Computational Prediction of Position Effects of Apparently Balanced Human Chromosomal Rearrangements

Cinthya J. Zepeda-Mendoza,[1,2,25] Jonas Ibn-Salem,[3,25] Tammy Kammin,[1] David J. Harris,[2,4] Debra Rita,[5] Karen W. Gripp,[6] Jennifer J. MacKenzie,[7] Andrea Gropman,[8] Brett Graham,[9] Ranad Shaheen,[10] Fowzan S. Alkuraya,[10,11] Campbell K. Brasington,[12] Edward J. Spence,[12] Diane Masser-Frye,[13] Lynne M. Bird,[13,14] Erica Spiegel,[15] Rebecca L. Sparkes,[16] Zehra Ordulu,[17] Michael E. Talkowski,[17,18,19,20,21] Miguel A. Andrade-Navarro,[3] Peter N. Robinson,[22] and Cynthia C. Morton[1,3,20,23,24,*]

Interpretation of variants of uncertain significance, especially chromosomal 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 three-dimensional chromosomal structural data and transcriptional regulatory information to predict position effects and their association with pathogenic phenotypes in 17 subjects with apparently balanced chromosomal abnormalities. We found 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 analysis of position effect. These results highlight the important interplay between chromosomal structure and disease and demonstrate the need to utilize chromatin conformational data for the prediction of position effects in the clinical interpretation of non-coding chromosomal rearrangements.

Introduction

The importance of the integrity of chromosomal 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 catalog 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 chromosomal 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 chromosomal abnormalities (BCAs), such as translocations and inversions, often found outside of the hypothesized disease-related genes.[19]

Position effects were first identified in *Drosophila melanogaster*, in which chromosomal inversions placing *white*+ near centric heterochromatin caused mosaic red and 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.[19,21,22] Examples of position-effect genes include paired box gene 6 (*PAX6* [MIM: 607108]), for which downstream chromosomal translocations affect its *cis*-regulatory control and produce aniridia (MIM: 106210);[23,24] twist family bHLH transcription factor 1 (*TWIST1* [MIM: 601622]), where downstream translocations and inversions are associated with Saethre-Chotzen syndrome (MIM: 101400);[25]

[1]Department of Obstetrics, Gynecology, and Reproductive Biology, Brigham and Women's Hospital, Boston, MA 02115, USA; [2]Harvard Medical School, Boston, MA 02115, USA; [3]Johannes Gutenberg University, Mainz 55122, Germany; [4]Boston Children's Hospital, Boston, MA 02115, USA; [5]Cytogenetics Lab, ACL laboratories, Rosemont, IL 60018, USA; [6]Nemours Alfred I. DuPont Hospital for Children, Wilmington, DE 19803, USA; [7]Department of Pediatrics, McMaster University, Hamilton, ON L8S 4L8, Canada; [8]Children's National Medical Center, Washington, DC 20010, USA; [9]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [10]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 12713, Saudi Arabia; [11]Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh 11533, Saudi Arabia; [12]Clinical Genetics Division, Department of Pediatrics, Levine Children's Hospital at Carolinas Medical Center, Charlotte, NC 28203, USA; [13]Genetics and Dysmorphology, Rady Children's Hospital San Diego, San Diego, CA 92123, USA; [14]University of California, San Diego, La Jolla, CA 92093, USA; [15]Maternal Fetal Medicine, Columbia University Medical Center, New York, NY 10032, USA; [16]Department of Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada; [17]Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA; [18]Departments of Neurology and Psychiatry and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; [19]Department of Neurology, Harvard Medical School, Boston, MA 02115, USA; [20]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [21]Stanley Center for Psychiatric Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [22]Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA; [23]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA; [24]Division of Evolution and Genomic Science, School of Biological Sciences, Manchester Academic Health Science Centre, Manchester M13 9NT, UK
[25]These authors contributed equally to this work
*Correspondence: cmorton@bwh.harvard.edu
http://dx.doi.org/10.1016/j.ajhg.2017.06.011.

paired-like homeodomain 2 (*PITX2* [MIM: 601542]), for which translocations are associated with Axenfeld-Rieger syndrome type 1 (MIM: 180500);[26,27] and SRY-box 9 (*SOX9* [MIM: 608160]), where translocation breakpoints located up to 900 kb upstream and 1.3 Mb downstream are associated with campomelic dysplasia (MIM: 114290);[28] as well as 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,[31,32] 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 chromosomal 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 chromosomal rearrangements, we analyzed 17 subjects with de novo non-coding BCAs classified as variants of uncertain significance (VUSs) from the Developmental Genome Anatomy Project (DGAP).[18,50–53] 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 misregulated expression in DGAP-derived lymphoblastoid cell lines (LCLs). These results suggest that many VUSs are likely to be further interpretable via long-range effects and warrant routine assessment and integration in clinical diagnosis.

Material and Methods

Selection of Subjects with BCAs

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

Clinical Descriptions of DGAP Subjects

The clinical presentation of the 17 subjects ranged 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 according to the International System for Human Cytogenomic Nomenclature (ISCN2016) (Table 1). Detailed clinical descriptions, as well as nomenclature developed to describe chromosomal rearrangements by using next-generation sequencing,[58] are included in the Supplemental Note: Case Reports. Reported ages of DGAP subjects are from the time of enrollment. All reported genomic coordinates reference GRCh37.

Analysis of Genes Bordering the Rearrangement Breakpoints

The presence of annotated genes or pseudogenes and lincRNAs was assessed in ± 3 and ± 1 Mb windows neighboring each subject's translocation and inversion breakpoints and within the reported H1-hESC topologically associated domains (TADs)[35] where the breakpoints were located. The gene annotation file was obtained from Ensembl GRCh37,[59] and we used the Human lincRNA Catalog.[60] Haploinsufficiency (HI) and triplosensitivity scores were assigned according to the report by Huang et al.[61] and version hg19 of ClinGen[62] data downloaded on September 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 changes in gene expression; DNase I hypersensitive sites (DHSs) 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] On the basis of 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 ENCODE project[67] data on CTCF binding sites, DHSs, and chromatin segmentation classifications (Broad ChromHMM) derived from a LCL (GM12878) and human stem cells (H1-hESC) accessed through the UCSC Genome Browser.[68] Enhancer positions were additionally obtained from Andersson et al.[69] for tissue and primary cells and from the VISTA Enhancer Browser version hg19.[70] Finally, lists of transcription factor (TF) binding sites and gene promoters were obtained from Ensembl 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.[35] and the WashU EpiGenome Browser.[71] A high-resolution dataset of chromatin loops and domains was obtained from Rao et al.[38] for IMR90 and GM12878 cells. Lastly, we used distal DHS/enhancer-promoter connections[63] (DHSs that could be candidate enhancers given their association with distal promoters) to assess disrupted predicted *cis*-regulatory interactions by the BCAs. Genomic overlaps between the rearrangement breakpoints, functional elements, and disrupted chromatin

Table 1. Description of the 17 Analyzed DGAP Subjects with Non-coding BCAs

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*[a,c]
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*[c,d,e] and *COCH*[d,e]
DGAP176	46,Y,inv(X)(q13q24)mat	DHS, CTCF	2	1	2	*ACSL4*[b,d] and *COL4A5*[b,c,d,e]
DGAP249	46,XX,t(2;11)(q33;q23)dn	E, DHS	2	2	2	*SATB2*[b,c,d,e] and *SORL1*[e]
DGAP252	46,XY,t(3;18)(q13.2;q11.2)dn	–	2	2	2	*RBBP8*[a] and *GATA6*[b,c,d,e]
DGAP275	46,XX,t(7;12)(p13;q24.33)dn	DHS	1	1	2	*ANKLE2*[e] and *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*[b,c,d]
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*[b,c,d]
DGAP329	46,XX,t(2;14)(q21;q24.3)dn	–	1	2	2	*ZEB2*[b,c,d,e]

Corresponding clinical karyotypes, including overlap between breakpoints and regulatory elements (E, enhancer; DHS, DNaseI hypersensitive site; CTCF, CTCF binding site), and TADs from H1-hESC, IMR90, and GM12878 (1, one breakpoint within the TAD; 2, both BCA breakpoints are located within the TAD) are reported. 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 (see footnotes).
[a]ClinGen known recessive genes.
[b]ClinGen genes with emerging and sufficient evidence suggesting that HI is associated with clinical phenotype.
[c]HI scores less than 10.
[d]Within H1-ESC TAD.
[e]Disrupted DHS/enhancer-promoter interactions.

interactions were calculated with custom Perl scripts, the BEDtools suite,[72] and the Genomic Association Tester (GAT).[73]

Ontological Analysis of Genes Neighboring Breakpoints

We calculated phenotype similarity between potential position-effect genes and DGAP subjects 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.[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 with DGAP subjects and those 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 the significance of this similarity. The rest was based on randomly permuting the associations of phenotypes to genes; to this effect, the phenotype-gene associations were shuffled 100 times randomly, and the similarity between these random phenotypes and the studied clinical findings was calculated. The second background control was based on shifting the breakpoint location along the chromosome; each breakpoint was shifted by −9, −6, −3, +3, +6, and +9 Mb, and the similarity of genes in proximity to the shifted breakpoints was computed.

Real-Time qPCR

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

Assessment of DGAP Breakpoints Overlapping Non-coding Structural Variants in Public Databases

To find similar subjects with non-coding structural rearrangements and compare their annotated clinical phenotypes with those observed in DGAP subjects, we searched DECIPHER[75]

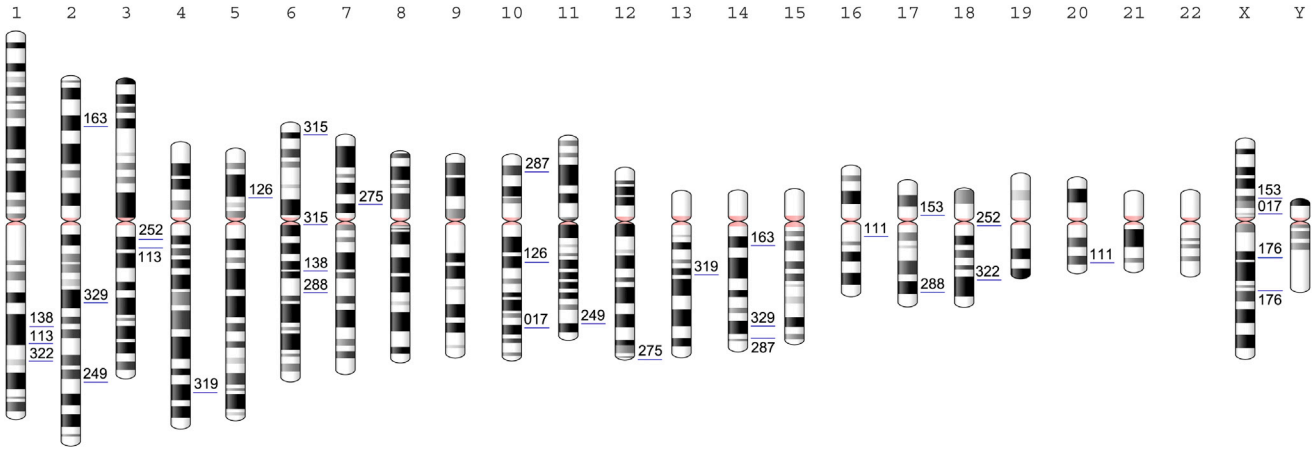

Figure 1. Chromosome Locations of the 17 Analyzed DGAP Subjects with Non-coding BCAs
Breakpoint positions are marked with a blue line and the corresponding DGAP number. All chromosomes are aligned by the centromere (marked in pink) and are indicated above by their corresponding chromosome number.

version 2015-07-13, as well as dbVar from the NCBI Variation Viewer 1.5.[76] Both databases are comprehensive community-supported repositories of clinical subjects 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 chromosomal rearrangements from individuals with abnormal developmental and cognitive phenotypes; many of these have yet to be investigated in detail. Each studied DGAP BCA has two breakpoint positions (because 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 subjects with both breakpoints in non-coding regions only and exclude lincRNAs and pseudogenes; a total of 17 subjects, 15 with translocations and 2 with inversions, fulfilled our criteria (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 of overlapping annotated promoters and TF binding sites (GAT TF $p = 0.0003$, promoter $p = 0.0001$; Tables S2 and S3). 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 H1-hESC (GAT $p = 0.0036$) but not for the annotated tissue or primary cell enhancers from Andersson et al.[69] (Table S4). Eight breakpoints overlapped cell-type-specific DHSs (Tables 1 and S5); these corresponded to DGAP subjects DGAP017, DGAP176, DGAP249, DGAP275, DGAP288, and DGAP322. Of these, DGAP176 and DGAP275 breakpoints overlapped DHSs at both BCA breakpoint sites. In addition, three DGAP subjects had rearrangements overlapping CTCF binding sites in H1-hESC (DGAP111, DGAP176, and DGAP287) and none in GM12878 cells (Table 1 and Table S6). Except for those in two subjects in H1-hESC (DGAP17 and DGAP176) and four subjects in GM12878 (DGAP017, DGAP126, DGAP163, and DGAP176), all rearrangements fell 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%) overlapped repeated elements at a significant level (GAT $p = 0.0002$; Table S8), which could 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 fell within previously reported TADs in H1-hESC and IMR90 cell lines (Tables 1 and S9).[35] However, this overlap was not significant for either cell line (GAT H1-hESC $p = 0.0537$ and IMR90 $p = 0.28$). We found that the breakpoints disrupted dozens, hundreds, or even thousands of chromatin contacts when they were assessed at 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 consistently lacked disrupted chromatin contacts, which is expected because it overlaps a repetitive satellite region, so no chromatin contacts could be mapped to the segment (Tables S9 and 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 of 34 breakpoints were

contained within GM12878 sub-compartments (Tables 1 and S10); interestingly, 28 of these were classified as members of the B compartment, which is less gene dense and less expressed than the A compartment. On the other hand, 18 and 24 breakpoints were 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 disrupted several significant short and long-range chromatin interactions in the GM12878 Hi-C data (Table S12).

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

Identification of Genes with Potential Position Effects

To identify genes that could be generating the complex DGAP phenotypes via position effects from chromosomal 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 3,081 genes were contained within the ±3 and ±1 Mb windows for all subjects; 106 of these genes (~3.4%) had an HI score of <10%, which is a predictor of HI,[61] and 55 and 2 genes had ClinGen emerging evidence suggesting that dosage HI and triplosensitivity, respectively, are associated with clinical phenotypes (Table S15).

To further refine our search for genes that might exhibit position effects, we performed an unbiased correlation between DGAP subjects' 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 subjects and neighboring genes by 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 phenotypes of the DGAP subject and neighboring gene. The similarity significance is then calculated on the basis of randomly permuting the associations between phenotypes and genes and on 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 obtaining information on dosage sensitivity and phenotypic similarity, we complemented our analysis with assessment of enhancer-promoter interactions to make our candidate selection more specific. A typical mechanism by which chromosomal rearrangements cause position effects is through disruptions in the association between genes and 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, whereas 60 genes were separated from their predicted interactions with 10 or up to 91 enhancers (Table S14).

For the 17 analyzed DGAP BCAs, 645 genes had 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 according to their reported dosage sensitivity and disrupted regulatory interactions and selected a phenomatch cutoff value capable of detecting pathogenic and likely pathogenic genes in 57 published DGAP subjects from Redin et al.[53] By accounting for the top quartile values of the reported phenomatch scores per subject and adding up their data on dosage sensitivity and disrupted regulatory interactions, we consistently ranked the reported pathogenic and likely pathogenic genes in the upper decile for 52 of the 57 DGAP control subjects (~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 subject, whereas 19 of them were positioned in the second-tier rank. Only five genes could not be found in the top decile positions, because 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 subjects and 102 second-tier candidates for the 17 analyzed DGAP subjects within ±1 Mb analysis windows (Tables 1 and S15). This is a significant reduction in comparison with the initial 645 possible candidates (~3.8% of the neighboring genes in the ±3 Mb windows when top and second-tier candidates are considered and 0.5% when only top candidates are considered). Of note, only 9 of the 16 top-ranking candidates were included within the same TAD as the BCA breakpoint (H1-hESC TADs from Dixon et al.[35]), whereas the rest were located farther away. Nine top-ranking genes had an HI score < 10%,[61] whereas ClinGen HI data revealed that 4 of these 16 genes are associated with autosomal-recessive phenotypes, and an additional seven have sufficient or some evidence of

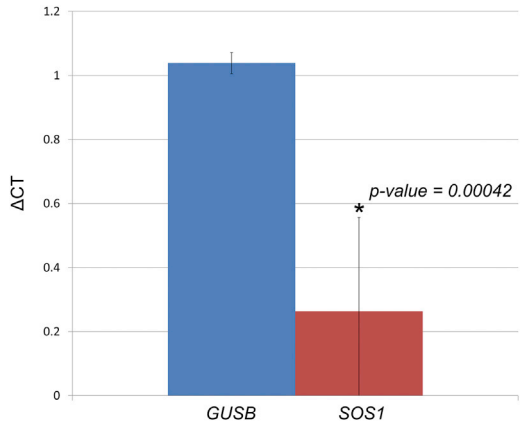

Figure 2. Assessment of Gene-Expression Changes for DGAP163-Derived LCLs

Each column compares the ΔCt results of three culture replicates (with four technical replicates each) with those of 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*.

HI. 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).

Together, these cases represent more plausible candidates in the search for position-effect genes with functional consequences in the subjects' phenotypes. For example, *GRIK2* could explain the intellectual disability observed in DGAP138; *SOS1*, forkhead box G1 (*FOXG1* [MIM: 164874]), and cochlin (*COCH* [MIM: 603196]) could be related to the neurological and developmental delay and hearing loss in 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]) might underlie the delayed speech and language development observed in DGAP249; RB binding protein 8 endonuclease (*RBBP8* [MIM: 604124]) might 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]) might contribute to the extreme short stature observed in DGAP275; and zinc finger E-box binding homeobox 2 (*ZEB2* [MIM: 605802]) could 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 real-time qPCR analyses revealed *SOS1* as having lower expression in DGAP163-derived LCLs than in three normal sex-matched control lines (Figure 2). Expression of 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] revealed 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 cellular or animal models that reproduce the observed BCAs for further analysis.

Identification of Subjects with Shared Non-coding Chromosomal Alterations and Phenotypes

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

We found 494 DECIPHER subjects whose rearrangements overlap our 34 non-coding BCA breakpoints (Table S19). Of these, 489 have rearrangements that overlap one or more annotated genes (Table S20). Only five DECIPHER subjects fulfilled our non-coding selection criteria (Table S21): subjects 1985 and 1989, whose rearrangement positions overlap one of DGAP017's breakpoints in chromosome 10 but have several other gene-altering genomic rearrangements; subject 289720, who has a 161.44 kb deletion in chromosome 10 described as likely benign and shares a sequence breakpoint with DGAP126; subject 289865, who has a rearrangement overlapping a breakpoint in chromosome 10 of DGAP126, and similarly subject 289720, who has an additional pathogenic gene-altering rearrangement; and lastly subject 293610, in whom a pathogenic duplication of 364.43 kb in chromosome 17 shares a breakpoint with DGAP288. Only two of the five DECIPHER subjects have reported clinical phenotypes. DECIPHER subject 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 subject 293610 has reported gonadal tissue discordant for external genitalia or chromosomal sex and a non-obstructive azoospermia clinical phenotype;[80] neither feature was observed until puberty, and both 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 dbVar, 675 non-coding structural rearrangements including CNVs, deletions, inversions, and translocations overlapped 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 and 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 ~613 kb duplications 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 detected variants are associated with phenotypes observed in the DGAP subjects, 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 subjects with similar clinical features and chromosomal 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 from de novo germline or somatic mutations, can give rise to different phenotypes through several mechanisms. Chromosomal 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 chromosomal rearrangements has shifted to also assessing regions neighboring these variants.[49] This search for position effects has been particularly important in the analysis of chromosomal 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 disease-causing disruption of the regulatory chromatin environment;[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 chromosomal rearrangements.

We tested the possibility of utilizing chromatin contact information to dissect chromosomal rearrangements that disrupt non-coding chromosome regions in clinical cases. We focused on 17 DGAP subjects (12 of whom have available clinical microarray information) with different rare presentations and de novo non-coding BCAs classified as VUSs. Of these, 15 had translocations, and two had inversions. These subjects represent ~11% of the total number of sequenced DGAP subjects, 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 information, predicted enhancer-promoter interactions, phenomatch scores, and HI and triplosensitivity information for all genes surrounding the BCA breakpoints at different window sizes (±3 and ± 1 Mb and BCA-containing TAD positions), we discovered 16 genes that are top-ranking position-effect candidates for 11 DGAP subjects' clinical phenotypes (Table 1).

We observed that eight of the sequenced DGAP BCA breakpoints, corresponding to six DGAP subjects (DGAP017, DGAP176, DGAP249, DGAP275, DGAP288, and DGAP322), overlapped reported annotated and predicted enhancers and DHSs. 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 DHSs 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, DGAP249, and DGAP287 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 subjects (DGAP113, DGAP126, DGAP138, DGAP153, DGAP163, DGAP252, DGAP315, DGAP319, and DGAP329), six with normal arrays and two with benign CNVs, for whom no overlap with genomic or other regulatory elements was detected. These subjects 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, and GM12878) at different Hi-C window resolutions, 32 breakpoints in H1-hESC TADs,[35] and the separation of 193 genes from 1–91 of their predicted enhancers

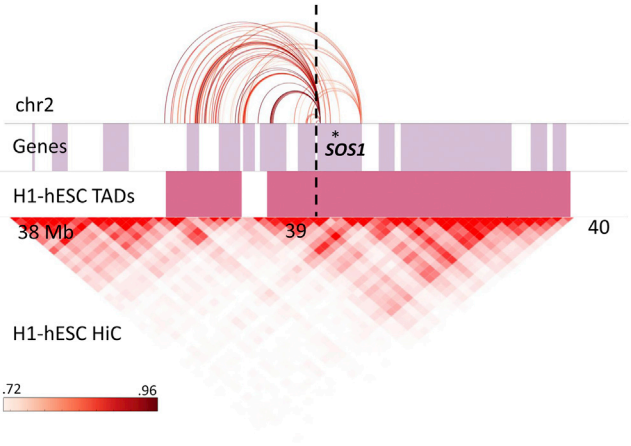

Figure 3. Disrupted DHS/Enhancer-Promoter Interactions Predicted for *SOS1*

Gene position is indicated by an asterisk. The color-graded rectangle represents the correlation values for the interactions reported by Dixon et al.[40] The dashed line indicates the translocation breakpoint position in chromosome 2. Lilac rectangles represent genes, and pink rectangles show TAD positions annotated in H1-hESC.

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 and 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 facial dysmorphism, immunodeficiency, livedo, and short stature syndrome (MIM: 615139) in an autosomal-recessive manner,[83] could contribute to the extreme short stature observed in this DGAP subject. Furthermore, *ZEB2*, related to Mowat-Wilson syndrome (MOWS [MIM: 235730]) in an autosomal-dominant manner (MIM: 235730), could potentially explain the hypotonia and neurological features observed in DGAP329 but not all of the dysmorphic features of MOWS. Overall, assessing the validity of our position-effect predictions and the disruption of important chromatin regulatory elements will require rigorous analysis of more candidate genes. Nonetheless, insight into the molecular pathway of disorders could 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 with those observed in the DGAP subjects. Evidence of HI and triplosensitivity, inclusion in TAD regions, and HI scores build upon this selection and can help laboratories and clinicians focus on candidates of their interest in subsequent analyses. As of now, the "top-ranking" candidates have the most evidence supporting their selection; however, the 17 analyzed DGAP subjects have 102 second-tier candidates within ± 1 Mb analysis windows, and these could very 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: (1) we would need to collect more examples, which might not be easy to find and require a tremendous curation effort, and (2) we need to understand the possibility, suggested by our results, that more than one gene could contribute to the clinical presentation of the DGAP subjects either simultaneously or throughout development. Moreover, many of the candidates have recessive inheritance modes, making 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 via 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 provides valuable interpretation to many VUSs through long-range position effects. The correct prediction of 52 of 57 known pathogenic genes in DGAP subjects used as positive control individuals 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 toward predicting pathogenic consequences of non-coding variation observed in prenatal samples. For example, given its position and chromatin contact alterations, we correctly predicted the involvement and decreased expression of *SOX9* in the cleft palate Pierre-Robin sequence (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 chromosomal 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. Prediction of enhancer adoption 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. Although the existence of other subjects with phenotypes related to those of the DGAP subjects does not prove the involvement of neighboring genes in the etiology of these phenotypes, it is a step toward predicting pathogenic effects by starting from a simple computational analysis, pointing to a better phenotypic categorization during the clinical examination of affected individuals. By making our position-effect prediction method available to the human genetics community (see Web Resources), we hope to study additional subjects with complete phenotypic information and be able to better refine the rules for predicting position effects on gene expression and discover new mechanisms of pathogenicity.

Supplemental Data

Supplemental Data include a Supplemental Note detailing clinical case reports, a Supplemental Note detailing the karyotypes of DGAP subjects, 1 figure, and 22 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2017.06.011.

Acknowledgments

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 C.C.M. and M.E.T.).

Received: March 20, 2017
Accepted: June 19, 2017
Published: July 20, 2017

Web Resources

3D Genome Browser, http://promoter.bx.psu.edu/hi-c
ClinGen GRCh37 data, ftp://ftp.ncbi.nlm.nih.gov/pub/dbVar/clingen
DECIPHER, https://decipher.sanger.ac.uk
ENCODE, https://www.encodeproject.org
Ensembl GRCh37, http://grch37.ensembl.org
GTEx Portal, https://www.gtexportal.org/home
Harvard Biopolymers Facility, https://genome.med.harvard.edu
Human lincRNAs Catalog, http://portals.broadinstitute.org/genome_bio/human_lincrnas
Human Phenotype Ontology, http://human-phenotype-ontology.github.io
NCBI Genome Decoration Page, https://www.ncbi.nlm.nih.gov/genome/tools/gdp/
NCBI Variation Viewer, https://www.ncbi.nlm.nih.gov/variation/view
OMIM, http://www.omim.org
Scripts used for predicting position effects, https://github.com/ibn-salem/position_effect
UCSC Genome Browser, https://genome.ucsc.edu
WashU EpiGenome Browser, http://epigenomegateway.wustl.edu/

References

1. Lejeune, J., Gautier, M., and Turpin, R. (1959). [Study of somatic chromosomes from 9 mongoloid children]. C. R. Hebd. Seances Acad. Sci. *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 Genet. *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., Månér, 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. Stankiewicz, P., and Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. Annu. Rev. Med. *61*, 437–455.
12. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F., Peltonen, L., et al.; International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.
13. 1000 Genomes Project Consortium (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 Res. *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* (R1), R102–R110.
22. Spielmann, M., and Mundlos, S. (2016). Looking beyond the genes: the role of non-coding variants in human disease. Hum. Mol. Genet. *25*, R157–R165.
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., Barisić, I., Ligutic, I., and Nizetić, 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., and Aiden, E.L. (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. Lupiáñez, 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. Gröschel, 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., Köhler, 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 Genet. *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 Evol. Biol. *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 Res. *42*, D749–D755.
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 Genet. *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.; ClinGen (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., Suvà, 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.; and ENCODE Project Consortium (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 Res. *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.; FANTOM Consortium (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–D92.
71. Zhou, X., and Wang, T. (2012). Using the Wash U Epigenome Browser to examine genome-wide sequencing data. Curr. Protoc. Bioinformatics *Chapter 10*, 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 *29*, 2046–2048.
74. Köhler, 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 Res. *42*, D966–D974.
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 Res. *41*, D936–D941.
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.; and GTEx Consortium (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., Debré, M., Nitschke, P., Bole-Feysot, C., Legeai-Mallet, L., et al. (2012). Polymerase ε1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome"). J. Exp. Med. *209*, 2323–2330.

The American Journal of Human Genetics, Volume *101*

Supplemental Data

Computational Prediction of Position Effects of Apparently Balanced Human Chromosomal Rearrangements

Cinthya J. Zepeda-Mendoza, Jonas Ibn-Salem, Tammy Kammin, David J. Harris, Debra Rita, Karen W. Gripp, Jennifer J. MacKenzie, Andrea Gropman, Brett Graham, Ranad Shaheen, Fowzan S. Alkuraya, Campbell K. Brasington, Edward J. Spence, Diane Masser-Frye, Lynne M. Bird, Erica Spiegel, Rebecca L. Sparkes, Zehra Ordulu, Michael E. Talkowski, Miguel A. Andrade-Navarro, Peter N. Robinson, and Cynthia C. Morton

Supplemental Note: Case Reports

DGAP017

46,X,t(X;10)(p11.2;q24.3)dn.arr(1-22,X)x2

Newborn female with a bicornuate uterus, diaphragmatic hernia, thenar hypoplasia, pulmonary hypoplasia, absent right olfactory lobe, loose skin, scoliosis, small thorax, hypoplastic labia, right clinodactyly and camptodactyly, as well as a scaphoid abdomen. This collection of features was reminiscent of Fryns syndrome (FRNS [MIM: 229850]). This case was obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (GM00972).[1] An Affymetrix Genome-Wide Human SNP Array 6.0 performed at Coriell is reportedly normal.

DGAP111

46,XY,t(16;20)(q11.2;q13.2)dn.arr[hg18] 1q23.3(159763523_159905125)x3

Six-year-old male with congenital heart disease (one atrial septal defect, seven small ventricular septal defects), eye anomaly (Duane syndrome), poor growth, developmental delay, chronic constipation, left undescended testis, history of scoliosis (resolved), history of weak ankles and feet requiring braces (resolved), and asthma. Microarray analysis of DNA extracted from the DGAP111 EBV-transformed cell line contributed from DGAP to the NIGMS Human Genetic Cell Repository (GM22709, Coriell) was performed on the Affymetrix Genome-Wide Human SNP Array 6.0 and revealed a duplication of ~141.6 Kb in 1q23.3 (159763523-159905125) that was interpreted as likely benign.

DGAP113

46,XY,t(1;3)(q32.1;q13.2)dn

One-year-old male with bilateral congenital cataracts (TORCH screen, positive IgG and negative IgM for rubella, cytomegalovirus, herpes simplex virus; rubella virus isolation from urine and lens was negative), and mild developmental delay. Cranial magnetic resonance imaging (MRI) revealed prominent extra-axial cerebrospinal fluid spaces of uncertain significance, and the subject has marked macrocephaly (head circumference >95th percentile).[2] No microarray was performed.

DGAP126

46,XX,t(5;10)(p13.3;q21.1)dn.arr[hg18] 7q34(142030226_142154515)x1

Ten-year-old female with significant developmental delay with regression, autistic tendencies, and receptive and expressive language delay, disruptive behavior disorder, enuresis, dysthymia, sleep disturbance, self-injurious behaviors, and agitation. She had delays in gross and fine motor skills. No dysmorphic features were observed. Microarray analysis of DNA extracted from the DGAP126 EBV-transformed cell line contributed from DGAP to the NIGMS Human Genetic Cell Repository (GM18825, Coriell) was performed on the Affymetrix Genome-Wide Human

SNP Array 6.0 and revealed a deletion of ~124.3 Kb in region 7q34 (142030226-142154515) that was interpreted to be benign.

DGAP138

46,XY,t(1;6)(q23;q13)dn.arr(1-22)x2,(X,Y)x1

Seven-year-old male with intellectual disability, fat distribution around trunk, gastroesophageal reflux, feeding problems (gastrostomy), seizure disorder, movement disorder (random, writhing type movements), wheelchair-dependence, Pierre-Robin sequence (mild micrognathia and cleft of the soft palate) (PRBNS [MIM: 261800]), microcephaly, pseudogynecomastia, and low growth hormone and high cortisone levels. Normal microarray results were reported from of DNA extracted from the DGAP138 EBV-transformed cell line contributed from DGAP to the NIGMS Human Genetic Cell Repository (GM20568, Coriell) on the Affymetrix Genome-Wide Human SNP Array 6.0.

DGAP153

46,X,t(X;17)(p11.23;p11.2)dn.arr(1-22,X)x2

Eight-year-old female with dysmorphic features (including mild synophrys, a flat philtrum and thin upper lip vermilion), mild developmental delay, sleep disturbance, and behavior problems (including temper tantrums, self-biting, and agitation). Deletion testing was negative for Smith-Magenis syndrome (SMS [MIM: 182290]). No cryptic aneusomies were reported to be detected by clinical aCGH. The DGAP153 EBV-transformed cell line was contributed to the NIGMS Human Genetic Cell Repository (GM20572, Coriell).

DGAP163

46,XY,t(2;14)(p23;q13)dn.arr(1-22)x2,(X,Y)x1

Four-year-old male with severe global developmental delay, absent speech, dysmorphic/distinctive facies, hypospadias (repaired), seizures as an infant (now seizure free), myopia, nystagmus, small left retinal coloboma, and conductive hearing loss (history of otitis media). MRI showed periventricular white matter changes of unknown origin (no record of anoxic event), and recent electroencephalograms (EEGs) were normal. Fluorescence *in situ* hybridization (FISH) for SMS, DiGeorge syndrome (DGS [MIM: 188400]) and Velocardiofacial syndrome (VCFS [MIM: 192430]) was reportedly normal, as was aCGH using a 1M Agilent array with a resolution of 6.3 Kb.

DGAP176

46,Y,inv(X)(q13q24)mat

Four-year-old male with congenital, severe, bilateral sensorineural hearing loss, cognitive impairment, plagiocephaly, lax joints, and coordination difficulties. Dysmorphic features include macrocephaly, broad forehead, hypertelorism, downslanting palpebral fissures, epicanthic folds, flat midface, rounded nasal tip, flat nasal root, downturned corners of the mouth, simple helix of

left ear, and full lips. He also had fifth finger clinodactyly and bridged palmar creases. No mutations were detected in the coding regions of gap junction protein beta 2 (*GJB2* [MIM: 121011]) or gap junction protein beta 6 (*GJB6* [MIM: 604418]). The mother is mosaic for inv(X)(q13q24) and 45,X but is reportedly healthy.[3] No microarray was performed.

DGAP249

46,XX,t(2;11)(q33;q23)dn.arr(1-22,X)x2

Seven-year-old female with a history of global developmental delay. She has gross and fine motor delays, atypical oral motor skills and limited exploration of sensory materials. At four years she had an abnormal sleep-deprived EEG and increased bilateral electrocortical excitability; at six years EEG results were significantly abnormal with bifrontal symptoms consistent with epileptiform disturbance recorded in the interictal state. She has decreased visual motor integration, and a composite intellectual coefficient (IQ) of 71. Normal clinical microarray results were reported.

DGAP252

46,XY,t(3;18)(q13.2;q11.2)dn.arr(1-22)x2,(X,Y)x1

Four-month-old male whose prenatal course was complicated by polyhydramnios with an accompanying abnormal prenatal ultrasound and MRI, revealing an abnormal cerebellum, dilated cisterna magna, right lung apex cyst, intra-abdominal cysts and bilateral abnormal feet. Delivery was at term with two right posterior mediastinal cysts identified as a foregut duplication cyst and a bronchogenic cyst by pathology after surgical excision. Three ileal cysts were identified as duplication cysts with complete muscularis propria, small bowel/colon, and gastric oxyntic type mucosa by pathologic examination after excision. Cerebellar hypoplasia was noted by MRI of his brain at one day of age. A wide anterior fontanelle (three finger widths) was observed, and his head was reportedly mildly turricephalic with a high forehead and a round bony protrusion of his skull at the occipital base. Normal clinical microarray results (CMA-HR + SNP (v.8.3)) were reported.

DGAP275

46,XX,t(7;12)(p13;q24.33)dn.arr(1-22,X)x2

Nine-year-old female with severe unexplained short stature (<4 SDs) and normal radiographs. An extensive endocrine workup revealed a normal growth hormone axis and no evidence of precocious puberty. She was non-dysmorphic and had normal cognitive development. A normal clinical Affymetrix Cytoscan SNP microarray was reported.

DGAP287

46,XY,t(10;14)(p13;q32.1)dn.arr(1-22)x2,(X,Y)x1

Four-year-old male with a history of global developmental delay and asymmetric spastic diplegia. He is ataxic, non-verbal, and drools frequently. He is non-dysmorphic, and a brain MRI

was normal. Normal clinical Affymetrix Cytoscan HD SNP microarray results were reported.

DGAP288

46,XX,t(6;17)(q13;q21)dn.arr(1-22,X)x2

Prenatal case enrolled in study at 15 weeks, following ultrasound at 11 weeks revealing a cystic hygroma and chorionic villus sampling (CVS) at 12 weeks revealing the t(6;17) apparently balanced chromosome translocation. Normal clinical Affymetrix Cytoscan HD SNP microarray results were reported at 13 weeks. Micrognathia was seen on ultrasound at 18 weeks. At 19 weeks, DGAP sequencing results revealed no genes disrupted by the translocation, and the pregnancy was continued. Polyhydramnios and micrognathia were noted at 28 weeks. Fetal MRI at 34 weeks revealed a small jaw index consistent with micrognathia and retrognathia, glossoptosis, and cleft palate without cleft lip; findings were suspicious for PRBNS. Following delivery at 39 weeks, initial exams revealed a cleft palate. She was placed on continuous positive airway pressure, but otherwise was considered well.

DGAP315

46,XX,inv(6)(p24q11)dn.arr(1-22,X)x2

Fifteen-year-old female with severe static encephalopathy of unknown etiology. She uses a wheelchair, is microcephalic, nonverbal, and has severe generalized spasticity with poorly controlled epilepsy. She had a normal echo and eye examination and reportedly normal aCGH results.

DGAP319

46,XX,t(4;13)(q31.3;q14.3)dn.arr(1-22,X)x2

Thirteen-year-old female with intellectual disability, and height, weight, and head circumference below the 3rd percentile. She has a grade II-IV systolic murmur, abnormal facies, finger and toe abnormalities. This case was obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (GM00972).[1] This case was previously reported.[4] The Affymetrix Genome-Wide Human SNP Array 6.0 performed at Coriell is reportedly normal.

DGAP322

46,XY,t(1;18)(q32.1;q22.1).arr(1-22)x2,(X,Y)x1

Male subject of unknown age with genitourinary malformations, third degree hypospadias, labialized scrotum with palpable descended testes, mild developmental delay, growth delay, and apparently intact hormonal axis. This case was obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (GM16438).[1,5] The Affymetrix Genome-Wide Human SNP Array 6.0 performed at Coriell is reportedly normal.

DGAP329

46,XX,t(2;14)(q21;q24.3)dn.arr[GRCh37/hg19] 18q22.3(72545050_72692202)x1 pat

Five-year-old female with a progressive neurologic disorder. She has nearly constant choreoathetosis, dystonia (including painful neck dystonia), and myoclonic movements, which are exacerbated by fatigue and emotional stress and are worsening with time. She is profoundly hypotonic and non-ambulatory. She is nonverbal but able to follow simple commands. She had a reported normal clinical CytoSure ISCA 8x60K v2.0 microarray, although a paternally inherited 150 Kb deletion at 18q22.3 from her phenotypically normal father was detected.

Supplemental Note: Nucleotide-Level Nomenclature for DGAP karyotypes

Karyotypes of DGAP cases are described using a revised nomenclature that incorporates next-generation sequencing positions from Ordulu et al.[6]

DGAP017
46,X,t(X;10)(p11.2;q24.3)dn.arr(1-22,X)x2.seq[GRCh37/hg19] t(X;10)(10pter->10q25.1(107,711,256)::TATCCTTTG::Xp11.22(51,702,992)->Xpter;10qter->10q25.1(107,714,387)::GAGAAAAC::Xp11.22(51,707,815)->Xqter)dn

DGAP111
46,XY,t(16;20)(q11.2;q13.2)dn.arr[hg18] 1q23.3(159763523_159905125)x3.seq[GRCh37/hg19] (16,20)cx,der(16)(16pter->16q11.2(46,396,774)::16q11.2(46,397,625-46,397,900)::16q11.2(46,408,942-464093{69-70})::20q13.2(53,969,64{0-1}-53,970,162)::20q13.2(53,970,203)->20qter),der(20)(20pter->20q13.2(53,969,63{5-6})::16q11.2(46,403,29{1-2})->16qter)dn

DGAP113
46,XY,t(1;3)(q32.1;q13.2)dn.seq[GRCh37/hg19] t(1;3)(1pter->1q31.3(198,076,14{1})::3q13.13(110,275,76{4})->3qter;3pter->3q13.13(110,275,769)::AGAA::1q31.3(198,076,137)->1qter)dn

DGAP126
46,XX,t(5;10)(p13.3;q21.1)dn.arr[hg18] 7q34(142030226_142154515)x1.seq[GRCh37/hg19] t(5;10)(10qter->10q21.3(67,539,99{7-5})::5p13.3(29,658,44{0-2})->5qter;10pter->10q21.3(67,539,99{0})::5p13.3(29,658,42{6})->5pter)dn

DGAP138
46,XY,t(1;6)(q23;q13)dn.arr(1-22)x2,(X,Y)x1.seq[GRCh37/hg19] t(1;6)(1pter->1q31.2(193,491,602)::6q16.2(100,159,181)->6qter;6pter->6q16.2(100,159,182)::A::1q31.2(193,491,602)->1qter)dn

DGAP153
46,X,t(X;17)(p11.23;p11.2)dn.arr(1-22,X)x2.seq[GRCh37/hg19] t(X;17)(17pter->17p11.2(20,682,69{0-1})::Xp11.3(44,372,16{4-5})->Xqter;17qter->17p11.2(20,682,68{7-4})::Xp11.3(44,372,1{72-69})->Xpter)dn

DGAP163
46,XY,t(2;14)(p23;q13)dn.arr(1-22)x2,(X,Y)x1.seq[GRCh37/hg19] t(2;14)(14qter->14q13(31,717,834)::G::2p23(39,206,240-39,206,384)::2p23(39,206,414)->2qter;14pter-

>14q13(31,717,73{3})::2p23(39,206,24{2})->2pter)dn

DGAP176
46,Y,inv(X)(q13q24)mat.seq[GRCh37/hg19] inv(X)(pter->q13(82,275,014)::ATCAATTTA::q24q13(108,129,970-82,320,86{7-5})::q24(108,149,24{9-7})->qter)mat

DGAP249
46,XX,t(2;11)(q33;q23)dn.arr(1-22,X)x2.seq[GRCh37/hg19] t(2;11)(2pter->2q33.1(199,943,78{1-9})::11q24.1(121,642,3{46-54})->11qter;11pter->11q24.1(121,638,616)::AGATCT::2q33.1(199,943,805)->2qter)dn

DGAP252
46,XY,t(3;18)(q13.2;q11.2)dn.arr(1-22)x2,(X,Y)x1.seq[GRCh37/hg19] t(3;18)(3pter->3q13.11(104,627,622)::TCAATACCTTTTA::18q11.2(19,498,398)->18qter;18pter->18q11.2(19,498,400)::AAAAATGGC::3q13.11(104,627,629)->3qter)dn

DGAP275
46,XX,t(7;12)(p13;q24.33)dn.arr(1-22,X)x2.seq[GRCh37/hg19] t(7;12)(12qter->12q24.33(132,983,131)::TC::7p12.3(46,111,841)->7qter;12pter->12q24.33(132,983,129)::7p12.3(46,111,839)->7pter)dn

DGAP287
46,XY,t(10;14)(p13;q32.1)dn.arr(1-22)x2,(X,Y)x1.seq[GRCh37/hg19] t(10;14)(14qter->14q32.13(95,212,573)::AGTAAAGGGTTGGGTTAC::10p14(10,161,500-10,161,740)::TCG::10p14(10,161,685)->10qter;14pter->14q32.13(95,212,572)::TATCAG::10p14(10,161,498)->10pter)dn

DGAP288
46,XX,t(6;17)(q13;q21)dn.arr(1-22,X)x2.seq[GRCh37/hg19] t(6;17)(6pter->6q21(112,976,04{2-4})::17q24.3(69,728,01{7-9}->17qter;17pter->17q24.3(69,728,006)::CCCTTTA::6q21(112,976,031)->6qter)dn

DGAP315
46,XX,inv(6)(p24q11)dn.arr(1-22,X)x2.seq[GRCh37/hg19] inv(6)(qter->q11.1(63,115,715)::p24.3q11.1(9,394,991-63,115,685)::T::p24.3(9,394,994)->pter)dn

DGAP319
46,XX,t(4;13)(q31.3;q14.3)dn.arr(1-22,X)x2.seq[GRCh37/hg19] t(4;13)(4pter->4q32.2(161,913,247)::13q21.1(59,345,837)->13qter;13pter->13q21.1(59,345,83{5-

6})::4q32.2(161,913,24{7-8})->4qter)dn

DGAP322

46,XY,t(1;18)(q32.1;q22.1)dn.arr(1-22)x2,(X,Y)x1.seq[GRCh37/hg19] t(1;18)(1pter->1q32.2(208,544,055)::ACTCCTCCAACTCCTATGTAGTTG::18q22.1(63,566,045)->18qter;18pter->18q22.1(63,566,053)::TACA::1q32.2(208,544,091)->1qter)dn

DGAP329

46,XX,t(2;14)(q21;q24.3)dn. arr[GRCh37/hg19] 18q22.3(72545050_72692202)x1 pat.seq[GRCh37/hg19] t(2;14)(2pter->2pter->2q22.3(145,110,93{6})::14q31.1(83,574,72{4})->14qter;14pter->14q31.1(83,574,71{5-9}::2q22.3(14,511,09{37-41}->2qter)dn

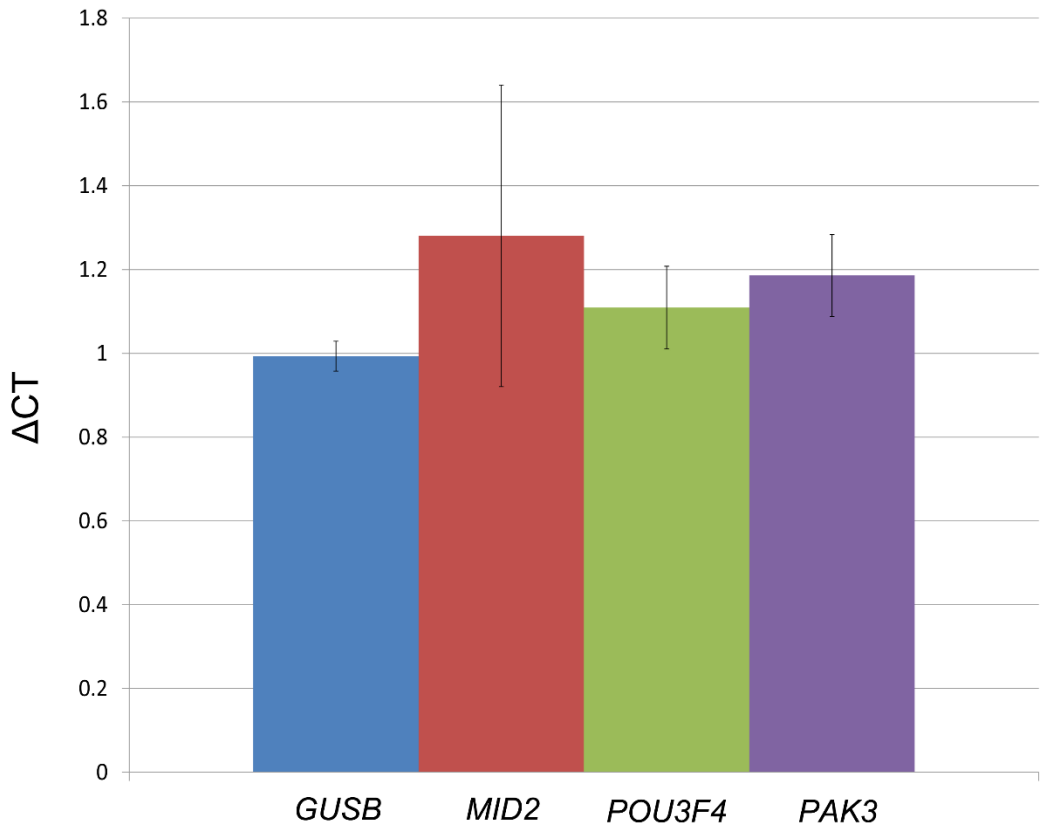

Figure S1. Assessment of gene expression changes for DGAP176-derived LCLs
Control gene expression is shown in blue and surveyed genes are marked in different colors. Each column represents the ΔCT results of three culture replicates, with four technical replicates each, compared to three sex-matched control cell lines. Error bars indicate the standard deviation calculated from the biological replicates per gene.

Supplemental References

1. Tang, Z., Berlin, D.S., Toji, L., Toruner, G.A., Beiswanger, C., Kulkarni, S., Martin, C.L., Emanuel, B.S., Christman, M., and Gerry, N.P. (2013). A dynamic database of microarray-characterized cell lines with various cytogenetic and genomic backgrounds. G3 3, 1143–1149.

2. Lachke, S.A., Higgins, A.W., Inagaki, M., Saadi, I., Xi, Q., Long, M., Quade, B.J., Talkowski, M.E., Gusella, J.F., Fujimoto, A., et al. (2012). The cell adhesion gene PVRL3 is associated with congenital ocular defects. Hum. Genet. 131, 235–250.

3. Anger, G.J., Crocker, S., McKenzie, K., Brown, K.K., Morton, C.C., Harrison, K., and MacKenzie, J.J. (2014). X-linked deafness-2 (DFNX2) phenotype associated with a paracentric inversion upstream of POU3F4. Am. J. Audiol. 23, 1–6.

4. Jenkins, E.C., Curcuru-Giordano, F.M., Krishna, S.G., and Cantarella, J. (1975). De novo occurrence of 46,XX,t(4;13) (q31;q14) in a mentally retarded girl. Ann. Genet. 18, 117–120.

5. Frizell, E.R., Sutphen, R., Diamond, F.B., Jr., Sherwood, M., and Overhauser, J. (1998). t(1;18)(q32.1;q22.1) associated with genitourinary malformations. Clin. Genet. 54, 330–333.

6. 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.

7. 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.

8. 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 Mamm. Genomes identified by analysis of chromatin interactions. Nature 485, 376–380.

9. 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.

10. 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.

11. 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.