Computational Prediction of Position Effects of Apparently Balanced Human Chromosomal Rearrangements

DOI:
10.1016/j.ajhg.2017.06.011

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
Zepeda-Mendoza, C. J., Ibn-Salem, J., Kammin, T., Harris, D. J., Rita, D., Gripp, K. W., MacKenzie, J. J., Gropman, A. L., Graham, B. H., Shaheen, R., Alkuraya, F. S., Brasington, C. K., Spence, E. J., Masser-Frye, D., Bird, L. M., Spiegel, E., Sparkes, R. L., Ordulu, Z., Talkowski, M. E., ... Morton, C. C. (2017). Computational Prediction of Position Effects of Apparently Balanced Human Chromosomal Rearrangements. *American Journal of Human Genetics*. https://doi.org/10.1016/j.ajhg.2017.06.011

Published in:
American Journal of Human Genetics

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

Download date:26. Aug. 2022

Computational Prediction of Position Effects of Apparently Balanced Human Chromosome Rearrangements

Cinthya J. Zepeda-Mendoza[1,3,*], Jonas Ibn-Salem[4,*], Tammy Kammin[1], David J. Harris[3,5], Debra Rita[6], Karen W. Gripp[7], Jennifer J. MacKenzie[8], Andrea Gropman[9], Brett Graham[10], Ranad Shaheen[11], Fowzan S. Alkuraya[11,12], Campbell K. Brasington[13], Edward J. Spence[13], Diane Masser-Frye[14], Lynne M. Bird[14,15], Erica Spiegel[16], Rebecca L. Sparkes[17], Zehra Ordulu[18], Michael E. Talkowski[18-24], Miguel A. Andrade-Navarro[4], Peter N. Robinson[25], Cynthia C. Morton[1-3,23,26†]

Departments of [1]Obstetrics, Gynecology and Reproductive Biology and [2]Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA

[3]Harvard Medical School, Boston, MA 02115, USA

[4]Johannes Gutenberg University, Mainz 55122, Germany

[5]Boston Children's Hospital, Boston, MA 02115, USA

[6]ACL laboratories. Cytogenetics Lab, Rosemont, IL 60018, USA

[7]Nemours Alfred I. DuPont Hospital for Children; Wilmington, DE 19803, USA

[8]Department of Pediatrics, McMaster University, Hamilton, ON L8S 4L8, Canada

[9]Children's National Medical Center, Washington, DC 20010, USA

[10]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[11]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 12713, Saudi Arabia

[12]Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh 11533, Saudi Arabia

[13]Clinical Genetics Division, Department of Pediatrics. Levine Children's Hospital at Carolinas Medical Center. Charlotte, NC 28203, USA

[14]Genetics and Dysmorphology, Rady Children's Hospital San Diego, San Diego, CA 92123, USA

[15]University of California, San Diego, La Jolla, CA 92093, USA

[16]Maternal Fetal Medicine, Columbia University Medical Center, New York, NY 10032, USA

[17]Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada

Departments of [18]Pathology, [19]Neurology and [20]Psychiatry and [21]Center for Genomic Medicine and, Massachusetts General Hospital, Boston, MA 02114, USA

[22]Department of Neurology, Harvard Medical School, Boston, MA 02115, USA

[23]Program in Medical and Population Genetics and [24]Stanley Center for Psychiatric Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[25]The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

[26]Division of Evolution and Genomic Science, School of Biological Sciences, Manchester Academic Health Science Centre, Manchester M13 9NT, UK

*These authors contributed equally to this work

[†]Corresponding author: Cynthia C. Morton, Ph.D. Brigham and Women's Hospital, New Research Building, Rm. 160D, 77 Avenue Louis Pasteur. Boston, MA 02115. Tel: 617-525-4535; Fax: 617-525-4533. Email: cmorton@partners.org

Abstract

Interpretation of variants of uncertain significance, especially chromosome rearrangements in non-coding regions of the human genome, remains one of the biggest challenges in modern molecular diagnosis. To improve our understanding and interpretation of such variants, we used high-resolution 3-dimensional chromosome structure data and transcriptional regulatory information to predict position effects and their association with pathogenic phenotypes in 17 subjects with apparently balanced chromosome abnormalities. We find that the rearrangements predict disruption of long-range chromatin interactions between several enhancers and genes whose annotated clinical features are strongly associated with the subjects' phenotypes. We confirm gene expression changes for a couple of candidate genes to exemplify the utility of our position effect analysis. These results highlight the important interplay between chromosome structure and disease, and demonstrate the need to utilize chromatin conformation data for the prediction of position effects in the clinical interpretation of cases of non-coding chromosome rearrangements.

Introduction

The importance of the integrity of chromosome structure and its association with human disease is one of the oldest and most studied topics in clinical genetics. As early as 1959, cytogenetic studies in humans linked specific genetic or genomic disorders and intellectual disability syndromes to changes in chromosomal ploidy, translocations, and DNA duplications and deletions[1-4]. The discovery of copy-number variants (CNVs) by microarray and sequencing technologies expanded the catalogue of genetic variation between individuals to test such associations at higher resolution.[5-14] Over the years, analysis of disease-related structural

rearrangements has illuminated genes that are mutated in various human developmental disorders.[15-18] Such chromosome aberrations can directly disrupt gene sequences, affect gene dosage, generate gene fusions, unmask recessive alleles, reveal imprinted genes, or result in alterations of gene expression through additional mechanisms such as position effects.[15] The latter is particularly important for the study of apparently balanced chromosome abnormalities (BCAs), such as translocations and inversions, often found outside of the hypothesized disease-causing genes (reviewed in [19]).

Position effects were first identified in *Drosophila melanogaster*, where chromosomal inversions placing *white+* near centric heterochromatin caused mosaic red/white eye patterns.[20] In humans, BCAs can induce position effects through disruption of a gene's long-range transcriptional control (*i.e.*, enhancer-promoter interactions, insulator influence, etc.), or its placement in regions with different local chromatin environments as observed in the classical *Drosophila* position effect variegation (reviewed in [19, 21, 22]). Examples of position effect genes include paired box gene 6 (*PAX6* [MIM: 607108]), for which downstream chromosome translocations affect its *cis*-regulatory control and produce aniridia (AN [MIM: 106210]);[23, 24] twist family bHLH transcription factor 1 (*TWIST1* [MIM: 601622]), where downstream translocations and inversions are associated with Saethre-Chotzen syndrome (SCS [MIM: 101400]);[25] paired like homeodomain 2 (*PITX2* [MIM: 601542]) for which translocations are associated with Axenfeld-Rieger syndrome type 1 (RIEG1 [MIM: 180500]);[26, 27] SRY-box 9 (*SOX9* [MIM: 608160]), where translocation breakpoints located up to 900 Kilobases (Kb) upstream and 1.3 Megabases (Mb) downstream are associated with campomelic dysplasia (CMPD [MIM: 114290]),[28] in addition to several others.[19, 29, 30]

The availability of genome sequencing in the clinical setting has generated a need for

rapid prediction and interpretation of structural variants, especially those pertaining to *de novo* non-coding rearrangements in individual subjects. With the development and subsequent branching of the chromosome conformation capture (3C) technique ([31], reviewed in [32]), regulatory issues such as alteration of long-range transcriptional control and position effects can now be predicted in terms of chromosome organization. The high resolution view of chromosome architecture in diverse human cell lines and tissues[33-40] has allowed molecular assessment of the disruption of regulatory chromatin contacts by pathogenic structural variants and single nucleotide changes; examples include the study of limb malformations,[41] leukemia,[42] and obesity,[43] among others.[44-49] These examples underscore the importance of chromatin interactions in quantitative and temporal control of gene expression, which can greatly enhance our power to predict pathologic consequences.

To test the feasibility of prediction and clinical interpretation of position effects of non-coding chromosome rearrangements, we analyzed 17 subjects from the Developmental Gene Anatomy Project (DGAP)[18, 50-53] with *de novo* non-coding BCAs classified as variants of uncertain significance (VUS). Using publicly available chromatin contact information, annotated and predicted regulatory elements, and correlation between phenotypes observed in DGAP subjects and those associated with neighboring genes, we reliably predicted candidate genes exhibiting mis-regulated expression in DGAP-derived lymphoblastoid cell lines (LCLs). These results suggest that many VUS are likely to be further interpretable via long-range effects, and warrant their routine assessment and integration in clinical diagnosis.

Materials and Methods

Selection of subjects with apparently balanced chromosome abnormalities

BCA breakpoints and clinical data were obtained from DGAP cases for which whole-genome sequencing was performed using a previously described large-insert jumping library approach.[18, 50-54] A total of 151 cases were filtered to select only subjects whose translocation or inversion breakpoints fall within intergenic regions (GRCh37) and did not overlap known long intergenic non-coding RNAs (lincRNAs) or pseudogenes, as these elements have been shown to exert functional roles (reviewed in [55; 56, 57]). Of 151 DGAP subjects, only 17 fulfilled our selection criteria, 12 of whom had available and reportedly normal clinical array results, suggesting lack of large duplications or deletions.

Clinical descriptions of DGAP cases

The clinical presentation of the 17 subjects varied, ranging from developmental delay to neurological conditions, offering the opportunity to assess long-range position effects in different phenotypes. Subjects' karyotypes are presented in the main text using the International System for Human Cytogenetic Nomenclature (ISCN2016) (Table 1). Detailed case descriptions are included in the Supplemental Note: Case Reports, as well as a nomenclature developed to describe chromosome rearrangements using next-generation sequencing.[58] Reported ages of DGAP subjects are from time of enrollment. All reported genomic coordinates use GRCh37.

Analysis of genes bordering the rearrangement breakpoints

The presence of annotated genes or pseudogenes and lincRNAs was assessed in windows of ±3 and ±1 Mb neighboring each subject's translocation and inversion breakpoints, and within reported H1-hESC topologically associated domains (TADs)[35] where the breakpoints were located. The gene annotation file was obtained from Ensembl GRCh37 archive,[59] and we used

the Human Body Map lincRNAs catalog.[60] Haploinsufficiency (HI) and triplosensitivity scores were assigned using Huang *et al.*, 2010[61] and version hg19 of ClinGen[62] data downloaded on 9/20/2016.

Assessment of disrupted functional elements and chromatin interactions bordering rearrangement breakpoints

The disruption of regulatory elements such as enhancers, promoters, locus control regions, and insulators can lead to disease-related gene expression changes; DNase I hypersensitive (DHS) sites have been used as markers for the identification of such elements.[63] In addition, the alteration of TAD boundaries has been previously shown to cause a rewiring of enhancers with pathological consequences;[41, 46, 64] CCCTC-Binding Factor (CTCF) binding sites have been found to be enriched in TAD boundaries,[35] and several mutations of boundary-defining sites have been associated with cancer.[65, 66] Based on these observations, we assessed the number of regulatory elements that were potentially disrupted by the analyzed DGAP breakpoints. We compared the breakpoint positions of the selected DGAP subjects against data corresponding to CTCF binding sites, DHS sites, and chromatin segmentation classifications (Broad ChromHMM) derived from a lymphoblastoid cell line (GM12878) and human stem cells (H1-hESC), obtained from the Encyclopedia of DNA Elements (ENCODE) project[67] and accessed through the University of California Santa Cruz Genome Browser.[68] Enhancer positions were additionally obtained from Andersson *et al.*, 2014[69] for tissue and primary cells, and the VISTA Enhancer browser, human version hg19.[70] Finally, lists of transcription factor (TF) binding sites and gene promoters were obtained from the Ensembl database human version GRCh37.[59] Hi-C interaction data and TAD positions for H1-hESC, GM06990, and IMR90 at 20 Kb, 40 Kb, 100 Kb, and 1

Mb resolution were obtained from Dixon *et al*., 2012[35] and the WashU EpiGenome Browser.[71] A high-resolution dataset of chromatin loops and domains was obtained from Rao *et al.*, 2014 for IMR90 and GM12878 cells.[38] Lastly, distal DHS/enhancer–promoter connections[63] were used to assess disrupted predicted cis-regulatory interactions by the BCAs. Genomic overlaps between the rearrangement breakpoints, functional elements and disrupted chromatin interactions were calculated using custom Perl scripts, the BEDtools suite[72] and the genomic association tester (GAT) tool.[73]

Ontological analysis of genes neighboring breakpoints

Phenotype similarity between potential position effect genes and DGAP cases was calculated by converting the phenotypes of the 17 subjects to Human Phenotype Ontology (HPO)[74] terms and calculating their phenomatch score as described in Ibn-Salem *et al*., 2014.[48] The phenomatch score quantifies the information content of the most specific HPO term that is part of or a common ancestor (more general term) of a set of phenotypes. Our set of phenotypes is constituted by the HPO terms associated to DGAP cases and the ones annotated to candidate position effect genes within windows of ±3 and ±1 Mb of sequence in proximity to the breakpoints. We used two background models to assess significance of this similarity. The first is based on randomly permuting the associations of phenotypes to genes; to this effect, the phenotype-gene associations are shuffled 100 times randomly and the similarity of these random phenotypes to the studied case clinical findings is calculated. The second background control is based on shifting the breakpoint location along the chromosome; each breakpoint is shifted by -9, -6, -3, +3, +6, and +9 Mb and the similarity of genes in proximity to the shifted breakpoints is computed.

Quantitative real-time PCR

LCLs derived from DGAP236-02m, DGAP244-02m and DGAP245-02m were used as karyotypically normal male controls. These are karyotypically normal fathers of enrolled DGAP cases with no history of disease. LCL 17402 (DGAP163) was used to test differential gene expression for SOS Ras/Rac guanine nucleotide exchange factor 1 (*SOS1* [MIM: 182530]), and LCL 18060 (DGAP176) was used to test midline 2 (*MID2* [MIM: 300204]), p21 (RAC1) activated kinase 3 (*PAK3* [MIM: 300142]), and POU class 3 homeobox 4 (*POU3F4* [MIM: 300039]) expression using quantitative polymerase chain reaction (qPCR). Glucuronidase beta (*GUSB* [MIM: 611499]) was used as a housekeeping control. qPCR experiments were performed by the Harvard Biopolymers Facility using TaqMan probes Hs00264887_s1 (*POU3F4*), Hs00201978_m1 (*MID2*), Hs00176828_m1 (*PAK3*), Hs00893134_m1 (*SOS1*), and Hs00939627_m1 (*GUSB*). Data were analyzed using the ΔCT method.

Assessment of DGAP breakpoints overlapping with non-coding structural variants in public databases

To find similar non-coding structural rearrangement subjects and compare their annotated clinical phenotypes to those observed in DGAP cases, we searched the DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER)[75] version 2015-07-13, as well as the dbVar database from the National Center for Biotechnology Information (NCBI) Variation Viewer 1.5.[76] Both databases are comprehensive community-supported repositories of clinical cases with novel and extremely rare genomic variants.

Results

Genomic characterization of non-coding breakpoints

To study the structural and evolutionary context of BCAs and their impact on nuclear architecture and gene expression, we used data generated by DGAP,[18, 50-53] the largest collection of sequenced balanced chromosome rearrangements from individuals with abnormal developmental and cognitive phenotypes, many of which have yet to be investigated in detail. Each studied DGAP BCA has two breakpoint positions (as two distinct chromosome regions are involved in their generation), which we labeled with the DGAP#_A and DGAP#_B identifiers. We filtered DGAP data to select cases with both breakpoints in non-coding regions only, and excluding lincRNAs and pseudogenes; a total of 17 cases fulfilled our criteria, 15 translocations and 2 inversions (Figure 1 and Table S1). These subjects are phenotypically distinct, and most of them presented with congenital developmental and neurological conditions not recognized as a known syndrome or genomic disorder (see clinical descriptions in Supplemental Note: Case Reports).

Further analysis revealed that BCA breakpoints were significantly depleted for overlapping annotated promoters or transcription factor (TF) binding sites (GAT TF p=0.0003, promoter p=0.0001, Table S2,3). Only one breakpoint (DGAP249_B) overlapped a ChromHMM enhancer in GM12878 cells (Table 1); the others had no overlap with annotated or predicted enhancers in the analyzed datasets, and this depletion was significant for VISTA (GAT p=0.0364) and Hi-ESC (GAT p=0.0036) but not for the annotated tissue and primary cell enhancers from Andersson *et al*., 2014[69] (Table S4). Eight breakpoints overlapped cell-type specific DHS sites (Table 1 and Table S5); these corresponded to DGAP cases 017, 176, 249, 275, 288 and 322; of these, DGAP176 and DGAP275 overlapped DHS sites at both BCA

breakpoint sites. In addition, three DGAP cases overlapped CTCF binding sites in H1-hESC (DGAP cases 111, 176, and 287) and none in GM12878 cells (Table 1 and Table S6). Except for two cases in H1-hESC (DGAP17 and DGAP176), and four cases in GM12878 (DGAP 017, 126, 163 and 176), all rearrangements fall within ChromHMM repressed chromatin regions, but this association was not significant (GAT p= 0.40 for GM12878 and p= 0.15 for H1-hESC, Table S2F). Interestingly, 22 of the 34 breakpoints (~65%) overlap repeated elements at a significant level (GAT p=0.0002, Table S8), which may indicate a non-allelic homologous recombination process in their generation.[77, 78]

Noticeably, either one or two breakpoints from all the non-coding DGAP BCAs fall within previously reported TADs in H1-hESC and IMR90 cell lines (Table 1 and Table S9).[35] However, this overlap was not significant for both cell lines (GAT H1-ESC p=0.0537 and IMR90 p=0.28). We found that the breakpoints disrupt dozens, hundreds, or even thousands of chromatin contacts when assessed at the 20 and 40 Kb resolution in Hi-C data of H1-hESC and IMR90 cells, as well as chromatin contacts at 100 Kb and 1 Mb resolution in GM06990 cells (Table S11). Breakpoint DGAP111_A had a consistent absence of disrupted chromatin contacts, which is expected as it overlaps a repetitive satellite region so no chromatin contacts could be mapped to the segment (Table S9 and Table S11). With the availability of higher resolution data, it is possible to detect whether BCA breakpoints disrupt smaller chromatin domains and loops not detected in previous studies. When analyzing high resolution IMR90 and GM12878 Hi-C data,[38] we discovered that 32 out of 34 breakpoints are contained within GM12878 sub-compartments (Table 1 and Table S10); interestingly, 28 of these are classified as members of the B compartment, which is less gene dense and less expressed compared to the A compartment. On the other hand, 18 and 24 breakpoints are contained within GM12878 and

IMR90 arrowhead domains, respectively (Table S10), which are regions of enhanced contact frequency that tile the diagonal of each chromatin contact matrix. In addition, the breakpoints disrupt several significant short and long-range chromatin interactions in the GM12878 Hi-C data (Table S12).

Overall, the observation of breakpoint-associated DHS sites suggests the alteration of underlying regulatory elements with potential pathogenic outcomes, while the predicted extensive disruption of chromatin contacts and the alteration of TAD boundaries by the BCAs may affect long-range regulatory interactions of neighboring genes (see Discussion).

Identification of genes with potential position effects

To identify genes which could be generating the complex DGAP phenotypes via position effects from chromosome rearrangements, we analyzed all annotated genes within windows of ±3 and ±1 Mb proximal and distal to the breakpoints, and within the BCA-containing H1-hESC reported TAD positions. A total of 3081 genes were contained within the ±3 and ±1 Mb windows for all cases; 106 of these genes (~3.4%) have an HI score of <10%, which is a predictor of haploinsufficiency,[61] and 55 and two genes have ClinGen emerging evidence suggesting that dosage haplo/triplo-sensitivity, respectively, is associated with clinical phenotype (Table S15).

To further refine our search for genes which may exhibit position effects, we performed an unbiased correlation between DGAP case phenotypes and the clinical traits associated with genes bordering each breakpoint. To this end, we used the HPO dataset,[74] which provides a standardized vocabulary of phenotypic abnormalities encountered in human disease, and currently contains ~11,000 terms and over 115,000 annotations to hereditary diseases. We translated DGAP clinical features to HPO terms (Table S16), and calculated phenotype similarity

between DGAP cases and neighboring genes using the phenomatch score.[48] The phenomatch score distinguishes between general and very specific phenotypic descriptions by quantifying the information content of the most specific HPO terms that are common to, or a common ancestor of, the DGAP case and neighboring gene phenotypes. The similarity significance is then calculated based on randomly permuting the associations of phenotypes to genes, and in shifting the DGAP translocation and inversion breakpoint positions along the chromosome. We obtained phenomatch scores ranging from 0.003 to 91.48 for 179 genes within the ±3 and ±1 Mb windows, as well as within the TAD positions (Table S15).

In addition to dosage sensitivity and phenotypic similarity information, we complemented our analysis with assessment of enhancer-promoter interactions to make our candidate selection more specific. A typical mechanism by which chromosome rearrangements cause position effects is through disruptions in the association of genes with their regulatory regions.[19, 29] We therefore reasoned that genes and enhancers included in predicted enhancer-promoter interactions would be strong position effect candidates. We used the ENCODE distal DHS/enhancer–promoter connections[63] to assess disrupted predicted *cis*-regulatory interactions by the DGAP breakpoints within a ±500 Kb window. The analysis revealed 193 genes that were separated from their predicted candidate enhancers, potentially altering gene expression (Table S13). A total of 133 candidate genes were separated from <10 of their predicted enhancers, while 60 genes were separated from their predicted interactions with 10 or up to 91 enhancers (Table S14).

For the 17 analyzed DGAP BCAs, there are a total of 645 genes with either evidence of dosage sensitivity, disrupted enhancer-promoter interactions, or significant phenotypic similarity. This represents ~21% of the genes contained within the ±3 Mb windows, clearly an undesirable

number for timely clinical interpretation and functional analyses. To filter the most promising candidates, we ranked them using their reported dosage sensitivity, disrupted regulatory interactions, and by selecting a phenomatch cut-off value capable of detecting pathogenic and likely pathogenic genes in 57 published DGAP cases from Redin *et al*., 2017.[53] By taking into consideration the top quartile values of the reported phenomatch scores per case and adding up their dosage sensitivity and disrupted regulatory interaction data, we consistently ranked the reported pathogenic and likely pathogenic genes in the upper decile for 52 out of the 57 control DGAP cases (~91%) when considering candidates within the TAD and ±1 Mb analysis windows (Table S17). 32 of these genes were the top-ranking candidates in their corresponding DGAP case, while 19 of them were positioned in the second-tier rank. Only five genes could not be found in the top decile ranking positions as they had one or no lines of evidence supporting their inclusion.

Applying this ranking strategy to the 17 non-coding BCAs, we predict 16 top-ranking candidates for 11 DGAP cases and 102 second-tier candidates for the 17 analyzed DGAP cases within ±1 Mb analysis windows (Table 1 and Table S15). This is a significant reduction compared to the initial 645 possible candidates (~3.8% of the neighboring genes in the ±3 Mb windows considering top and second-tier candidates, and only 0.05% considering top candidates only). Of note, only nine of the 16 top-ranking candidates are included within the same TAD as the BCA breakpoint (H1-hESC TADs from [35]), while the rest are located farther away. Nine top-ranking genes had an HI score <10%,[61] while ClinGen HI data revealed that four of these 16 genes are associated with autosomal recessive phenotypes, and an additional seven have sufficient or some evidence for haploinsufficiency. Only one candidate gene for DGAP138, glutamate ionotropic receptor kainate type subunit 2 (*GRIK2* [MIM: 138244]) was a confirmed

triplosensitive annotated gene in ClinGen (Table S15).

Taken together, these cases represent more plausible candidates in the search for position effect genes with functional consequences in the subjects' phenotypes. Examples include *GRIK2* which could explain the intellectual disability observed in DGAP138; *SOS1*, forkhead box G1 (*FOXG1* [MIM: 164874]) and cochlin (*COCH* [MIM: 603196]) may be related to the neurological and developmental delay as well as hearing loss of DGAP163; acyl-CoA synthetase long-chain family member 4 (*ACSL4* [MIM: 300157]) and *POU3F4* could be involved in DGAP176's cognitive impairment and hearing loss; SATB homeobox 2 (*SATB2* [MIM: 608148]) may underlie the delayed speech and language development observed in DGAP249; RB binding protein 8 endonuclease (*RBBP8* [MIM: 604124]) may be involved in DGAP252's craniofacial dysmorphic features; *SOX9* most likely explains the cleft palate observed in DGAP288; DNA polymerase epsilon catalytic subunit (*POLE* [MIM: 174762]) may contribute to the extreme short stature observed in DGAP275, and zinc finger E-box binding homeobox 2 (*ZEB2* [MIM: 605802]) can potentially explain the hypotonia and neurological features observed in DGAP329. *SOX9* had been previously proposed to explain DGAP288's phenotype, and as predicted by our method, a decrease in its expression was observed in RNA derived from DGAP288's umbilical cord blood.[49] Additional quantitative real-time PCR analyses revealed *SOS1* as having reduced expression in DGAP163-derived LCLs compared to three normal sex-matched controls (Figure 2).Expression assessment for second-tier candidates *PAK3*, *MID2* and *POU3F4* in DGAP176 LCLs did not deviate substantially from their control expression values (Figure S1); further searches into the Genotype-Tissue Expression (GTEx) project[79] reveal that *PAK3*, *MID2* and *POU3F4* have low expression in LCLs, which would have made assessing changes in expression of these genes technically difficult. This points to the importance of the

availability of tissues and cell lines relevant to the studied phenotypes, or the capacity to generate animal models that reproduce the observed BCAs for further analysis.

Identification of subjects with shared non-coding chromosome alterations and phenotypes

The identification of subjects with shared non-coding chromosome alterations and phenotypes as described herein would further support our idea of these rearrangements exerting their pathogenic outcomes through long-range position effects. To identify such subjects, we searched the DECIPHER[75] and dbVar databases,[76] both comprehensive community-supported repositories of clinical cases with novel or extremely rare genomic variants.

We found 494 DECIPHER cases overlapping our 34 non-coding BCA breakpoints (Table S19). Of these, 489 had rearrangements that overlapped one or more annotated genes (Table S20). Only five DECIPHER cases fulfilled our non-coding selection criteria (Table S21): cases 1985 and 1989, both of which overlap one of DGAP017's breakpoints in chromosome 10, but which have several other gene-altering genomic rearrangements; case 289720, a subject with a 161.44 Kb deletion in chromosome 10 described as likely benign and sharing a sequence breakpoint with DGAP126; case 289865 overlapping a breakpoint in DGAP126 in chromosome 10, very similar to case 289720, however with the presence of an additional pathogenic gene-altering rearrangement; and lastly case 293610, a pathogenic duplication of 364.43 Kb in chromosome 17 sharing a breakpoint with DGAP288. Only two of the five DECIPHER cases have reported clinical phenotypes. DECIPHER case 289720 presents with intellectual disability and psychosis, both pertaining to the superclasses of behavioral and neurodevelopmental abnormalities under the HPO classification. Interestingly, DGAP126 has abnormal aggressive, impulsive or violent behavior and auto-aggression, as well as language and motor delays, which

also fall under the classification of behavioral and neurodevelopmental abnormalities. DECIPHER case 293610 has reported gonadal tissue discordant for external genitalia or chromosomal sex as well as a non-obstructive azoospermia clinical phenotype;[80] both features are not observed until puberty, and are associated with the female-to-male sex disorder observed for CNVs altering the *SOX9* genomic landscape. Although DGAP288 is still an infant, there is no report of sex reversal.

From the dbVar database, 675 non-coding structural rearrangements including CNVs, deletions, inversions, and translocations overlap DGAP breakpoints (Table S22). Of these, only five variants had associated clinical information, including variant nsv534336, a 530 Kb duplication overlapping the DGAP017 breakpoint in chromosome 10, classified as "uncertain significance"[81] and exhibiting a growth delay phenotype; nsv931775, a benign ~381.8 Kb deletion overlapping the DGAP113 breakpoint on chromosome 3, associated with developmental delay and/or other significant developmental or morphological phenotypes;[81] nsv534571, an ~639.7 Kb duplication of uncertain significance associated with muscular hypotonia and overlapping the DGAP287 breakpoint on chromosome 10; and variants nsv532026 and nsv917014, two duplications of ~613 Kb classified as "uncertain significance" and "likely benign," respectively, overlapping the DGAP315 breakpoint in chromosome 6, and associated with developmental delay and/or other significant developmental or morphological phenotypes as well as autism and global developmental delay. All the detected variants are associated with phenotypes observed in the DGAP cases, especially DGAP017's hypoplasia, the developmental delay observed in DGAP113, and DGAP315's significant developmental or morphological phenotypes.

Strictly speaking, these phenotypes are disparate, but fall under similar phenotypic

categories, which could enable identification of long-range effect genes between different cases with similar clinical features and chromosome rearrangements. These comparisons highlight the importance of establishing detailed, specific, and unbiased guidelines for assigning phenotypes when performing computational phenotype comparisons.

Discussion

Structural variation of the human genome, either inherited or arising by *de novo* germline or somatic mutations, can give rise to different phenotypes through several mechanisms. Chromosome rearrangements can alter gene dosage, promote gene fusions, unmask recessive alleles, or disrupt associations between genes and their regulatory elements. The traditional clinical focus of studying genes disrupted by chromosome rearrangements has shifted to also assess regions neighboring these variants.[49] This search for positional effects has been particularly important in the analysis of chromosome rearrangements associated with different clinical conditions and disrupting non-annotated genomic regions.[21, 22]

The study of chromatin conformation has been requisite in the analysis of such non-coding rearrangements. DNA is organized in the three-dimensional nucleus at varying hierarchical levels that are important for the regulation of gene expression,[32] with primary roles in embryonic development and disease.[82] Several studies have analyzed the impact of structural variants in disruption of the regulatory chromatin environment leading to disease;[41, 42, 44-46, 48] these studies have set the precedent for integrative analyses of disrupted chromatin conformation to expedite functional annotations of non-coding chromosome rearrangements.

We tested the possibility of utilizing chromatin contact information to dissect chromosome rearrangements which disrupt non-coding chromosome regions in clinical cases.

We focused on 17 subjects from DGAP, 12 with available clinical microarray information, with different rare presentations and *de novo* non-coding BCAs classified as VUS. Of these, 15 corresponded to translocations and two were inversions. These cases represent ~11% of the total number of sequenced DGAP cases, which makes our predictions even more significant for future potential treatment or management of subjects who would not otherwise obtain a clinical diagnosis. Utilizing publicly available annotated genomic and regulatory elements, chromatin conformation capture information, predicted enhancer-promoter interactions, phenomatch scores, as well as haploinsufficiency and triplosensitivity information for all genes surrounding the BCA breakpoints at different window sizes (±3 and ±1 Mb as well as BCA-containing TAD positions), we discovered 16 genes for 11 DGAP cases that are top-ranking position effect candidates for the subjects' clinical phenotypes (Table 1).

We observed that eight of the sequenced DGAP BCA breakpoints, corresponding to six DGAP cases (DGAP017, 176, 249, 275, 288 and 322), overlapped reported annotated and predicted enhancers and DHS sites. Disruption of these regulatory elements could potentially cause improper gene expression or repression through altered enhancer-promoter interactions or interactions with other DHS-associated elements such as insulators and locus control regions, among others. In fact, four of the breakpoints that disrupt annotated DHS sites and enhancers have been shown to establish chromatin contacts with our top position effect candidate genes in the region in Hi-C data of H1-hESC cells at 40 Kb resolution (Table S18). For example, the DGAP275_B breakpoint is involved in a chromatin interaction that puts it into physical proximity with *POLE* and *ANKLE2,* DGAP288_B contacts *SOX9*, and DGAP176_B interacts with *ACSL4*. Three additional breakpoints from DGAP111, 249 and 287 overlap CTCF binding sites. CTCF binding sites are enriched in TAD boundaries,[35] and the elimination of these binding

sites could potentially induce gene expression or other functional changes through alteration of the structural regulatory landscape of the region.[41]

There are nine DGAP cases (DGAP113, 126, 138, 153, 163, 252, 315, 319 and 329), six with normal arrays and two with benign CNVs, for which no overlap with genomic or other regulatory elements was detected. These cases thus represent events in which position effects are most likely caused by alteration of the underlying chromatin structure itself. This hypothesis is supported by detection of a vast number of disrupted chromatin contacts in four different cell lines (H1-hESC, IMR90, GM06990, GM12878) at different Hi-C window resolutions, 32 breakpoints included in H1-hESC TADs,[35] and the separation of 193 genes from one and up to 91 of their predicted enhancers after the occurrence of the BCAs (Table S14). For example, *SOS1*, one of the most significant candidates in explaining DGAP163's global developmental delay, dysmorphic/distinctive facies and hearing loss, as observed in Noonan Syndrome 1 (NS1 [MIM: 163950]), is separated from its interaction with 88 predicted enhancers (Figure 3), and exhibited a decrease in expression in DGAP163-derived LCLs. However, NS1 is caused by autosomal dominant mutations in *SOS1*; we hypothesize that the reduced expression of *SOS1* might affect the RAS/MAPK signaling pathway and generate clinical features not completely overlapping those of NS1; however, this possibility remains to be functionally tested and complemented with analyses of genomic single nucleotide variants. A similar approach could be explored for DGAP275, where we hypothesize that *POLE*, associated with the facial dysmorphism, immunodeficiency, livedo, and short stature syndrome (FILS [MIM: 615139]) in an autosomal recessive manner,[83] may contribute to the extreme short stature observed in this DGAP subject; and *ZEB2*, etiologic for Mowat-Wilson syndrome (MOWS [MIM: 235730]) in an autosomal dominant manner (OMIM#235730), may potentially explain the hypotonia and

neurological features observed in DGAP329 but not present all of the dysmorphic features or medical/non-neurologic phenotype of MOWS. Overall, more candidate genes will need to be analyzed rigorously to assess the validity of our position effect predictions and the disruption of important chromatin regulatory elements. Nonetheless, insight into the molecular pathway of disorders may be forthcoming from our approach and of value in the management of some individuals.

All predicted candidate genes have different lines of evidence supporting their selection, starting with a significant phenomatch score that correlates annotated gene phenotypes to those observed in the DGAP cases. HI and triplosensitivity evidence, inclusion in TAD regions, as well as HI scores build upon this selection, and can help laboratories and clinicians focus in subsequent analyses on candidates of their interest. As of now, the "top-ranking" candidates have the highest number of evidence supporting their selection; however, there are also 102 second-tier candidates for the 17 analyzed DGAP cases within ±1 Mb analysis windows which may well play a functional role. Presently, we are unable to give "weights" to any of these selection criteria (*i.e.*, a gene with a high phenomatch score and no evidence of HI is "more significant" than a gene with a medium phenomatch score and evidence of HI) mainly for two reasons: (i) we would need to collect more examples, which might not be easy to find and require a tremendous curation effort, and (ii) we need to understand the possibility, suggested by our results, that more than one gene may be contributory in the clinical presentation of the DGAP subjects, either acting simultaneously or throughout development. Moreover, many of the candidates have recessive inheritance modes, which make it necessary to assess the mutational status of both alleles as well as additional sequence variants not captured by our BCA breakpoint sequencing and the microarrays. Future in-depth exome, DNA and RNA sequencing as well as Hi-C

experiments will provide a comprehensive view of the contribution of sequence variants, disruption of chromatin contacts, and changes in gene expression in the DGAP disease etiologies, such that guidelines might be developed as to which candidates should be followed up first and further studied with comprehensive functional validation using animal models and human cell lines that reproduce the BCA breakpoints.

Overall our results suggest that the integration of phenomatch scores, altered chromatin contacts, and other clinical gene annotations provide valuable interpretation to many variants of uncertain significance through long-range position effects. The correct prediction of 52 out of 57 known pathogenic genes in DGAP cases used as positive controls supports such integration. Our computational analysis is rapid and can provide additional information to benefit the clinical assessment of both coding and non-coding genome variants. The latter is an important step towards prediction of pathogenic consequences of non-coding variantion observed in prenatal samples. For example, based on its position and chromatin contact alterations, we correctly predicted the involvement and decreased expression of *SOX9* in the cleft palate Pierre-Robin sequence (PRBNS [MIM: 261800]) association in DGAP288.[49]

Lastly, we would like to note that predicting the pathogenic outcome of disrupted chromatin contacts is not a straightforward endeavor: it has been shown that a single gene promoter can be targeted by several enhancers,[63] therefore compensating for the perturbed interactions by the chromosome rearrangements. In addition, rearrangements can reposition gene promoters and enhancers outside of their preferred chromatin environments, leading to improper gene activation by enhancer adoption.[41] Our method currently identifies instances in which known and predicted enhancer/promoter interactions are disrupted by the rearrangement breakpoints and thus lead to decreased candidate gene expression. Enhancer adoption prediction

will be incorporated once mathematical models of TAD formation upon changes in genomic sequence are refined and available to the greater scientific community. Presently, our predictions are as good as the availability of pathogenic gene annotations, chromatin conformation data, clinical phenotype information, and the presence of similar rearrangements in databases such as DECIPHER and dbVar. While the existence of other subjects with related phenotypes to the DGAP cases does not prove the involvement of neighboring genes in the etiology of these phenotypes, it is a step forward towards prediction of pathogenic effects starting from a simple computational analysis, pointing to a better phenotypic categorization when clinically examining affected individuals. By making our position effect prediction method available to the human genetics community, we hope to study additional cases with complete phenotypic information and be able to refine better the rules for the prediction of position effects on gene expression and discover new mechanisms of pathogenicity.

Acknowledgements

We offer heartfelt gratitude to all DGAP research participants and their families, and to countless genetic counselors, clinical geneticists, cytogeneticists, and physicians for their ongoing support of our study and for referrals to our project. This study was funded by the National Institutes of Health (GM061354 to CCM and MET). The authors declare no conflicts of interest.

Tables

Table 1. Description of the 17 analyzed DGAP cases with non-coding BCAs. Corresponding clinical karyotypes are reported, with overlap of breakpoints with regulatory elements (E = enhancer, DHS = DNaseI hypersensitive sites, CTCF = CTCF binding sites), and TADs from H1-hESC, IMR90, and GM12878 (1= one breakpoint within TAD, 2=both BCA breakpoints are located within TAD). Top-ranking position effect genes are provided for the ±1 Mb windows surrounding the BCA breakpoints; each gene is highlighted with different evidence supporting its inclusion (a = ClinGen known recessive genes, b= ClinGen genes with emerging and sufficient evidence suggesting haploinsufficiency is associated with clinical phenotype, c = HI scores less than 10, d = within H1-ESC TAD, e = DHS enhancer-promoter disrupted interactions).

Subject ID	Reported Karyotype	Disruption of Functional Element	Breakpoints within TADs			Top-ranking Candidates ±1 Mb
			hESC	IMR90	GM12878	
DGAP017	46,X,t(X;10)(p11.2;q24.3)	DHS	2	2	1	-
DGAP111	46,XY,t(16;20)(q11.2;q13.2)dn	CTCF	1	1	2	*ORC6*[a]
DGAP113	46,XY,t(1;3)(q32.1;q13.2)dn	-	2	2	2	*ASPM*[a]
DGAP126	46,XX,t(5;10)(p13.3;q21.1)dn	-	2	1	2	-
DGAP138	46,XY,t(1;6)(q23;q13)dn	-	2	2	2	*GRIK2*[ac]
DGAP153	46,X,t(X;17)(p11.23;p11.2)dn	-	1	1	1	-
DGAP163	46,XY,t(2;14)(p23;q13)dn	-	2	2	2	*SOS1*[cde], *COCH*[de]
DGAP176	46,Y,inv(X)(q13q24)mat	DHS, CTCF	2	1	2	*ACSL4*[bd], *COL4A5*[bcde]
DGAP249	46,XX,t(2;11)(q33;q23)dn	E, DHS	2	2	2	*SATB2*[bcde], *SORL1*[e]
DGAP252	46,XY,t(3;18)(q13.2;q11.2)dn	-	2	2	2	*RBBP8*[a],*GATA6*[bcde]
DGAP275	46,XX,t(7;12)(p13;q24.33)dn	DHS	1	1	2	*ANKLE2*[e], *POLE*[e]
DGAP287	46,XY,t(10;14)(p13;q32.1)dn	CTCF	2	2	2	-
DGAP288	46,XX,t(6:17)(q13;q21)dn	DHS	2	2	2	*SOX9*[bcd]
DGAP315	46,XX,inv(6)(p24q11)dn	-	1	1	2	-
DGAP319	46,XX,t(4;13)(q31.3;q14.3)dn	-	2	1	2	-
DGAP322	46,XY,t(1;18)(q32.1;q22.1)	DHS	1	2	2	*IRF6*[bcd]
DGAP329	46,XX,t(2;14)(q21;q24.3)dn	-	1	2	2	*ZEB2*[bcde]

Figures

Figure 1. Chromosome locations of the 17 analyzed DGAP cases with non-coding BCAs. Breakpoint positions are marked with a blue line and the corresponding DGAP number. All chromosomes are aligned by the centromere (marked in pink) and are indicated above by their corresponding chromosome number.

Figure 2. Assessment of gene expression changes for DGAP163-derived LCLs. Each column represents the ΔCT results of three culture replicates, with four technical replicates each, compared to three sex-matched control cell lines. Error bars indicate the standard deviation calculated from the biological replicates. The Mann-Whitney U test p-value is provided for the comparison between expression values of *SOS1* and the control *GUSB*.

Figure 3. Disrupted enhancer-promoter DHS interactions predicted for *SOS1* (gene position indicated by asterisk). The color graded rectangle represents the correlation values for the interactions as reported by ENCODE. The dashed line indicates the translocation breakpoint position in chromosome 2. Lilac colored rectangles represent genes, and pink rectangles show TAD positions annotated in H1-hESC.

Web Resources

The scripts used in this study to predict position effects can be downloaded from: https://github.com/ibn-salem/position_effect

OMIM, http://www.omim.org

Ensembl GRCh37 archive, http://grch37.ensembl.org

Human lincRNAs catalog, http://portals.broadinstitute.org/genome_bio/human_lincrnas

Haploinsufficiency scores, https://decipher.sanger.ac.uk

ClinGen GRCh37 data, ftp://ftp.ncbi.nlm.nih.gov/pub/dbVar/clingen

University of California Santa Cruz Genome Browser, https://genome.ucsc.edu

Human Phenotype Ontology, http://human-phenotype-ontology.github.io

Harvard Biopolymers Facility, https://genome.med.harvard.edu

dbVar Variation Viewer, https://www.ncbi.nlm.nih.gov/variation/view

3D Genome Browser, http://promoter.bx.psu.edu/hi-c

ENCODE, https://www.encodeproject.org

WashU EpiGenome Browser, http://epigenomegateway.wustl.edu/

GTEx portal, https://www.gtexportal.org/home

References

1. Lejeune, J., Gautier, M., and Turpin, R. (1959). [Study of somatic chromosomes from 9 mongoloid children]. Comptes rendus hebdomadaires des seances de l'Academie des sciences 248, 1721-1722.
2. Ford, C.E., Jones, K.W., Polani, P.E., De Almeida, J.C., and Briggs, J.H. (1959). A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). Lancet 1, 711-713.
3. Jacobs, P.A., and Strong, J.A. (1959). A case of human intersexuality having a possible XXY sex-determining mechanism. Nature 183, 302-303.
4. Stankiewicz, P., and Lupski, J.R. (2002). Genome architecture, rearrangements and genomic disorders. Trends in genetics : TIG 18, 74-82.
5. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nat Genet 36, 949-951.
6. Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. Science 305, 525-528.
7. Hinds, D.A., Kloek, A.P., Jen, M., Chen, X., and Frazer, K.A. (2006). Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38, 82-85.
8. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E., and Pritchard, J.K. (2006). A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38, 75-81.
9. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al. (2010). Origins and functional impact of copy number variation in the human genome. Nature 464, 704-712.
10. Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., et al. (2007). Paired-end mapping reveals extensive structural variation in the human genome. Science 318, 420-426.
11. McCarroll, S.A., Hadnott, T.N., Perry, G.H., Sabeti, P.C., Zody, M.C., Barrett, J.C., Dallaire, S., Gabriel, S.B., Lee, C., Daly, M.J., et al. (2006). Common deletion polymorphisms in the human genome. Nat Genet 38, 86-92.
12. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature 444, 444-454.
13. Genomes Project, C., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature 467, 1061-1073.
14. Carvalho, C.M., and Lupski, J.R. (2016). Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 17, 224-238.
15. Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10, 451-481.
16. Theisen, A., and Shaffer, L.G. (2010). Disorders caused by chromosome abnormalities. Appl Clin Genet 3, 159-174.
17. Nambiar, M., and Raghavan, S.C. (2011). How does DNA break during chromosomal translocations? Nucleic acids research 39, 5813-5825.
18. Higgins, A.W., Alkuraya, F.S., Bosco, A.F., Brown, K.K., Bruns, G.A., Donovan, D.J., Eisenman, R., Fan, Y., Farra, C.G., Ferguson, H.L., et al. (2008). Characterization of apparently

balanced chromosomal rearrangements from the developmental genome anatomy project. Am J Hum Genet 82, 712-722.
19. Kleinjan, D.A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76, 8-32.
20. Weiler, K.S., and Wakimoto, B.T. (1995). Heterochromatin and gene expression in Drosophila. Annu Rev Genet 29, 577-605.
21. Zhang, F., and Lupski, J.R. (2015). Non-coding genetic variants in human disease. Hum Mol Genet 24, R102-110.
22. Spielmann, M., and Mundlos, S. (2016). Looking beyond the genes: the role of non-coding variants in human disease. Hum Mol Genet.
23. Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M., et al. (1995). Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum Mol Genet 4, 415-422.
24. Kleinjan, D.A., Seawright, A., Schedl, A., Quinlan, R.A., Danes, S., and van Heyningen, V. (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. Hum Mol Genet 10, 2049-2059.
25. Cai, J., Goodman, B.K., Patel, A.S., Mulliken, J.B., Van Maldergem, L., Hoganson, G.E., Paznekas, W.A., Ben-Neriah, Z., Sheffer, R., Cunningham, M.L., et al. (2003). Increased risk for developmental delay in Saethre-Chotzen syndrome is associated with TWIST deletions: an improved strategy for TWIST mutation screening. Hum Genet 114, 68-76.
26. Flomen, R.H., Vatcheva, R., Gorman, P.A., Baptista, P.R., Groet, J., Barisic, I., Ligutic, I., and Nizetic, D. (1998). Construction and analysis of a sequence-ready map in 4q25: Rieger syndrome can be caused by haploinsufficiency of RIEG, but also by chromosome breaks approximately 90 kb upstream of this gene. Genomics 47, 409-413.
27. Trembath, D.G., Semina, E.V., Jones, D.H., Patil, S.R., Qian, Q., Amendt, B.A., Russo, A.F., and Murray, J.C. (2004). Analysis of two translocation breakpoints and identification of a negative regulatory element in patients with Rieger's syndrome. Birth Defects Res A Clin Mol Teratol 70, 82-91.
28. Velagaleti, G.V., Bien-Willner, G.A., Northup, J.K., Lockhart, L.H., Hawkins, J.C., Jalal, S.M., Withers, M., Lupski, J.R., and Stankiewicz, P. (2005). Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. Am J Hum Genet 76, 652-662.
29. Kleinjan, D.J., and van Heyningen, V. (1998). Position effect in human genetic disease. Hum Mol Genet 7, 1611-1618.
30. Lupski, J.R., and Stankiewicz, P. (2005). Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet 1, e49.
31. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science 295, 1306-1311.
32. de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev 26, 11-24.
33. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-293.
34. Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 462, 58-64.

35. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-380.
36. Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature 489, 109-113.
37. Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 153, 1281-1295.
38. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665-1680.
39. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47, 598-606.
40. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature 518, 331-336.
41. Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161, 1012-1025.
42. Groschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A., Erpelinck, C., van der Velden, V.H., Havermans, M., Avellino, R., van Lom, K., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 157, 369-381.
43. Claussnitzer, M., Dankel, S.N., Kim, K.H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., et al. (2015). FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med 373, 895-907.
44. Visser, M., Kayser, M., and Palstra, R.J. (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res 22, 446-455.
45. Roussos, P., Mitchell, A.C., Voloudakis, G., Fullard, J.F., Pothula, V.M., Tsang, J., Stahl, E.A., Georgakopoulos, A., Ruderfer, D.M., Charney, A., et al. (2014). A role for noncoding variation in schizophrenia. Cell Rep 9, 1417-1429.
46. Giorgio, E., Robyr, D., Spielmann, M., Ferrero, E., Di Gregorio, E., Imperiale, D., Vaula, G., Stamoulis, G., Santoni, F., Atzori, C., et al. (2015). A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). Hum Mol Genet 24, 3143-3154.
47. Oldridge, D.A., Wood, A.C., Weichert-Leahey, N., Crimmins, I., Sussman, R., Winter, C., McDaniel, L.D., Diamond, M., Hart, L.S., Zhu, S., et al. (2015). Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. Nature 528, 418-421.
48. Ibn-Salem, J., Kohler, S., Love, M.I., Chung, H.R., Huang, N., Hurles, M.E., Haendel, M., Washington, N.L., Smedley, D., Mungall, C.J., et al. (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome Biol 15, 423.
49. Ordulu, Z., Kammin, T., Brand, H., Pillalamarri, V., Redin, C.E., Collins, R.L., Blumenthal, I., Hanscom, C., Pereira, S., Bradley, I., et al. (2016). Structural Chromosomal Rearrangements Require Nucleotide-Level Resolution: Lessons from Next-Generation Sequencing in Prenatal

Diagnosis. Am J Hum Genet 99, 1015-1033.
50. Ligon, A.H., Moore, S.D., Parisi, M.A., Mealiffe, M.E., Harris, D.J., Ferguson, H.L., Quade, B.J., and Morton, C.C. (2005). Constitutional rearrangement of the architectural factor HMGA2: a novel human phenotype including overgrowth and lipomas. Am J Hum Genet 76, 340-348.
51. Kim, H.G., Kishikawa, S., Higgins, A.W., Seong, I.S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K., et al. (2008). Disruption of neurexin 1 associated with autism spectrum disorder. Am J Hum Genet 82, 199-207.
52. Lu, W., Quintero-Rivera, F., Fan, Y., Alkuraya, F.S., Donovan, D.J., Xi, Q., Turbe-Doan, A., Li, Q.G., Campbell, C.G., Shanske, A.L., et al. (2007). NFIA haploinsufficiency is associated with a CNS malformation syndrome and urinary tract defects. PLoS genetics 3, e80.
53. Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.A., et al. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet 49, 36-45.
54. Talkowski, M.E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., Kirby, A., Liu, S., Muddukrishna, B., Ohsumi, T.K., et al. (2011). Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. Am J Hum Genet 88, 469-481.
55. Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17, 47-62.
56. Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., and Carter, D.R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17, 792-798.
57. Muro, E.M., and Andrade-Navarro, M.A. (2010). Pseudogenes as an alternative source of natural antisense transcripts. BMC evolutionary biology 10, 338.
58. Ordulu, Z., Wong, K.E., Currall, B.B., Ivanov, A.R., Pereira, S., Althari, S., Gusella, J.F., Talkowski, M.E., and Morton, C.C. (2014). Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. Am J Hum Genet 94, 695-709.
59. Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. Nucleic acids research 42, D749-755.
60. Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25, 1915-1927.
61. Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and predicting haploinsufficiency in the human genome. PLoS genetics 6, e1001154.
62. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen--the Clinical Genome Resource. N Engl J Med 372, 2235-2242.
63. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature 489, 75-82.
64. Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 347, 1017-1021.
65. Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suva, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene

activation in IDH mutant gliomas. Nature 529, 110-114.
66. Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science 351, 1454-1458.
67. Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74.
68. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome research 12, 996-1006.
69. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature 507, 455-461.
70. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35, D88-92.
71. Zhou, X., and Wang, T. (2012). Using the Wash U Epigenome Browser to examine genome-wide sequencing data. Current protocols in bioinformatics Chapter 10, Unit10.10.
72. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.
73. Heger, A., Webber, C., Goodson, M., Ponting, C.P., and Lunter, G. (2013). GAT: a simulation framework for testing the association of genomic intervals. Bioinformatics (Oxford, England) 29, 2046-2048.
74. Kohler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C., Brown, D.L., Brudno, M., Campbell, J., et al. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic acids research 42, D966-974.
75. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 84, 524-533.
76. Lappalainen, I., Lopez, J., Skipper, L., Hefferon, T., Spalding, J.D., Garner, J., Chen, C., Maguire, M., Corbett, M., Zhou, G., et al. (2013). DbVar and DGVa: public archives for genomic structural variation. Nucleic acids research 41, D936-941.
77. Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics 1, 4.
78. Cardoso, A.R., Oliveira, M., Amorim, A., and Azevedo, L. (2016). Major influence of repetitive elements on disease-associated copy number variants (CNVs). Hum Genomics 10, 30.
79. Consortium, G.T. (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580-585.
80. Vetro, A., Dehghani, M.R., Kraoua, L., Giorda, R., Beri, S., Cardarelli, L., Merico, M., Manolakos, E., Parada-Bustamante, A., Castro, A., et al. (2015). Testis development in the absence of SRY: chromosomal rearrangements at SOX9 and SOX3. Eur J Hum Genet 23, 1025-1032.
81. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 86, 749-764.
82. Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat Rev Genet 17, 661-678.

83. Pachlopnik Schmid, J., Lemoine, R., Nehme, N., Cormier-Daire, V., Revy, P., Debeurme, F., Debre, M., Nitschke, P., Bole-Feysot, C., Legeai-Mallet, L., et al. (2012). Polymerase epsilon1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome"). J Exp Med 209, 2323-2330.