromatin structure and gene expression.

This collaborative effort is focusing at our end on three different histone marks that pinpoint TSS of active and poised genes, closed chromatin and transcribed units, i.e. H3K4me3, H3K27me3 and H4K20me1, respectively. The consortium will also produce and analyze data for a fourth histone mark (H3K4me1), transcription factors (PU.1, MYC and TFIIB) and RNA polymerase II binding, as well as DNA methylation, gene and microRNA expression levels through RNA-seq and GRO-seq.

We are still optimizing the procedures we will use for the data analysis pipeline. However, the preliminary analysis showed correlating profiles between RNA polymerase II binding, transcription factors binding and histone modifications marking actively transcribed genes in the window of 5kb around transcription start site (TSS). Low levels of RNA polymerase II were, in contrast, correlated with H3K27me3 – a mark of silenced chromatin (Figure 4).
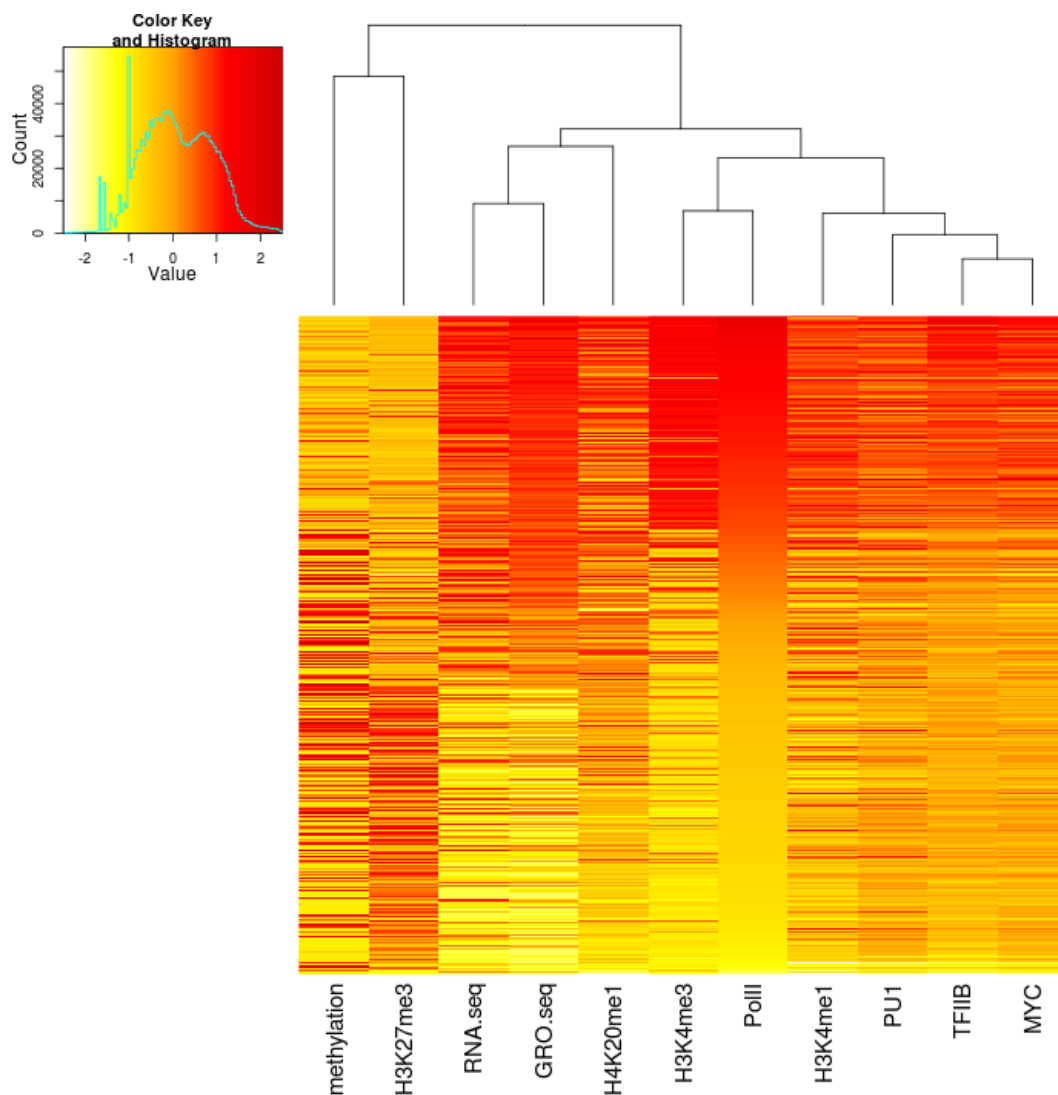


**Figure 4.** Heatmap presenting signal strength and clustering of individual assays in a 5kb window around TSS. Scale based on a Z-score; red represents strong and yellow weak signal, respectively. Columns are sorted according to PolII signal.

The study of the two trios was planned as a pilot phase of the project and should be completed in the following months. In parallel to the data analysis of the pilot phase, data production for additional samples is taking place: we are currently including 54 additional individuals of European descent in our study. For this phase of the project we will survey H3K4me3, H3K27ac (a marker of enhancer elements), H3K4me1, transcription factors (PU.1 and MYC) and RNA polymerase II binding and gene expression. The level of complexity of the functional elements regulation in the genome is immense, so adding more data would allow us to gain statistical power, therefore resulting in the ability to draw additional conclusion from the study.

## Conclusions

Genome biology is the field gaining more and more interest in the recent years. One of the reasons behind this, except the huge leap in the DNA sequencing technology, is an increasing awareness of the incompleteness of our knowledge in the field, which for many years was based on, so called, central dogma of molecular biology describing interactions between DNA, RNA and proteins. Although it is unquestionable that this dogma describes the general flow of sequence information, the complexity of the interactions between these basic carriers of information in molecular biology is vastly understated. To complement the understanding of the transfer of genetic information we need to unravel the complexity of the regulation of individual elements.

Our ability to identify the effect of large genomic variants, such as copy number variation, on chromatin structure is crucial to fully comprehend how genetic variation influences human health. Until recently, the common understanding of pathophysiology of copy number related diseases was that causative genes are most likely the genes affected by the copy number change. It was shown, however, that genes located outside of the copy number varying region could also differ in expression, thus potentially be causative. Here we studied one of the putative mechanisms of how CNVs can influence genes located outside of the rearrangement, namely modifications in the structure of chromatin, in patients with genomic disorders compared to healthy individuals. Changes in chromatin conformation can lead to altered expression of the genes encoded by the loci affected by the change, and this eventually can influence the phenotype of affected individual. This type of alteration in the phenotype would be driven directly by the gene potentially located far away from the initial genomic variant, while CNV effect would be indirect, although ultimately causative. In this thesis we show that copy number variations can influence the chromatin state throughout the whole chromosome on which they are located, and by this we add to the

understanding of the mechanisms of pathology related to this type of genomic disorders.

# Acknowledgments

I would like to thank Professor Alexandre Reymond for believing in me and giving me the opportunity to join his laboratory to perform this thesis. It was both – nice and productive – time spent there.

I am very grateful to Phil Shaw for all the technical advices, especially at the beginning of my work. It would be difficult for me to complete the experiments without him shearing his expertise.

I would also like to thank Eugenia Migliavacca and Nele Gheldof for fruitful collaboration and making this project possible to come to an end.

Many thanks to all the members of the lab. It was really fun to work with you all! Special thanks to Gerard Didelot, Andreas Gschwind and Katrin Mannik for hours of scientific – and not only scientific – discussions and to Emilie Ait Yahya Graison and Nele Gheldof for help with the summary of this thesis in French.

I would also like to thank my parents and my whole family for supporting my idea of leaving my home country to pursue scientific career.

And last but not least – huge THANK YOU to Michalina. For support in everyday matters, for the scientific discussions, for being there whenever needed!

# Bibliography

1. Kidd, J.M., et al., *Mapping and sequencing of structural variation from eight human genomes.* Nature, 2008. **453**(7191): p. 56-64.

2. Lupski, J.R., *Genomic rearrangements and sporadic disease.* Nature genetics, 2007. **39**(7 Suppl): p. S43-7.

3. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.

4. Itsara, A., et al., *Population analysis of large copy number variants and hotspots of human genetic disease.* American journal of human genetics, 2009. **84**(2): p. 148-61.

5. Girirajan, S., C.D. Campbell, and E.E. Eichler, *Human copy number variation and complex genetic disease.* Annual review of genetics, 2011. **45**: p. 203-26.

6. Cooper, G.M., et al., *A copy number variation morbidity map of developmental delay.* Nature genetics, 2011. **43**(9): p. 838-46.

7. Pober, B.R., *Williams-Beuren syndrome.* The New England journal of medicine, 2010. **362**(3): p. 239-52.

8. Fishman, I., et al., *Language and sociability: insights from Williams syndrome.* Journal of neurodevelopmental disorders, 2011. **3**(3): p. 185-92.

9. Van der Aa, N., et al., *Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome.* European journal of medical genetics, 2009. **52**(2-3): p. 94-100.

10. Elsea, S.H. and S. Girirajan, *Smith-Magenis syndrome.* Eur J Hum Genet, 2008. **16**(4): p. 412-21.

11. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.

12. McDonald-McGinn, D.M. and K.E. Sullivan, *Chromosome 22q11.2 deletion syndrome (DiGeorge syndrome/velocardiofacial syndrome).* Medicine, 2011. **90**(1): p. 1-18.

13. Merla, G., et al., *Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes.* Am J Hum Genet, 2006. **79**(2): p. 332-41.

14. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes.* Science, 2007. **315**(5813): p. 848-53.

15. Henrichsen, C.N., et al., *Segmental copy number variation shapes tissue transcriptomes.* Nat Genet, 2009. **41**(4): p. 424-9.

16. Cahan, P., et al., *The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells.* Nat Genet, 2009. **41**(4): p. 430-7.

17. Orozco, L.D., et al., *Copy number variation influences gene expression and metabolic traits in mice.* Hum Mol Genet, 2009. **18**(21): p. 4118-29.

18. Stryer, L., *Biochemistry.* 4th ed1995, New York: W. H. Freeman and Company.

19. Zhang, Z. and B.F. Pugh, *High-resolution genome-wide mapping of the primary structure of chromatin.* Cell, 2011. **144**(2): p. 175-86.

20. Margueron, R., P. Trojer, and D. Reinberg, *The key to development: interpreting the histone code?* Current opinion in genetics & development, 2005. **15**(2): p. 163-76.

21. Portela, A. and M. Esteller, *Epigenetic modifications and human disease.* Nature biotechnology, 2010. **28**(10): p. 1057-68.

22. Ernst, J. and M. Kellis, *Discovery and characterization of chromatin states for systematic annotation of the human genome.* Nature biotechnology, 2010. **28**(8): p. 817-25.

23. Ernst, J., et al., *Mapping and analysis of chromatin state dynamics in nine human cell types.* Nature, 2011.

24. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.

25. Dahl, J.A. and P. Collas, *A rapid micro chromatin immunoprecipitation assay (microChIP).* Nat Protoc, 2008. **3**(6): p. 1032-45.

26. Graubert, T.A., et al., *A high-resolution map of segmental DNA copy number variation in the mouse genome.* PLoS Genet, 2007. **3**(1): p. e3.

27. Cutler, G., et al., *Significant gene content variation characterizes the genomes of inbred mouse strains.* Genome Res, 2007. **17**(12): p. 1743-54.

28. She, X., et al., *Mouse segmental duplication and copy number variation.* Nat Genet, 2008. **40**(7): p. 909-14.

29. Conrad, D.F., et al., *Origins and functional impact of copy number variation in the human genome.* Nature, 2010. **464**(7289): p. 704-12.

30. Zhang, F., et al., *Copy number variation in human health, disease, and evolution.* Annu Rev Genomics Hum Genet, 2009. **10**: p. 451-81.

31. Fanciulli, M., E. Petretto, and T.J. Aitman, *Gene copy number variation and common human disease.* Clin Genet, 2009.

32. Lee, C. and S.W. Scherer, *The clinical context of copy number variation in the human genome.* Expert Rev Mol Med, 2010. **12**: p. e8.

33. Carvalho, C.M., F. Zhang, and J.R. Lupski, *Evolution in health and medicine Sackler colloquium: Genomic disorders: a window into human gene and genome evolution.* Proc Natl Acad Sci U S A, 2010. **107 Suppl 1**: p. 1765-71.

34. Mills, R.E., et al., *Natural genetic variation caused by small insertions and deletions in the human genome.* Genome research, 2011. **21**(6): p. 830-9.

35. Henrichsen, C.N., E. Chaignat, and A. Reymond, *Copy number variants, diseases and gene expression.* Hum Mol Genet, 2009. **18**(R1): p. R1-8.

36. Chaignat, E., et al., *Copy number variation modifies expression time courses.* Genome research, 2011. **21**(1): p. 106-13.

37. Ricard, G., et al., *Phenotypic consequences of copy number variation: insights from Smith-Magenis and Potocki-Lupski syndrome mouse models.* PLoS biology, 2010. **8**(11): p. e1000543.

38. Jacquemont, S., et al., *Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus.* Nature, 2011. **478**(7367): p. 97-102.

39. Kleinjan, D.A. and L.A. Lettice, *Long-range gene control and genetic disease.* Advances in genetics, 2008. **61**: p. 339-88.

40.     Reymond, A., et al., *Side effects of genome structural changes.* Curr Opin Genet Dev, 2007. **17**(5): p. 381-6.

41.     Harewood, L., et al., *The effect of translocation-induced nuclear reorganization on gene expression.* Genome Res, 2010. **20**(5): p. 554-64.

42.     Gheldof, N., et al., *Detecting Long-Range Chromatin Interactions Using the Chromosome Conformation Capture Sequencing (4C-seq) Method.* Methods in molecular biology, 2012. **786**: p. 211-25.

43.     Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* Nat Genet, 2006. **38**(11): p. 1348-54.

44.     Simonis, M., J. Kooren, and W. de Laat, *An evaluation of 3C-based methods to capture DNA interactions.* Nature methods, 2007. **4**(11): p. 895-901.

45.     Tolhuis, B., et al., *Interactions among Polycomb domains are guided by chromosome architecture.* PLoS genetics, 2011. **7**(3): p. e1001343.

46.     Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.

47.     de Wit, E., et al., *Global chromatin domain organization of the Drosophila genome.* PLoS genetics, 2008. **4**(3): p. e1000045.

48.     Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.

49.     O'Geen, H., et al., *Comparison of sample preparation methods for ChIP-chip assays.* BioTechniques, 2006. **41**(5): p. 577-80.

50.     Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome biology, 2009. **10**(3): p. R25.

51.     Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

52.     Zang, C., et al., *A clustering approach for identification of enriched domains from histone modification ChIP-Seq data.* Bioinformatics, 2009. **25**(15): p. 1952-8.

53.	Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome biology, 2010. **11**(10): p. R106.

54.	Pruitt, K.D., T. Tatusova, and D.R. Maglott, *NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.* Nucleic acids research, 2005. **33**(Database issue): p. D501-4.

55.	Noordermeer, D., et al., *The dynamic architecture of Hox gene clusters.* Science, 2011. **334**(6053): p. 222-5.

56.	Schoenfelder, S., et al., *Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells.* Nature Genetics, 2010. **42**(1): p. 53-61.

57.	Osborne, C.S., et al., *Active genes dynamically colocalize to shared sites of ongoing transcription.* Nat Genet, 2004. **36**(10): p. 1065-71.

58.	Dutly, F. and A. Schinzel, *Unequal interchromosomal rearrangements may result in elastin gene deletions causing the Williams-Beuren syndrome.* Human Molecular Genetics, 1996. **5**(12): p. 1893-8.

59.	Simonis, M., et al., *High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology.* Nature methods, 2009. **6**(11): p. 837-42.

60.	Soutoglou, E. and T. Misteli, *Mobility and immobility of chromatin in transcription and genome stability.* Current opinion in genetics & development, 2007. **17**(5): p. 435-42.

61.	Chambeyron, S. and W.A. Bickmore, *Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.* Genes Dev, 2004. **18**(10): p. 1119-30.

62.	Merla, G., et al., *Copy number variants at Williams-Beuren syndrome 7q11.23 region.* Human genetics, 2010. **128**(1): p. 3-26.

63.	Roh, T.Y., et al., *The genomic landscape of histone modifications in human T cells.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15782-7.

64.	Shu, W., et al., *Genome-wide analysis of the relationships between DNaseI HS, histone modifications and gene expression reveals distinct modes of chromatin domains.* Nucleic acids research, 2011. **39**(17): p. 7428-43.

65. Yalcin, B., et al., *Sequence-based characterization of structural variation in the mouse genome.* Nature, 2011. **477**(7364): p. 326-9.

66. Guryev, V., et al., *Distribution and functional impact of DNA copy number variation in the rat.* Nat Genet, 2008. **40**(5): p. 538-45.

67. Splinter, E., et al., *The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA.* Genes & development, 2011. **25**(13): p. 1371-83.

68. Pastinen, T., *Genome-wide allele-specific analysis: insights into regulatory variation.* Nature reviews. Genetics, 2010. **11**(8): p. 533-8.

69. Wasserman, N.F., I. Aneas, and M.A. Nobrega, *An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer.* Genome research, 2010. **20**(9): p. 1191-7.

70. McDaniell, R., et al., *Heritable individual-specific and allele-specific chromatin signatures in humans.* Science, 2010. **328**(5975): p. 235-9.

71. Sanyal, A., et al., *Chromatin globules: a common motif of higher order chromosome structure?* Current opinion in cell biology, 2011. **23**(3): p. 325-31.

72. Yaffe, E. and A. Tanay, *Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture.* Nature genetics, 2011. **43**(11): p. 1059-65.

73. Sung, M.H. and G.L. Hager, *More to Hi-C than meets the eye.* Nature genetics, 2011. **43**(11): p. 1047-8.

74. Ben-David, E., et al., *Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression.* Human molecular genetics, 2011. **20**(18): p. 3632-41.

75. Huang, X.L., et al., *A de novo balanced translocation breakpoint truncating the autism susceptibility candidate 2 (AUTS2) gene in a patient with autism.* American journal of medical genetics. Part A, 2010. **152A**(8): p. 2112-4.

76. Kalscheuer, V.M., et al., *Mutations in autism susceptibility candidate 2 (AUTS2) in patients with mental retardation.* Human genetics, 2007. **121**(3-4): p. 501-9.

77.   Lyle, R., et al., *Gene expression from the aneuploid chromosome in a trisomy mouse model of down syndrome.* Genome Res, 2004. **14**(7): p. 1268-74.

78.   Ait Yahya-Graison, E., et al., *Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes.* Am J Hum Genet, 2007. **81**(3): p. 475-91.

79.   Prandini, P., et al., *Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance.* Am J Hum Genet, 2007. **81**(2): p. 252-63.

80.   Beunders, G., et al., *A triplication of the Williams-Beuren syndrome region in a patient with mental retardation, a severe expressive language delay, behavioural problems and dysmorphisms.* Journal of medical genetics, 2010. **47**(4): p. 271-5.

81.   Somerville, M.J., et al., *Severe expressive-language delay related to duplication of the Williams-Beuren locus.* N Engl J Med, 2005. **353**(16): p. 1694-701.

82.   Henrichsen, C.N., et al., *Using transcription modules to identify expression clusters perturbed in williams-beuren syndrome.* PLoS computational biology, 2011. **7**(1): p. e1001054.

83.   Lai, C.S., et al., *A forkhead-domain gene is mutated in a severe speech and language disorder.* Nature, 2001. **413**(6855): p. 519-23.

84.   Osborne, L.R., *Animal models of Williams syndrome.* American journal of medical genetics. Part C, Seminars in medical genetics, 2010. **154C**(2): p. 209-19.

85.   Koturbash, I., et al., *Alterations in histone H4 lysine 20 methylation: implications for cancer detection and prevention.* Antioxid Redox Signal, 2012. **17**(2): p. 365-74.

86.   *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.