Comprehensive clinically oriented workflow for nucleotide level resolution and interpretation in prenatal diagnosis of de novo apparently balanced chromosomal translocations in their genomic landscape

DOI:
10.1007/s00439-020-02121-x

Document Version
Accepted author manuscript

Dezső David[1] João Freixo[2*] Joana Fino[1*] Inês Carvalho[2] Mariana Marques[1] Manuela Cardoso[1] Raul E. Piña-Aguilar[3,4] Cynthia C. Morton[3-7]

Comprehensive clinically oriented workflow for nucleotide level resolution and interpretation in prenatal diagnosis of *de novo* apparently balanced chromosomal translocations in their genomic landscape

[1]National Health Institute Doutor Ricardo Jorge, Department of Human Genetics, Lisbon, Portugal;

[2]Central Lisbon Hospital Center (CHLC), Department of Medical Genetics, Lisbon, Portugal;

[3]Harvard Medical School, Boston, MA, United States;

[4]Departments of Obstetrics and Gynecology and of [5]Pathology, Brigham and Women's Hospital, Boston, MA, United States;

[6]Broad Institute of MIT and Harvard, Cambridge, MA, United States;

[7]University of Manchester, Manchester Academic Health Science Center, Manchester, UK

*These authors contributed equally to the work

Corresponding author:

Dezső David, Ph.D.

Department of Human Genetics

National Health Institute Doutor Ricardo Jorge

Av. Padre Cruz, 1649-016 Lisbon, Portugal

email: dezso.david@insa.min-saude.pt

Fax: (+351) 217526410

Telephone: (+351) 217519322

Acknowledgments We dedicate this article to Maria Guida Boavida who founded the Human Genetics Laboratory as the precursor of the current Department of Human Genetics at the National Health Institute Doutor Ricardo Jorge, Lisbon, Portugal. She and her colleagues performed the first chromosomal prenatal diagnosis from cultured amniocytes in Portugal in July 1977 (no. 77180). However, the first translocation in PND, a paternally inherited Robertsonian t(13;14) (no. 80321), was only identified in June 1980. Almost 40 years later, we report application of an liGS-based approach for identification of dnBCAs breakpoints in PND.

First, we are grateful to the family members for their involvement in this study. We thank Dr. Michael Talkowski and members of his lab for library preparation and sequencing data generation. We thank Jader Cruz for echographic diagnosis of fetal heart defects, to Hildeberto Correia, Ana Cristina Alves and Bárbara Marques for conventional prenatal diagnosis, and to Márcia Rodrigues, Sofia Nunes, Teresa Kay, and Rui Gonçalves for their contribution to the project. C.C.M. is supported (GM061354) by the National Institutes of Health (USA) and the NIHR Manchester Biomedical Research Centre, UK. This research was supported by national funds through FCT - *Fundação para a Ciência e a Tecnologia*, Research Grant HMSP-ICT/0016/2013 of the Harvard Medical School - Portugal Program in Translational Research and Information.

.

Abstract

We present a comprehensive clinically oriented workflow for large-insert genome sequencing (liGS)-based nucleotide level resolution and interpretation of *de novo* (dn) apparently balanced chromosomal abnormalities (BCA) in prenatal diagnosis (PND). Retrospective or concomitant with conventional PND and liGS, molecular and newly developed clinically inspired bioinformatic tools (TAD-GConTool and CNV-ConTool) are applied to analyze and assess the functional and phenotypic outcome of dn structural variants (dnSVs). Retrospective analysis of four phenotype-associated dnSVs identified during conventional PND precisely reveal the genomic elements disrupted by the translocation breakpoints. Identification of autosomal dominant disease due to disruption of *ANKS1B* and *WDR26* by t(12;17)(q23.1;q21.33)dn and t(1;3)(q24.11;p25.3)dn breakpoints, respectively, substantiated the proposed workflow. We then applied this workflow to two ongoing prenatal cases with apparently balanced dnBCAs: 46,XX,t(16;17)(q24;q21.3)dn referred for increased risk on combined first trimester screening and 46,XY,t(2;19)(p13;q13.1)dn referred due to a previous trisomy 21 pregnancy. Translocation breakpoints in the t(16;17) involve *ANKRD11* and *WNT3* and disruption of *ANKRD11* resulted in KBG syndrome confirmed in postnatal follow-up. Breakpoints in the t(2;19) are within *ATP6V1B1* and the 3' UTR of *CEP89*, and are not interpreted to cause disease. Genotype-phenotype correlation confirms the causative role of *WDR26* in the Skraban-Deardorff and 1q41q42 microdeletion phenocopy syndromes, and that disruption of *ANKS1B* causes *ANKS1B* haploinsufficiency syndrome. In sum, we show that an liGS-based approach can be realized in PND care providing additional information concerning clinical outcomes of dnBCAs in patients with such rearrangements.

Keywords: Balanced chromosomal abnormality (BCA); PND care; liGS; KBG syndrome; ANKS1B haploinsufficiency syndrome; Skraban-Deardorff syndrome; TAD-GConTool; CNV-ConTool

Introduction

A causal relationship between a balanced chromosomal abnormality (BCA) and a congenital anomaly is predicted in up to 40% of cases presenting a clinical phenotype associated BCA (Redin et al. 2017). Recognition of *de novo* (dn) BCAs leading to disorders constitutes a formidable challenge in prenatal diagnosis (PND). Conventional low-resolution karyotyping remains the standard approach for assigning rearrangement breakpoints of cytogenetically visible dnBCAs in the prenatal setting. Breakpoints of some BCAs have been localized though molecular cytogenomic approaches including FISH, but high-resolution chromosomal microarrays (CMA) are generally insensitive to BCAs (David et al. 2003). More recently, massively parallel sequencing-based methods have been used facilitating nucleotide level resolution of BCAs (Chen et al. 2008; Talkowski et al. 2011).

Long-insert genome sequencing (liGS), with high physical coverage and low sequence depth, has been applied within an actionable timeframe of a PND, for precise identification of BCA breakpoints (Talkowski et al. 2012; Ordulu et al. 2016). BCAs must also be evaluated in the context of copy-number variation (CNV) burden, and the relevance and expanding knowledge of topologically associated domains (TADs) in mechanisms of disease (Dixon et al. 2012; Lupiáñez et al. 2015).

In the present study, we apply the liGS approach for identification of structural variant (SV) breakpoints in four retrospectively analyzed dnBCAs identified during conventional PND and in two ongoing PNDs with dnBCAs. Two bioinformatic tools to assist prediction of the phenotypic outcome of SVs and CNVs in the routine clinical setting were developed including evaluation of the local genomic landscape in which these dnBCAs occurred. Finally, we consider the predictability of the phenotypic outcome of these dnBCAs identified during PND.

Materials and methods

Patients, karyotyping and CMA

Two fetal and four adolescent probands with dnBCAs identified by fetal karyotyping during a conventional PND protocol and their family members were analyzed. Karyotyping and CMA are described in Supplementary Material and Methods.

liGS library preparation, sequencing, bioinformatic analysis and resolution

liGS library preparation, sequencing, and bioinformatic analysis of sequencing data were carried out essentially as described by Talkowski et al. (2011) and Collins et al. (2017). Briefly, after aligning read-pairs against the reference genome, chimeric and improper read-pairs were selected, categorized, clustered and filtered against a so-called blacklist, a list of genomic regions with systematic short-read mappability biases, with an overlap cut-off ≥ 30% (Collins et al. 2017). Based on cluster analysis, different types of balanced and unbalanced SVs such as translocations, insertions (ins), inversions (inv), complex (cx) SV, deletions (del) and tandem duplications (dup) can be identified by liGS. The resolution of liGS is equivalent to the median insert size plus twice the S.D., *i.e.*, ~4.5 kb. SVs identified in 689 participants with autism spectrum disorder were used as an SV reference dataset (SVref dataset; Collins et al. 2017). Deletions and tandem duplications identified by depth-of-coverage and improper cluster analysis were cross validated (Klambauer et al. 2012; Collins et al. 2016) and analyzed using our CNV-ConTool.

As long as read-pair clusters do not overlap low-complexity regions, our clinically oriented pipeline includes all translocations, ins, del and dup above 30 kb, and cx SV above 10 kb. A more detailed description is available in Supplementary Material.

Identification of cluster-specific split-reads, CNVs and bioinformatic tools

For identification of cluster-specific split-reads encompassing BCA or SV breakpoints, a custom Python algorithm was developed and applied. This process uses read-pairs with one of the reads mapped within a breakpoint cluster and the respective paired read unmapped. Detailed description of this algorithm is available in Supplementary Material.

To assist prediction of the phenotypic outcome of SV and CNVs, two bioinformatic tools were developed. TAD-Gene Content Tool (TAD-GConTool) using TAD data from Dixon et al. (2012) and Moore et al. (2015), identifies breakpoint spanning and flanking TADs and retrieves a series of protein-coding and non-coding RNA genes and genomic elements localized within the TADs, as well as associated structural and functional information. Additionally, this tool has the ability to construct the sequence-based nomenclature of the SVs according to the International System for Human Cytogenomic Nomenclature (ISCN) 2016. This tool will be updated in concert with revisions to ISCN 2016, anticipated to be ISCN 2020. CNV-Content Tool (CNV-ConTool) was developed to search for overlap between patient-specific CNVs and those from public databases. This second tool also retrieves data on genes affected by these CNVs. Detailed descriptions of both bioinformatic tools are available in Supplementary Material and Methods.

Both TAD-GConTool and CNV-ConTool can be accessed online at www.dgrctools-insa.min-saude.pt. Source codes are available at https://github.com/DGRC-PT/.

Amplification of junction fragments

Amplification conditions for junction and control fragments of BCA and proband-specific CNVs are summarized in Supplementary Table 1.

Lymphoblastoid cell lines (LCLs), RNA extractions and expression studies

Establishment of LCLs from peripheral blood lymphocytes, extraction of RNAs from LCLs, peripheral blood and amniocytes, and quality assessment and quantification of RNAs were performed essentially as described previously (David et al. 2003).

Genome-wide assessment of gene expression levels in LCL or amniocytes of the probands and controls were performed using the Affymetrix Human Transcriptome Array 2.0 (HTA 2.0, ThermoFisher Scientific). Sample and array processing and data analysis were carried out according to the manufacturer's instructions and are detailed in the Supplementary Material and Methods.

Variant interpretation and disease prediction

Variants produced by liGS were interpreted according to ACMG sequence based variant criteria (Richards et al. 2015) or CNV criteria (Riggs et al. 2019). For disease prediction (Table 1) bespoke criteria were developed to guide clinicians in the interpretation of sequencing results and ACMG variant classification, as follows:

Disease causing - a structural variant resulting in loss-of-function (LoF) of an annotated gene transcript causing an autosomal dominant (AD) clinically relevant or major developmental disorder, where LoF is a known mechanism of the disease;

Low potential of disease - a structural variant resulting in LoF of an annotated gene transcript solely causing autosomal recessive (AR) disorders or not associated with an AD clinically relevant or major developmental disorder causing gene localized within the breakpoint topological associating domains (bpTADs);

Non-disease causing - a structural variant resulting in no disruption of protein-coding genes within the bpTADs, no human pathology reported to be associated with genomic elements localized within the bpTADs or no statistically significant GWAS data and/or data supporting at least partial overlap between the genetic traits associated with the affected genomic region and the patient phenotype; and

Disease plausible - a structural variant resulting in disruption of an annotated gene transcript intolerant to LoF variants but not yet associated with human disorders, affected gene reported with an important biological function, or convergent genomic and biological evidence (GWAS, gene expression, phenotypic data and other) supporting at least partial involvement of the disrupted gene in the patient clinical phenotype.

Results

Patient medical histories

A 39 year-old female presented with an elevated risk for aneuploidy following 1st trimester combined tests with increased nuchal translucency (4.1 mm, >95th percentile). Chorionic villus sampling (CVS) was performed at 14 weeks gestation for karyotyping of the fetus (designated DGRC0016). Neither parent had any relevant family medical history.

Ultrasound examination at 20 weeks revealed hypoplastic nasal bone and atrioventricular septal defect (AVSD) with ventricular septal defect (VSD) confirmed by fetal echocardiography. Besides AVSD and fetal growth restriction (5th centile), no other fetal anomalies were observed on subsequent evaluations.

Postnatal echocardiogram confirmed the reported AVSD with small VSD but without hemodynamic compromise. At 20 months of age, DGCR0016 presented with developmental delay, most evident in the postural control and locomotor areas, growth restriction and the characteristic facial gestalt to fulfill the criteria for a clinical diagnosis of KBG syndrome (KBGS) (Supplementary Table 2) (Low et al. 2016).

The 40 year-old mother of the second prenatal proband, DGCR0019, had a history of previous pregnancy termination due to trisomy 21 and was referred for amniocentesis at 17 weeks of gestation due to maternal anxiety. Postnatal medical examination of the newborn was phenotypically normal. Besides slightly hypohidrotic skin noticed at four months of age, no other health problem was noted.

Clinical phenotypes of retrospectively analyzed probands are described in Supplementary Results (DGRC0006 - t(8;14), DGRC0013 - inv(13), DGRC0025 - t(12;17), and DGRC0030 - t(1;3)) and summarized in Table 1. Clinical features of probands DGRC0006 and DGRC0013 do not match a specific genetic diagnosis, whereas DGRC0025 (Supplementary Fig. 1 and Supplementary Table 3) and DGRC0030 (Supplementary Fig. 2 and Supplementary Table 4) present clinical phenotypes matching a recently reported *ANKS1B* haploinsufficiency syndrome (Carbonell et al. 2019) and Skraban-Deardorff syndrome (SKDEAS OMIM #617616) (Skraban et al. 2017), respectively.

Conventional prenatal diagnosis

Cytogenetic analysis of the CVS of DGRC0016 revealed a *de novo* apparently balanced reciprocal translocation, 46,XX,t(16;17)(q24;q21.3)dn (Fig. 1a, b). CMA identified an 810 kb *de novo* deletion at 8q24.21 interpreted to be a variant of uncertain significance based on a total score of -0.15 (1A, 2H, 3A, 4I and 5A criteria) obtained from the ACMG CNV interpretation guidelines (Riggs et al. 2019). Karyotyping of the amniotic fluid cells of DGRC0019 revealed a *de novo* apparently balanced reciprocal translocation, 46,XY,t(2;19)(p13;q13.1)dn (Fig. 2a, b).The balanced nature of the translocations was confirmed by CMA and breakpoints mapped on average with a 7 Mb resolution by karyotyping.

Detection of SVs from liGS data

Two prenatal and four retrospective probands were sequenced using Illumina short-read (25 bp) sequencing of liGS libraries. Metrics for the libraries are summarized in Supplementary Table 5. Physical coverage was between 42 to 88-fold whereas sequence depth was about one-fold. Chimeric and improper read-pairs ranged between approximately 4 to 8%.

SVs were identified at liGS resolution of ~4.5 kb, but clinical reported at resolution of ≥ 30 kb. A summary of identified chimeric read-pair clusters denoting translocations, ins, inv and cx SV are shown in Supplementary Table 6. At clinical resolution, on average three fully resolved, novel or non-polymorphic (<1%) SVs were discovered, whereas at liGS resolution, an average of five additional novel or non-polymorphic SVs were identified.

Likewise, a summary of identified del and dup is shown in Supplementary Table 7. At a clinical resolution of ≥ 30 kb, an average of 18 del and dup were identified per proband, but only four del and six dup are considered novel or non-polymorphic (<1%) based on the SVref dataset (Collins et al. 2017). At liGS resolution, after filtering, an additional 14 del and 15 dup were detected per patient.

Identification of breakpoints at nucleotide resolution

liGS of the fetal DNA sample DGRC0016 identified the 16q breakpoint within a 70 bp region (chr16:89,401,663-89,401,732) at 16q24.3, and the 17q breakpoint was delimited to a

2,300 bp region (chr17:46,781,986-46,784,286) at 17q21.31 (Fig. 1). A split-read was found at the der(17) breakpoint. Sequencing of the second fetal DNA sample DGRC0019 identified the 2p breakpoint within a 485 bp fragment (chr2:70,941,289-70,941,773) at 2p13.3, and the 19q breakpoint was mapped within a 132 bp sequence (chr19:32,878,469-32,878,600) at 19q13.11. Junction fragments for both cases were amplified and Sanger sequenced (Supplementary Table 1 and Supplementary Figs. 3 and 4).

The karyotype of DGRC0016 was revised to t(16;17)(16pter→16q24.3::17q21.31→17qter;17pter→17q21.31::16q24.3→16qter)dn, and according to next-gen cytogenetics nomenclature (Ordulu et al. 2014) is described as 46,XX,t(16;17)(q24;q21.3)dn.seq[GRCh38] t(16;17)(16pter→16q24.3(89,401,715)::17q21.31(46,784,035)→17qter;17pter→17p21.31(46,781,998::16q24.3(89,401,718)→16qter)dn. The translocation was reclassified as unbalanced due to the 2,036 bp deletion identified at the 17q21.31 breakpoint (Supplementary Fig. 3).

The karyotype of DGRC0019 was revised to t(2;19)(19qter→19q13.11::2p13.3→2qter;19pter->19q13.11::2p13.3→2pter)dn, and according to next-gen cytogenetics nomenclature is described as 46,XY,t(2;19)(p13;q13.1)dn.seq[GRCh38] t(2;19)(19qter(-)→19q13.11(32,878,515)::2p13.3(+)(70,941,507)→2qter;19pter→19q13.11(+)(32,878,512)::CATA::2p13.3(-)(70,941,502)→2pter)dn.

Characterization of breakpoint regions

In DGRC0016, the 16q24.3 breakpoint at position chr16:89,401,715 disrupts IVS3 of *ANKRD11* (Ankrd11 repeat domain 11, OMIM *611192), whereas the 17q21.31 breakpoint at position chr17:46,781,998 disrupts IVS1 of *WNT3* (Wnt family member 3, OMIM *165330) (Fig. 1). Haploinsufficiency of *ANKRD11* causes AD KBGS (OMIM #148050) (Sirmaci et al. 2011).

Homozygous pathogenic variants in *WNT3* are associated (but not yet independently confirmed) with recessive tetra-amelia syndrome-1 (TETAMS1, OMIM #165330), a severe malformation syndrome that includes complete absence of all four limbs and other severe

anomalies (Niemann et al. 2004). As both *ANKRD11 and WNT3* are transcribed on the negative strand, the translocation results in two chimeric genes (Supplementary Fig. 5). Although the chimeric gene at the der(16) breakpoint lacks *ANKRD11* exons 1-3, it has an intact *ANKRD11* open-reading frame downstream of *WNT3* exon 1, translational initiation codon and *WNT3* 5' regulatory region.

Regarding the gene content of the 16q24.3 breakpoint-spanning TAD (brTAD) in human embryonic stem cells (hESC) (Dixon et al. 2012), the mitochondrial metalloprotease protein coding gene, *SPG7* (paraplegin matrix AAA peptidase subunit) associated with AR/AD adult-onset spastic paraplegia 7 (SPG7, OMIM #607259) is localized 89 kb from the breakpoint (Supplementary Fig. 3 and Supplementary Table 8) (Sánchez-Ferrero et al. 2013). Concerning the 17q21.31 in the brTAD (Fig. 3), in addition to the disrupted *WNT3*, the myosin light chain 4 gene (*MYL4*) is localized 425 kb distal to the breakpoint and is etiologic in dominant atrial fibrillation, familial, 18 (ATFB18, OMIM #617280) with an age of onset of 35 years (Orr et al. 2016). Further distal in the brTAD is *ITGB3* or platelet glycoprotein IIIa, which has been reported to cause AR or AD platelet-related mild bleeding disorders (BDPLT16, OMIM #187800).

In DGRC0019, the 2p13.3 breakpoint at position chr2:70,941,502 disrupts IVS1 of *ATP6V1B1* (ATPase H+ transporting V1 subunit B1, OMIM *192132), whereas the 19q13.11 breakpoint at position chr19:32,878,515 is located within the 3' UTR of the *CEP89* transcript NM_032816 (centrosomal protein 89, OMIM *615470) (Fig. 2). The disrupted ATPase is a component of the vacuolar ATPases, a multi-subunit enzyme that mediates acidification of eukaryotic intracellular organelles. Pathogenic variants within this gene are reported to cause an AR distal renal tubular acidosis with progressive nerve deafness (OMIM #267300) (Karet et al. 1999). A homozygous deletion comprising *CEP89* and *SLC7A9* has been reported in a patient with isolated mitochondrial complex IV deficiency, intellectual disability and multisystemic problems (van Bon et al. 2013). *SLC7A9*, causing cystinuria (OMIN #220100) with AR and AD inheritance with incomplete penetrance, was identified within the hESC and LCL GM12878 brTADs 8.75 kb proximal

from the 19q13.11 breakpoint (Supplementary Fig. 6d and Supplementary Tables 9 and 10) (Rao et al. 2014; Leclerc et al. 2002).

In DGRC0006, the 8q12.3 breakpoint disrupts IVS1 of a large intergenic non-coding (Linc) RNA *LINC01414* or RP11-32K4.1 with a brain-specific expression pattern and unreported biological function (Supplementary Fig. 7). The 14q31.2 breakpoint is in a large gene poor region.

In DGRC0013, IVS1 of *FLT1* (Fms related tyrosine kinase 1, OMIM *165070) is disrupted (Supplementary Fig. 8). *FLT1* is a tyrosine kinase receptor for vascular endothelial growth factors (VEGF) with important roles in angiogenesis and vasculogenesis. Although this receptor has been implicated in development and homeostasis of many organs, it is not yet associated with a human disorder (Tjwa et al. 2003). *Flt1* knockout mice models show increased angiogenesis, left ventricle wall thickening and enlargement of the left ventricle cavity, only the last of which is consistent with the DGRC0013 phenotype (Fong et al. 1995; Mei et al. 2015). However, it is not unsurprising that a disruption of a single allele in *FLT1* is not totally representative of the loss-of function phenotype in the knockout mouse. No *Flt1* knockout mice study showed abnormalities of the tricuspid valve as did DGRC0013, but the repression of VEGF was described as part of the mechanism for heart valve morphogenesis (Chang et al. 2004). Moreover, the enhanced expression of *FLT1* in atrioventricular valves, per FANTOM CAT browser, correlates with the reported valve abnormality in the patient (Hon et al. 2017).

In DGCR0025, *ANKS1B* (Ankrd11 repeat domain 11, OMIM *611192) IVS9 is disrupted (Supplementary Fig. 9). *ANKS1B* is a tyrosine kinase effector of activity-dependent post-synaptic signaling and a component of the postsynaptic density complex (Jordan et al. 2007). *ANKS1B* shows an enriched brain-specific expression pattern. Recently, monogenic heterozygous microdeletions in *ANKS1B* have been reported to cause a spectrum of neurodevelopmental phenotypes (Carbonell et al. 2019).

Finally, in DGRC0030 the breakpoints disrupt exon 12 of *WDR26* (WD repeat-containing protein 26; OMIM *617424) and IVS1 of *ATP2B2* (ATPase plasma membrane

Ca2+ transporting 2; OMIM *108733) (Supplementary Fig. 10). Pathogenic variants in these genes are reported to cause AD SKDEAS and AD non-syndromic sensorineural hearing impairment, respectively (Skraban et al. 2017; Smits et al. 2019).

Genomic imbalances

The median size of del and dup at clinical resolution is 64 and 49 kb, respectively. Two deletions, 53.512 kb at 3p24.1 (27,354,680-27,408,191) and 836.049 kb at 8q24.21 (129,061,233-129,897,281), identified in DGRC0016 (Supplementary Figs. 11, 12 and Supplementary Table 11) were not found in public CNV databases. The deletion at 3p24.1, classified as a VUS with a total score of -0.45 (1A, 3A, 4J(-0.30), 5C(-0.15)) according to ACMG CNV criteria (Riggs et al. 2019), is present in the proband's phenotypically normal mother and brother, and therefore unlikely to contribute to an abnormal phenotype. As for the 836.047 kb *de novo* deletion, none of the affected genes has been considered to cause a reported phenotype and the deletion is interpreted as VUS according to ACMG CNV criteria (Riggs et al. 2019). Moreover, although several genetic traits have been associated by GWAS with the affected genomic region, none of these represent developmental disorders (Supplementary Table 12). Posteriorly, this SV was considered as unrelated to the patient's reported clinical features. Regarding DGRC0019, with the exception of a 12,033 bp deletion within the olfactory receptor family 5 subfamily B pseudogene region (chr11q12.1:58336732-58348764), no other proband-specific alteration was detected (Supplementary Table 13).

Proband-specific del and dup identified in the retrospectively analyzed probands are summarized in Supplementary Table 14, and inv, ins and cx SV in Supplementary Table 15. Most likely, none of these SVs has a pathogenic implication.

Expression studies

From the disrupted genes in prenatal probands, only *ANKRD11* and *CEP89* are ubiquitously expressed in LCLs. *WNT3* shows skin enriched expression, whereas *ATP6V1B1* has kidney, lung and skin enhanced expression (Supplementary Figs. 13 and 14). Expression array profiling of the t(16;17) proband's LCLs shows that due to low sensitivity of this HTA 2.0

array, the whole gene expression level of *ANKRD11* is roughly the same as that in controls (7.77 vs. 7.85, SD 0.09) whereas that of *WNT3* is increased (5.86 vs. 4.6, SD 0.01) (Supplementary Table 16). The increased *WNT3* exon 5 signal intensity (33.58) may explain the observed whole gene elevated *WNT3* expression (Supplementary Fig. 15). Expression levels of the remaining genes from both brTAD were roughly similar to controls (Supplementary Table 16).

HTA 2.0 expression data of cultured human amniocytes are not available in the literature. Therefore, expression data of the t(2;19) proband's cultured amniocytes were compared to LCLs as control. Noticeable altered expression above the threshold of the microarray was not observed at the level of gene, exon or exon splicing (data not shown).

Predictability of the phenotypic outcome of dnBCA

The pathogenicity of an SV should be assessed separately for each breakpoint and jointly as a single alteration. In the absence of established guidelines or criteria for classification of SV, we based our classification on ACMG criteria for sequence variants (Richards et al. 2015). Variant classification and clinical interpretation of BCAs is summarized in Table 1.

During PND of DGRC0016, the t(16;17) rearrangement at 16q24.3 was classified as PM6 (ACMG criterion PM6 - assumed *de novo*, but without confirmation of paternity and maternity) and the absence of *ANKRD11* exons 1-3 in the der(16) as PSV1 (Table 1). Therefore, the 16q24.3 rearrangement was interpreted as a likely pathogenic variant, most likely leading to a KBG syndrome-like phenotype. Postnatally, it was further classified as PP4 (PP4 - patient's phenotype and family history highly specific for a disease with a single gene etiology). Moreover, although KBG syndrome is typically milder and less frequently diagnosed in females, the patient's clinical features meet the diagnostic criteria for KBG (Richards et al. 2015) (Supplementary Table 2) and therefore the ACMG interpretation was upgraded to pathogenic (Table 1).

During PND of DGRC0019, the t(2;19) rearrangement at 2p13.3 was also classified as PM6 and according to our interpretation criteria (Table 1) was predicted to have a "Low potential of disease."

For DGRC0006, none of the affected genes or identified genomic alterations is associated with pathologies nor show overlap with the patient's phenotype. Furthermore, GWAS data do not reach genome-wide statistical significance (Supplementary Table 17). Therefore, based on our criteria we consider this variant as "Non-disease causing" (Table 1).

In DGRC0013 and DGRC0025**,** dnBCA breakpoints directly disrupt genes with a low ratio of observed / expected (oe) number of LoF variants indicating a strong LoF intolerance (Table 1 and Supplementary Tables 18 and 19) but neither are curated in ClinGen. Both SVs (*i.e*., involving *FLT1* and *ANKS1B*) can only be scored to PM6, but based on our criteria are predicted to be "Disease plausible" (Table 1). Of note, the clinical phenotype of DGRC0025 matches a recently reported *ANKS1B* haploinsufficiency syndrome (Carbonell et al. 2019).

Finally, in DGRC0030**,** the SV disrupting the disease gene *WDR26* was classified as PVS1, PM6 and PP4 corresponding to pathogenic by ACMG criteria (Richards et al. 2015). Thus, for the *WDR26* variant our interpretation was "Disease causing" and the proband's clinical phenotype coincides with that of age-matched patients with SKDEAS (Skraban et al. 2017). *ATP2B2* is not curated as a dominant disease locus (Supplementary Table 20), but the *ATP2B2* variant is classified as "Disease plausible" (Table 1).

TAD analysis

It is now clearly established in the literature that disruption of TADs and the creation of neo-TADs are dominant mechanisms of SVs (Lupiáñez et al. 2015; Franke et al. 2016). The main source of knowledge of TAD maps are generated by chromosome conformation capture (Hi-C) data. Details of TADs involved in DGRC0016 are presented in Fig. 3. None of the t(16;17) breakpoints disrupt an interaction loop (data not shown) (Rao et al. 2014). Hi-C contact heatmaps of the t(2;19) breakpoint regions for LCLs and IMR90 are shown in Supplementary Fig. 6. Although the 2p13.3 breakpoint disrupts two interaction loops (Supplementary Fig. 6ab) (Rao et al. 2014), none of the involved genes shows LoF sensitivity.

DISCUSSION

Genome sequencing data of two ongoing fetal and four retrospective samples with dnBCA identified during conventional PND were analyzed by liGS, followed by comprehensive structural analyses of candidate genes from the disrupted bpTADs and prediction of the phenotypic outcome. Moreover, to facilitate implementation of this analysis, two new bioinformatic tools applicable in the clinical setting have been developed. Using this information and the developed bioinformatic tools, we propose an analytical workflow for identification and interpretation of *de novo* SVs in their genomic landscape (Fig. 4).

In DGCR0016, translocation breakpoints disrupt a single allele of *ANKRD11* and of *WNT3*, wherein haploinsufficiency of *ANKRD11* causes AD KBGS. KBGS was first reported by Herrmann et al. (1975) in three unrelated families with the surnames initials being K, B and G. The common phenotypic characteristics of this multiple congenital anomaly comprises, among others, a characteristic facial appearance (including protruding ears and hypertelorism), hand anomalies, neurologic involvement, and postnatal short stature (Skjei et al. 2007), which are consistent with the phenotype observed in the patient and fits KBGS diagnostic criteria (Low et al. 2016).

In DGCR0019, translocation breakpoints disrupt genes tolerant to LoF variants, *ATP6V1B1* and *CEP89*. No gene causing AD or developmental disorder was identified within the bpTADs. The predicted outcome was confirmed by absence of a postnatal clinical phenotype. Nevertheless, longer term follow-up would be warranted to exclude any later onset of a disorder that might be associated as recently demonstrated for prenatally detected dnBCAs (Halgren et al. 2018) or natural history of individuals with postnatal dnBCAs (Currall et al. 2018).

Of the four retrospectively analyzed dnBCAs, similarly to the aforementioned DGCR0016, disruption of *WDR26* predictably will lead to SKDEAS. The patient's clinical phenotype highlights that these phenocopies, SKDEAS and 1q41q42 deletion syndrome, are primarily caused by disruption of *WDR26*.

Although the breakpoints of dnBCAs inv(13) and t(12;17) disrupt genes not yet curated in ClinGen, the fact that these are significantly LoF intolerant genes involved in several biological processes, reinforced by convergent evidence, led us to predict that they are "Disease plausible". Furthermore, DGRC0025 clinical phenotype overlaps *ANKS1B* haploinsufficiency syndrome.

Finally, for the postnatal phenotype of DGRC0006, the t(8;14) variant is predicted as "Non-disease causing." Although pathogenic cx SVs smaller than our clinical resolution cutoff have been reported (Sanchis-Juan et al. 2018), at the higher resolution of liGS no additional presumably pathogenic SV was identified in DGRC0006. Exome sequencing (ES) has not been performed, and other non-genetic factors unrelated to the translocation may be responsible for the phenotype (*e.g.*, environmental or multifactorial factors).

Short-read sequencing by either genome sequencing (GS) or ES has been applied in the prenatal setting. ES in fetuses with structural anomalies was recently elucidated in a large-scale study (Lord et al. 2019; Petrovski et al. 2019) revealing a genetic etiology in about 10% of affected fetuses. However, short-read sequencing is not optimal for identification of SVs. The physical coverage of GS is relatively low, whereas ES is high but will miss breakpoints localized within non-coding sequence. The long-insert size of the liGS libraries, intended for identification of BCAs and CNVs, and low read size results in high physical but low sequence coverage.

The lack of transcriptome data on gestational age- and sex-matched first trimester cells from CVS and amniotic fluid is a current limitation for introduction of gene expression analysis in the clinical prenatal setting. Clearly, the future of prenatal diagnosis for SVs will require generation of gene expression data by RNA-Seq linked to Hi-C of CVS cells and amniocytes, as is available now in public databases for adult tissues. Presently, the interpretation of current PND of dnBCAs could be limited to disruption of major dominant genes leading to Mendelian disorders as occurred for *ANKRD11, WDR26 and ANKS1B.* Nonetheless, cytogenetics laboratories should be attentive to take into consideration the

architectural features of genomes to address fully the disease potential of a SVs (Lupiáñez et al. 2015).

In comparison to karyotyping and CMA analysis, we demonstrate the benefits of an liGS-based approach and our clinically inspired pipeline for identification of dnBCA breakpoints and interpretation of the genomic landscape on which these occurred in the prenatal setting. We show the predictability of the clinical outcome of these BCAs and plan to provide updated bioinformatic tools to facilitate data analysis and a workflow for implementation of genome sequencing in the diagnostic prenatal setting.

1

Compliance with Ethical Standards

Conflict of interest: The authors declare no conflicts of interest.

Informed consent: This study was approved by the Ethics Committee of the National Institute of Health Doutor Ricardo Jorge and was carried out according to the Principles of the Declaration of Helsinki of the World Medical Association. Samples were obtained after written informed consent of the participants or their legal representatives. Additional informed consent was obtained from all individual participants for whom identifying information is included in this article.

1

References

Chen W, Kalscheuer V, Tzschach A, et al. (2008) Mapping translocation breakpoints by next-generation sequencing. Genome Res 18:1143-1149.

Carbonell AU, Cho CH, Tindi JO, et al. (2019) Haploinsufficiency in the ANKS1B gene encoding AIDA-1 leads to a neurodevelopmental syndrome. Nat Commun 10:3529.

Chang CP, Neilson JR, Bayle JH, et al. (2004) A field of myocardial-endocardial NFAT signaling underlies heart valve morphogenesis. Cell 118:649-663.

Collins RL, Stone MR, Brand H, et al. (2016) CNView: a visualization and annotation tool for copy number variation from whole-genome sequencing. bioRxiv 049536; doi: https://doi.org/10.1101/049536.

Collins RL, Brand H, Redin CE, et al. (2017) Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. Genome Biol 18:36.

Currall BB, Chen M, Sallari RC, et al. (2018) Loss of LDAH associated with prostate cancer and hearing loss. Hum Mol Genet 27:4194-4203.

David D, Cardoso J, Marques B, et al. (2003) Molecular characterization of a familial translocation implicates disruption of HDAC9 and possible position effect on TGFB2 in the pathogenesis of Peters' anomaly. Genomics 81:489-503.

Dixon JR, Selvaraj S, Yue F, et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376-380.

Franke M, Ibrahim DM, Andrey G, et al. (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 538:265-269.

Fong GH, Rossant J, Breitman ML. (1995) Role of the FLT1 receptor tyrosine kinase in regulating the assembly of vascular endothelium. Nature 376:66-69.

Halgren C, Nielsen NM, Nazaryan-Petersen L, et al. (2018) Risks and recommendations in prenatally detected *de novo* balanced chromosomal rearrangements from assessment of long-term outcomes. Am J Hum Genet 102:1090-1103.

Herrmann J, Pallister PD, Tiddy W, Opitz JM. (1975) The KBG syndrome: A syndrome of short stature, characteristic facies, mental retardation, macrodontia and skeletal anomalies. Birth Defects Orig Artic 11:7-18.

Hon CC, Ramilowski JA, Harshbarger J, et al. (2017) An atlas of human long non-coding RNAs with accurate 5' ends. Nature 543:199-204.

Jordan BA, Fernholz BD, Khatri L, Ziff EB. (2007) Activity-dependent AIDA-1 nuclear signaling regulates nucleolar numbers and protein synthesis in neurons. Nat Neurosci 10:427-435.

Karet FE, Finberg KE, Nelson RD, et al. (1999) Mutations in the gene encoding B1 subunit of H+-ATPase cause renal tubular acidosis with sensorineural deafness. Nat Genet 21:84-90.

Klambauer G, Schwarzbauer K, Mayr A, et al. (2012) cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res 40:e69.

Leclerc D, Boutros M, Suh D, et al. (2002) SLC7A9 mutations in all three cystinuria subtypes. Kidney Int 62:1550-1559.

Lord J, McMullan DJ, Eberhardt RY, et al. (2019) Prenatal exome sequencing analysis in fetal structural anomalies detected by ultrasonography (PAGE): a cohort study. Lancet 393:747-757.

Low K, Ashraf T, Canham N, et al. (2016) Clinical and genetic aspects of KBG syndrome. Am J Med Genet A 170:2835-2846.

Lupiáñez DG, Kraft K, Heinrich V, et al. (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 161:1012-1025.

Mei L, Huang Y, Lin J, et al. (2015). Increased cardiac remodeling in cardiac-specific Flt-1 receptor knockout mice with pressure overload. Cell Tissue Res 362:389-398.

Moore BL, Aitken S, Semple CA. (2015) Integrative modeling reveals the principles of multi-scale chromatin boundary formation in human nuclear organization. Genome Biol 16:110.

Niemann S, Zhao C, Pascu F, et al. (2004) Homozygous WNT3 mutation causes tetra-amelia in a large consanguineous family. Am J Hum Genet 74:558-563.

Ordulu Z, Wong KE, Currall BB, et al. (2014) Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. Am J Hum Genet 94:695–709.

Ordulu Z, Kammin T, Brand H, 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.

Orr N, Arnaout R, Gula LJ, et al. (2016) A mutation in the atrial-specific myosin light chain gene (MYL4) causes familial atrial fibrillation. Nat Commun 7:11303.

Petrovski S, Aggarwal V, Giordano JL, et al. (2019) Whole-exome sequencing in the evaluation of fetal structural anomalies: a prospective cohort study. Lancet 393:758-767.

Rao SSP, Huntley MH, Durand NC, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159:1665-1680.

Redin C, Brand H, Collins RL, et al. (2017) The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet 49:36-45.

Richards S, Aziz N, Bale S, et al. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405-424.

Riggs ER, Andersen EF, Cherry AM, et al. (2019) Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med; doi: 10.1038/s41436-019-0686-8.

Sánchez-Ferrero E, Coto E, Beetz C, et al. (2013) SPG7 mutational screening in spastic paraplegia patients supports a dominant effect for some mutations and a pathogenic role for p.A510V. Clin Genet 83:257-262.

Sanchis-Juan A, Stephens J, French CE, et al. (2018) Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. Genome Med 10:95.

Sirmaci A, Spiliopoulos M, Brancati F, et al. (2011) Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. Am J Hum Genet 89:289-294.

Skjei KL, Martin MM, Slavotinek AM (2007) KGB syndrome: report of twins, neurological characteristics, and delineation of diagnostic criteria. Am J Med Genet 143A:292-300.

Skraban CM, Wells CF, Markose P, et al. (2017) WDR26 haploinsufficiency causes a recognizable syndrome of intellectual disability, seizures, abnormal gait, and distinctive facial features. Am J Hum Genet 101:139-148.

Smits JJ, Oostrik J, Beynon AJ, et al. (2019) De novo and inherited loss-of-function variants of ATP2B2 are associated with rapidly progressive hearing impairment. Hum Genet 138:61-72.

Talkowski ME, Ernst C, Heilbut A, 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.

Talkowski ME, Ordulu Z, Pillalamarri V, et al. (2012) Clinical diagnosis by whole-genome sequencing of a prenatal sample. N Engl J Med 367:2226-2232.

Tjwa M, Luttun A, Autiero M, Carmeliet P. (2003) VEGF and PlGF: two pleiotropic growth factors with distinct roles in development and homeostasis. Cell Tissue Res 314:5-14.

van Bon BW, Oortveld MA, Nijtmans LG, et al. (2013) CEP89 is required for mitochondrial metabolism and neuronal function in man and fly. Hum Mol Genet 22:3138-3151.

1

Figure legends

Fig. 1 Ideograms, partial karyotype and liGS-based localization of the t(16;17)(q24.3;q21.3)dn breakpoints at genomic and gene levels. **a, b** Ideograms and GTL-banded normal and derivative metaphase chromosomes. Chromosome 17 ideogram is shaded in yellow. Beside the derivative ideograms the karyotype and liGS-based resolution of the breakpoints are specified. Filled diamonds or arrows indicate chromosome breakpoints. Karyotype resolution indicates the size of the identified disrupted chromosome band. **c, d** Localization of the t(16;17)(q24.3;q21.3)dn breakpoints at genomic and gene level based on the translocation-specific chimeric cluster. Black and blue arrowheads depict chimeric reads aligned to chromosomes 16 and 17, respectively. Below, gene structure of the disrupted genes, reference transcript numbers and the translational initiation codons (ATG) are indicated. A split-read between positions chr17:46,781,986 and chr16:89,401,732 identified at the der(17) breakpoint is shown by a double arrowhead.

Fig. 2 Ideograms, partial karyotype and liGS-based localization of the t(2;19)(p13.3;q13.11)dn breakpoints at genomic and gene levels. **a, b** Ideograms and GTL-banded normal and derivative metaphase chromosomes. Chromosome 19 ideogram is shaded in yellow. Beside the derivative ideograms the karyotype and liGS-based resolution of the breakpoints are specified. Filled diamonds or arrows indicate chromosome breakpoints. Karyotype resolution indicates the size of the disrupted chromosome band established by metaphase analysis. **c, d** Localization of the t(2;19)(p13.3;q13.11)dn breakpoints at genomic and gene levels based on the translocation-specific chimeric cluster. Black and blue arrowheads depict chimeric reads aligned to chromosomes 2 and 19, respectively. Below, gene structure of the disrupted *ATP6B1*, reference transcript number and the translational initiation codon are indicated. Additionally, the genomic position of the 19q13.1 breakpoint within the 3'UTR of *CEP89* is shown.

Fig. 3 TADs spanning translocation breakpoints of t(16;17) in LCL GM12878, IMR90, and hESC, and chimeric TADs from derivative chromosomes. **a** 16q24 breakpoint region. **b** 17q21.3 breakpoint region. Chromosome 16 TADs are depicted in black or gray, whereas those of chromosome 17 in blue or light blue. Below the TAD tracks, CCCTC-binding factor (CTCF) sites from the analyzed region are shown according to the Chip-seq track in IMR90 fibroblasts (IMR90 CTCF IgG-rab ChIP-seq Signal from ENCODE/SYDH; ENCODE Project Consortium). Horizontal lines with folded gray arrowheads indicate the position of genes in sense and antisense orientations. Genes are color-coded according to their haploinsufficiency index (HI) available at https://decipher.sanger.ac.uk/ and their LoF intolerance, expressed as oe-ratio of LoF variants stated below the genes (http://gnomad.broadinstitute.org/). *ANKRD11* causing KBGS is marked with a hash mark (#). Data for IMR90 fibroblasts, hESCs and LCL GM12878 are according to Dixon et al. (2012) and Moore et al. (2015), respectively.

Fig. 4 Proposed workflow for nucleotide level resolution and interpretation of *de novo* structural rearrangements in their genomic landscape in prenatal diagnosis. **a** In the case of BCAs associated with fetal anomalies or *de novo* SVs, concomitantly with the conventional PND protocol, liGS should be performed for nucleotide level resolution of the rearrangement breakpoints in their genomic landscape. Inherited BCAs are referred to a clinical geneticist to lay out the follow-up required, including potential inclusion of liGS. Foreseeably, upon improvement of liGS-based methods, all prenatally identified non-polymorphic SV may be analyzed by such an approach.
b Long-insert based genome sequencing such as mate-pair, with short or medium reads, from 2 x 25 to 250 bp, can be used. Increasing the sequence coverage enables identification of SNV and indels from the same sequencing data.
c The proposed bioinformatic workflow:
i) Sequence data decoded in different types of read-pairs,
ii) Clusters denoting different types of balanced and unbalanced SVs (translocations, ins, inv, cx SV, del and dup) are identified at the liGS resolution,
iii) Identification of cluster-specific split-reads,

iv) Genomic regions comprising deletions and tandem duplications revealed by the two procedures merged,

v) Potentially pathogenic candidate genes, genomic loci and CNVs revealed by bioinformatic tools TAD-GConTool and CNV-ConTool, and

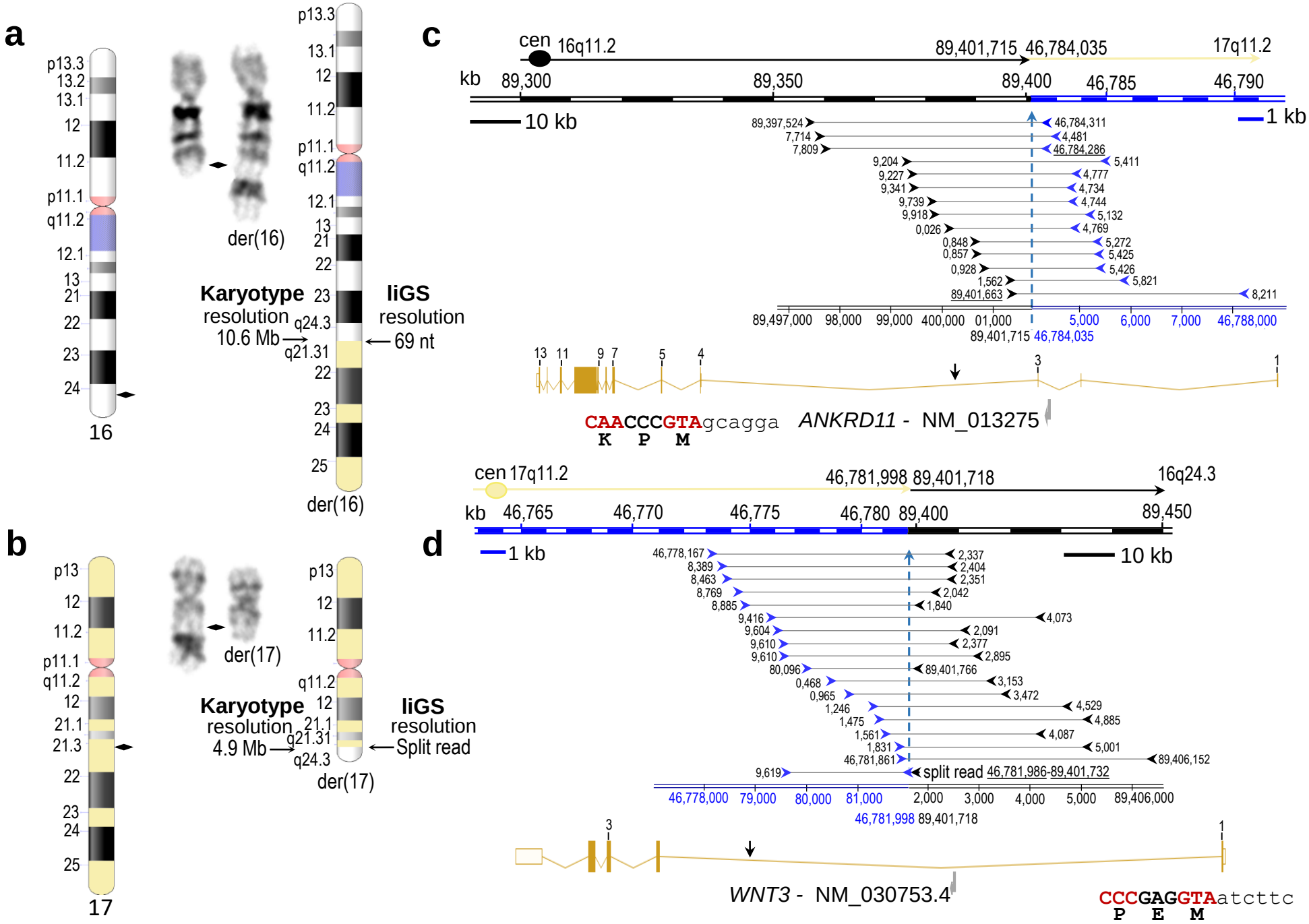
vi) Analysis of the disrupted or dysregulated genes and CNVs, orthogonal confirmation, validation, and preparation of report by a certified medical geneticist.

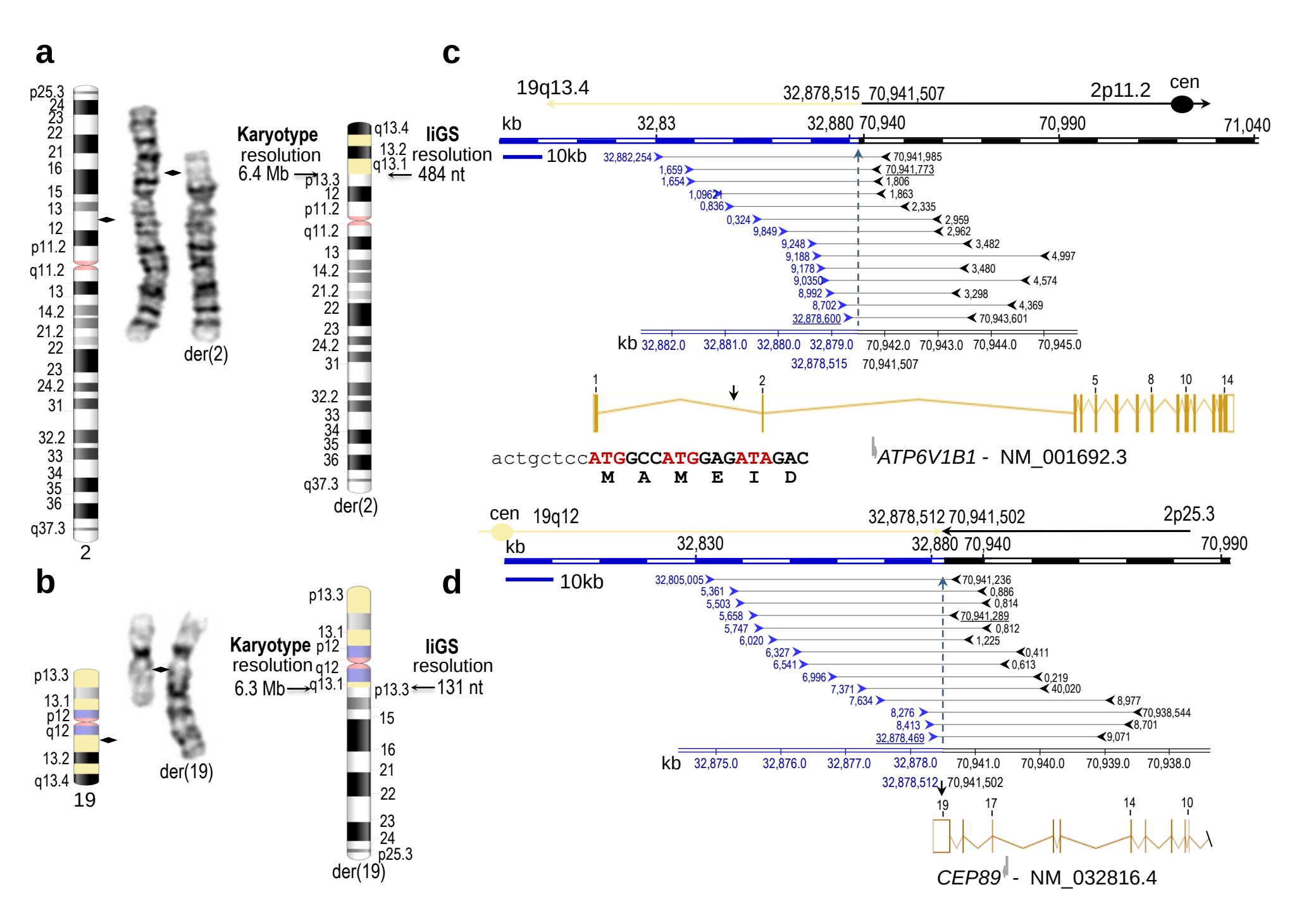
Table 1 Overview of phenotypes, sequencing results and interpretation, and clinical outcomes of patients with dnBCAs

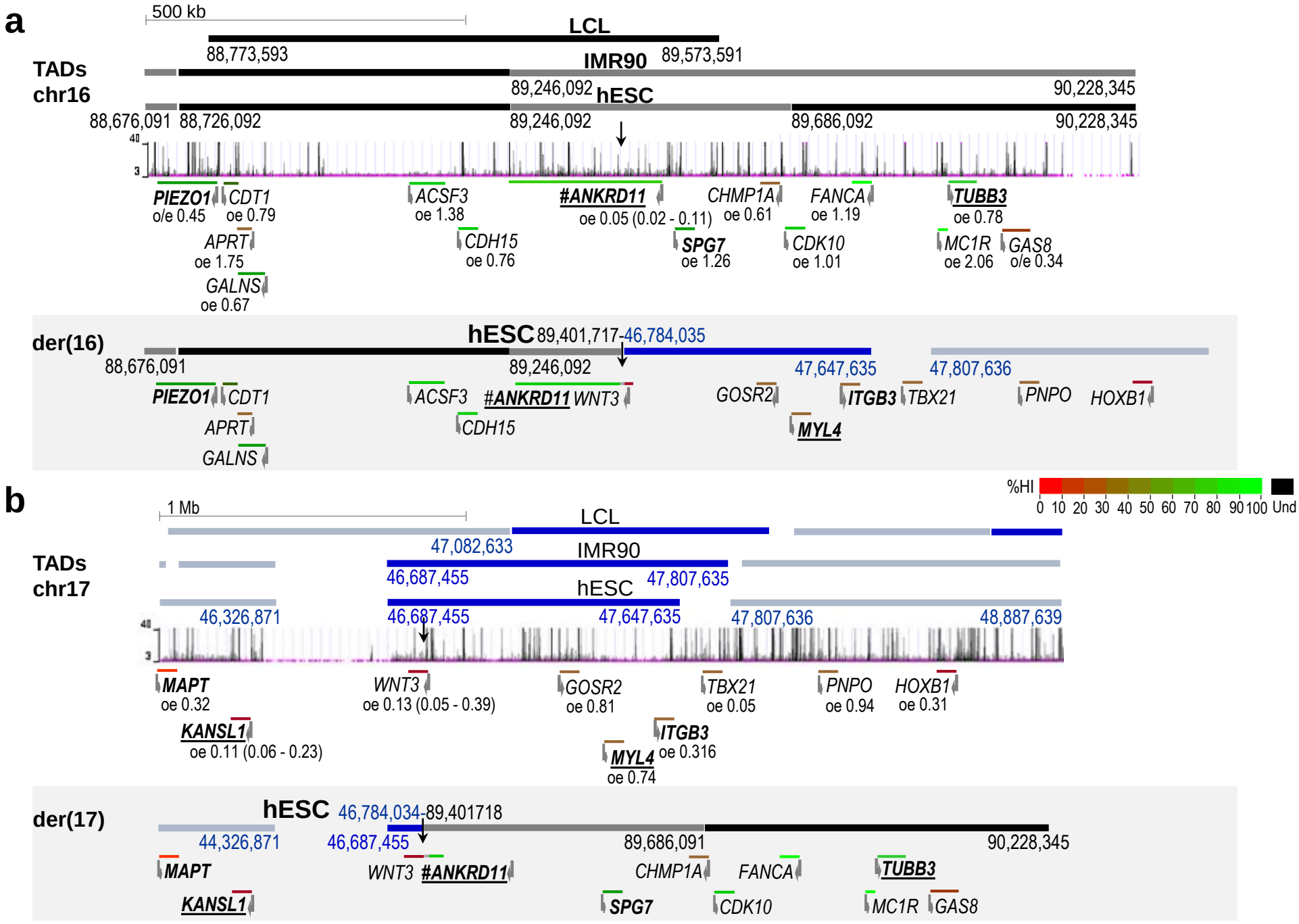
Subject	Indication for PND	Sequence based karyotype	Chromo-some band	Disrupted genes	gnomAD oe (90% CI)	ACMG variant interpretation[a]	Clinical interpretation and prediction[b]	Postnatal clinical phenotype
Prenatally analyzed probands								
DGRC0016	Increased risk for aneuploidy following 1st trimester combined tests and increased nuchal translucency	seq[GRCh38] t(16;17)(q24.3;q21.31)dn	16q24.3	*ANKRD11*	0.05 (0.02 - 0.11)	Prenatal: Likely pathogenic, Posnatal: Pathogenic	Disease causing	KBG syndrome (Supplementary Table 2)
			17q21.31	*WNT3*	0.13 (0.05 - 0.39)	Non-disease associated gene	Non-disease causing	
DGRC0019	Maternal anxiety and advanced maternal age	seq[GRCh38] t(2;19)(p13.3;q13.11)dn	2p13.3	*ATP6V1B1*	0.63 (0.43 - 0.93)	VUS	Low potential of disease	Healthy newborn
			19q13.11	*CEP89*	0.93 (0.72 - 1.21)	Non-disease associated gene	Low potential of disease	
Retrospectively analyzed probands								
DGRC0006	Advanced maternal age and previous child with psychomotor developmental delay	seq[GRCh38] t(8;14)(q12.3;q31.2)dn	8q12.3	*LOC102724623* or *RP11-32K4.1*[c]	nd	Non-disease associated gene	Non-disease causing	Small for gestational age (HP:0001518); Failure to thrive (HP:0001508); Feeding difficulties (HP:0011968); Moderate global developmental delay (HP:0011343); Precocious puberty in females (HP:0010465)
			14q31.2	none			Non-disease causing	
DGRC0013	Advanced maternal age	seq[GRCh38] inv(13)(q12.3q21.1)dn	13q12.3	*FLT1*	0.14 (0.09 - 0.24)	Non-disease associated gene	Disease plausible	Infantile muscular hypotonia (HP:0008947); Moderate global developmental delay (HP:0011343);Generalized joint laxity (HP:0002761); Recurrent upper tract respiratory infections (HP:0002788); Bilateral cryptorchidism (HP:0008689); Phimosis
			13q21.1	none			Low potential of disease	

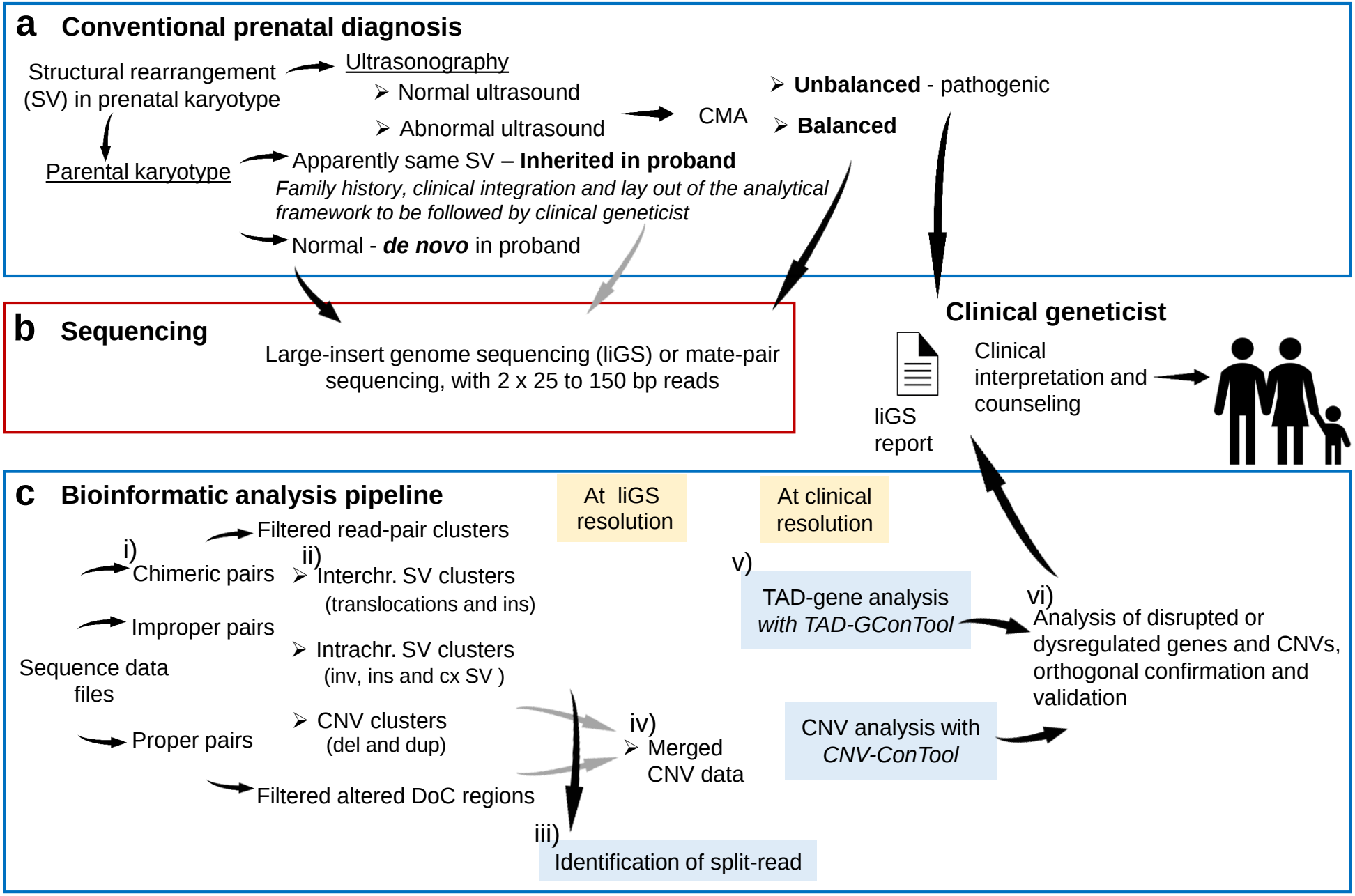
								(HP:0001741); Abnormality of the tricuspid valve (HP:0001702)
DGRC0025	Increased risk for aneuploidy following 1st trimester combined tests and increased nuchal translucency	seq[GRCh38] t(12;17)(q23.1;q21.33)dn	12q23.1	*ANKS1B*	0.10 (0.05 - 0.20)	Non-disease associated gene	Disease plausible	*ANKS1B* haploinsufficiency syndrome (Supplementary Table 3)
			17q21.33	none			Low potential of disease	
DGRC0030	Advanced maternal age	seq[GRCh38] t(1;3)(q42.11;p25.3)dn	1q42.11	*WDR26*	0.00 (0.00 - 0.08)	Pathogenic	Disease causing	Skraban-Deardorff syndrome (Supplementary Table 4)
			3p25.3	*ATP2B2*	0.06 (0.03 - 0.15)	Non-disease associated gene	Disease plausible	Nonsyndromic sensorineural hearing impairment

[a]ACMG variant classification according to ClinGen guidelines non-disease genes require first gene curation and then variant classification. [b]Disease and phenotype interpretation categorized as: i) disease causing, ii) disease plausible, iii) low potential of disease, and iv) non-disease causing, the criteria are described in Supplementary Information [c]Intergenic long non-coding RNAs (lncRNAs); The observed / expected (oe) score is from the Genome Aggregation Database (gnomAD; http://gnomad.broadinstitute.org/); nd= no data


a
p13.3
13.2
13.1
12
11.2
p11.1
q11.2
12.1
13
21
22
23
24
16
der(16)
Karyotype
resolution
10.6 Mb
q21.31
der(16)
liGS
resolution
69 nt
b
p13
12
11.2
p11.1
q11.2
12
21.1
21.3
22
23
24
25
17
der(17)
Karyotype
resolution
4.9 Mb
q21.31
q24.3
der(17)
liGS
resolution
Split read
c
cen
16q11.2
89,401,715
46,784,035
17q11.2
kb
89,300
89,350
89,400
46,785
46,790
10 kb
1 kb
89,397,524
7,714
7,809
46,784,311
4,481
46,784,286
9,204
5,411
9,227
4,777
9,341
4,734
9,739
4,744
9,918
5,132
0,026
4,769
0,848
5,272
0,857
5,425
0,928
5,426
1,562
5,821
89,401,663
8,211
89,497,000
98,000
99,000
400,000
01,000
5,000
6,000
7,000
46,788,000
89,401,715
46,784,035
13
11
9
7
5
4
3
1
CAACCCGTAgcagga
K P M
ANKRD11 - NM_013275
d
cen
17q11.2
46,781,998
89,401,718
16q24.3
kb
46,765
46,770
46,775
46,780
89,400
89,450
1 kb
10 kb
46,778,167
2,337
8,389
2,404
8,463
2,351
8,769
2,042
8,885
1,840
9,416
4,073
9,604
2,091
9,610
2,377
9,610
2,895
80,096
89,401,766
0,468
3,153
0,965
3,472
1,246
4,529
1,475
4,885
1,561
4,087
1,831
5,001
46,781,861
89,406,152
9,619
split read 46,781,986-89,401,732
46,778,000
79,000
80,000
81,000
2,000
3,000
4,000
5,000
89,406,000
46,781,998
89,401,718
3
1
WNT3 - NM_030753.4
CCCGAGGTAatcttc
P E M


a
Karyotype
resolution
6.4 Mb
liGS
resolution
484 nt
der(2)
der(2)
2
b
Karyotype
resolution
6.3 Mb
liGS
resolution
131 nt
der(19)
der(19)
19
c
19q13.4
32,878,515
70,941,507
2p11.2
cen
10kb
actgctccATGGCCATGGAGATAGAC
M A M E I D
ATP6V1B1 - NM_001692.3
cen
19q12
32,878,512
70,941,502
2p25.3
10kb
d
CEP89 - NM_032816.4


a
500 kb
LCL
88,773,593
IMR90
89,573,591
TADs
chr16
89,246,092
hESC
90,228,345
88,676,091
88,726,092
89,246,092
89,686,092
90,228,345
PIEZO1
o/e 0.45
CDT1
oe 0.79
ACSF3
oe 1.38
#ANKRD11
oe 0.05 (0.02 - 0.11)
CHMP1A
oe 0.61
FANCA
oe 1.19
TUBB3
oe 0.78
APRT
oe 1.75
CDH15
oe 0.76
SPG7
oe 1.26
CDK10
oe 1.01
MC1R
oe 2.06
GAS8
o/e 0.34
GALNS
oe 0.67
der(16)
hESC 89,401,717-46,784,035
88,676,091
89,246,092
47,647,635
47,807,636
PIEZO1
CDT1
ACSF3
#ANKRD11
WNT3
GOSR2
ITGB3
TBX21
PNPO
HOXB1
APRT
CDH15
MYL4
GALNS
b
%HI
0 10 20 30 40 50 60 70 80 90 100 Und
1 Mb
LCL
47,082,633
IMR90
46,687,455
47,807,635
TADs
chr17
hESC
46,326,871
46,687,455
47,647,635
47,807,636
48,887,639
MAPT
oe 0.32
WNT3
oe 0.13 (0.05 - 0.39)
GOSR2
oe 0.81
TBX21
oe 0.05
PNPO
oe 0.94
HOXB1
oe 0.31
KANSL1
oe 0.11 (0.06 - 0.23)
ITGB3
oe 0.316
MYL4
oe 0.74
der(17)
hESC 46,784,034-89,401718
44,326,871
46,687,455
89,686,091
90,228,345
MAPT
WNT3
#ANKRD11
CHMP1A
FANCA
TUBB3
KANSL1
SPG7
CDK10
MC1R
GAS8


a Conventional prenatal diagnosis
Structural rearrangement (SV) in prenatal karyotype
Ultrasonography
Normal ultrasound
Abnormal ultrasound
CMA
Unbalanced - pathogenic
Balanced
Parental karyotype
Apparently same SV – Inherited in proband
Family history, clinical integration and lay out of the analytical framework to be followed by clinical geneticist
Normal - de novo in proband
b Sequencing
Large-insert genome sequencing (liGS) or mate-pair sequencing, with 2 x 25 to 150 bp reads
Clinical geneticist
liGS report
Clinical interpretation and counseling
c Bioinformatic analysis pipeline
At liGS resolution
At clinical resolution
Filtered read-pair clusters
i)
Chimeric pairs
ii)
Interchr. SV clusters (translocations and ins)
Improper pairs
Intrachr. SV clusters (inv, ins and cx SV)
Sequence data files
CNV clusters (del and dup)
Proper pairs
Filtered altered DoC regions
iii)
Identification of split-read
iv)
Merged CNV data
v)
TAD-gene analysis with TAD-GConTool
CNV analysis with CNV-ConTool
vi)
Analysis of disrupted or dysregulated genes and CNVs, orthogonal confirmation and validation

Human Genetics Volume 139, 2020

SUPPLEMENTARY MATERIAL

Comprehensive clinically oriented workflow for nucleotide level resolution and interpretation in prenatal diagnosis of *de novo* apparently balanced chromosomal translocations in their genomic landscape

Dezső David[1] João P. Freixo[2*] Joana Fino[1*] Inês Carvalho[2] Mariana Marques[1] Manuela Cardoso[1] Raul E. Piña-Aguilar[3,4] Cynthia C. Morton[3-7]

[1]Department of Human Genetics, National Health Institute Doutor Ricardo Jorge, Lisbon, Portugal;
[2]Department of Medical Genetics, Central Lisbon Hospital Center (CHLC), Lisbon, Portugal;
[3]Harvard Medical School, Boston, MA, United States;
[4]Departments of Obstetrics and Gynecology, and of [5]Pathology, Brigham and Women's Hospital, Boston, MA, United States;
[6]Broad Institute of MIT and Harvard, Cambridge, MA, United States;
[7]University of Manchester, Manchester Academic Health Science Center, Manchester, UK

*These authors contributed equally to the work

Corresponding author:
Dezső David, Ph.D.,
Department of Human Genetics
National Health Institute Doutor Ricardo Jorge
Av. Padre Cruz, 1649-016 Lisbon, Portugal.
email: dezso.david@insa.min-saude.pt
Fax: (+351) 217526410
Telephone: (+351) 217519322

Contents

Supplementary Tables

Supplementary Material and Methods

Ethics statement, karyotyping and CMA

This study was approved by the Ethics Committee of the National Institute of Health Doutor Ricardo Jorge and was carried out according to the Principles of the Declaration of Helsinki of the World Medical Association. Samples were obtained after informed consent of the participants or their legal representatives. Secondary use of DNAs was approved by the Partners HealthCare IRB under the Developmental Genome Anatomy Project (DGAP) protocol.

Analysis of genomic DNAs using Affymetrix (now ThermoFisher Scientific (Waltham, MA, USA)) CytoScan 750K or CytoScan HD microarrays was performed according to the manufacturer's instructions with the analysis parameters of marker count 15 and size 35 kb.

Affymetrix Human Transcriptome Array 2.0 data analysis

Affymetrix Human Transcriptome Array 2.0 data analysis was performed using the Transcriptome Analysis Console (TAC -4.0) annotation file HTA-2_0.r3.na36.hg19.a1.transcript.csv. Gene level and alternative splicing analyses were performed using the Signal Space Transformation -Robust Multichip Average method for background correction and intensity normalization, according to the user manual.

The array detection limit is at a 2-fold change. Fold change (in linear space) of exons and their corresponding junctions (splicing index) were obtained based on normalized exons and junction probe set intensities, respectively.

Large-insert or large-insert jumping genomic sequencing library preparation and sequencing

Large-insert genomic sequencing (liGS) libraries were generated according to the liGS procedures described by Talkowski and coworkers (Talkowski et al. 2011). In brief, 5 µg of fetal DNA was randomly sheared for a target size of 3 kb, end-repaired, ligated with *Eco*P15I cap adapter, and after gel size selection of DNA fragments, circularized using a biotinylated internal adapter containing 2 nt overhangs. Post-circularization steps included DNase digestion of non-circularized products, *Eco*P15I digestion, end-repair of digested DNA fragments, and binding through the biotinylated internal adapter of DNA fragments containing short pieces of DNA from both ends of the inserts to streptavidin-coated magnetic beads.

The following steps of dA tailing, ligation of universal and barcode adapters and PCR amplification with specific primers were carried out on these streptavidin bead-bound DNA fragments. Finally, PCR products were separated from the streptavidin bound fragments and an approximately 200 bp amplicon was gel purified. Multiplex paired-end, 25/26-cycle sequencing of the resulting libraries was performed on an Illumina HiSeq 2000.

Bioinformatic analysis of sequencing data

Bioinformatic analysis of sequencing data was carried out as previously described by Talkowski et al. (2011) and Collins et al. (2017).

In brief, overall quality of sequencing data was assessed by FastQC v0.11.4 and raw read pairs (R1) and (R2) in reverse-forward (outward-facing) orientation converted to standard forward-reverse (inward-facing) orientation by using Seqtk Version 1.0-r82.

The converted mate-pair FASTQ files were aligned with Burrows-Wheeler short-read aligner (BWA v0.7.12) to the reference genome GRCh38/hg38.p9. The alignment output SAM file converted to the binary BAM file format using sambamba v0.6.5 (Li and Durbin 2009; Tarasov et al. 2015).

In a post-processing step, aligned read duplicates are marked and removed by picard tools v1.119 and sambamba v0.6.5, respectively. Then each read is locally realigned using the 1000 Genomes Project datasets as reference for insertions/deletions (indels) by the indelRealigner of the Genome Analysis Toolkit (gatk v3.8.1). Finally, base quality is recalibrated using GATK v3.8.1 BaseRecalibrator and the resulting alignment BAM file is coordinate and name sorted by sambamba v0.6.5.

Subsequently, the BAM file containing aligned and name sorted read-pairs was converted to SAM and submitted to the in-house python script improperCLAS.py. The script starts by sampling 10,000 proper-pairs to calculate median and standard deviation (SD) of the insert size (IS). Then, it categorizes the improper pairs as (i) interchromosomal (interchr) or chimeric, if the reads of the same pair are mapped in different chromosomes; (ii) inversion (inv) improper-pairs, if both reads of the same pair are mapped in the same orientation; (iii) deletion (del) for inward facing and (iv) duplication (dup) for outward facing read pairs with an IS larger than the median IS+3*IS SD, previously calculated.

Improper pairs were clustered together by mapping position using readPairCluster v0.1.0. (Talkowski et al. 2011), creating a set of clusters for each category: chimeric, inv, del and dup.

Clusters with ≥ 30% overlap with biased genome regions showing systematic short-read mappability biases were filtered out based on a so-called “blacklist”. The bulk of the blacklist includes genomic regions with consistently high sequencing depth, compiled by Layer et al. (2014). In addition, the blacklist includes annotated gaps in the reference genome assembly GRCh38/hg38.p9 as well as annotated centromeric, telomeric genomic regions.

Chimeric and inv clusters were divided (if possible) according to their pair orientation, defining two breakpoint specific sub-clusters. Chimeric sub-clusters with less than three read-pairs were discarded.

The chimeric sub-clusters were interpreted for the identification of translocations and interchr insertions (ins), while inv sub-clusters were interpreted for the identification of inv, intrachromosomal (intrachr) ins and complex (cx) structural variants (SV). According to Collins et al. (2017) cx variants involve two or more different distinct SV signatures or three or more breakpoints.

Simultaneously, depth-of-coverage (DoC) analysis was performed to allow a cross validation of the genomic imbalances. Concisely, proper read-pairs of the analyzed six

cases and 27 controls, were submitted to cn.MOPS v1.24 R package, with slight modifications. By default, cn.MOPS divides the genome into 1 kb bins and counts the number of reads per bin; the introduced modification allows cn.MOPS to count all the bins that a proper read-pair insert spans, not only the bins where the reads are mapped. DoC analysis was performed with 1 kb, 3 kb, 10 kb and 30 kb resolutions, allowing the identification of alterations larger than 3 kb, 9 kb, 30 kb and 90 kb, respectively. The results of the four different resolutions were combined using bedtools v2.27.1 merge (Quinlan and Hal 2010), and filtered with the blacklist, as described above for the clustering results.

The filtered DoC results and blacklist-filtered del/ tandem duplication (dup) clusters, were cross validated and searched in SV reference dataset (SVref dataset; Collins et al. 2017). Alterations acknowledged in two of the three (cluster, DoC, SVref dataset) were reported. Potentially pathogenic, novel or non-polymorphic (<1% frequency on SVref dataset) SV were analyzed in more detail using CNV-ConTool. Statistical significance was verified using CNView Collins et al. (2016).

The expected resolution of our analysis, liGS resolution, is equal to the median IS plus twice the SD, i.e., ~4.5 kb. However, as long as read-pair clusters do not overlap low-complexity regions, our clinically oriented pipeline includes all translocations and interch ins, unbalanced SV such as del and tandem dup above 30 kb, and inv, intrachr ins and cx variants above 10 kb.

Identification of cluster specific split-reads

A sequence read that overlaps a balanced or unbalanced SV breakpoint is considered a split-read. The detection of such reads enables straightforward identification of breakpoint junctions at nucleotide resolution. Currently, for short reads, standard alignment software is unable to perform such task. The use of short reads makes identification of split-reads difficult, because alignment of sequence reads to multiple genomic positions is inversely proportional to their size. Additionally, although theoretically possible, searching for split-reads at genome level would require a large amount of computational resources.

For identification of split-reads, an algorithm was developed assuming that: **a)** split-reads were not mapped in previous steps; **b)** a sub-cluster that defines the narrowest breakpoint interval has been identified; and **c)** the pair of the split-read is mapped within the breakpoint defining cluster and is marked in the SAM file as “unmapped-mate”.

To implement the algorithm, a custom python tool – Cluster specific split-read finder – was developed (source code is submitted on GitHub: https://github.com/DGRC-PT/cluster_specific_split_read_finder). The script depends on the BWA software for read mapping and on an application programming interface (API) for connection to the NCBI database through biopython v1.68 (Cock et al. 2009), to retrieve FASTA sequences of specific genomic regions.

The algorithm uses as mapping reference the genomic position of the narrowest breakpoint-spanning interval defined by each breakpoint specific sub-cluster, plus 25 bp on each side, or in tandem dup and ins cases 5 kb regions on each side of the duplicated/inserted region.

Additionally, it requires the reference genome version and the SAM file obtained from the initial mapping.

The algorithm can be divided in two parts: data selection, where the potential split-reads and breakpoint regions are selected and prepared for analysis; and read processing, where BWA tries to map iteratively the potentially split-reads against the breakpoint regions. An example with a translocation cluster is depicted in Fig. S1.

Data selection consists of:
i) retrieving the FASTA sequences for mapping reference through NCBI API and BWA indexing; and
ii) selection, from the SAM file, of unmapped mate read-pairs localized within the breakpoint defining clusters, for posterior processing.

Read processing includes:
i) alignment of the first and last 5 bp, designated as read chunks, of each unmapped read, against the reference sequences, and storage of the mapping data;
ii) repeated realignment of chunks after sequentially increasing their size by 1 bp until no read chunk has a possible alignment or until the chunk size reaches the unmapped mate read size; and
iii) validation of alignment results, outputting only those where read chunks of an unmapped read were mapped to different breakpoint regions, and the sum of the length of the chunks is equal or greater than 80% length of the unmapped read. To contemplate the possibility of ins in the breakpoint regions, the output also includes reads where only one of the chunks mapped, since that chunk comprises at least 68% of the read length.

Results are further validated based on the plausibility of their genomic positions, their orientation relative to each other, to the karyotype and to the reciprocal breakpoint.

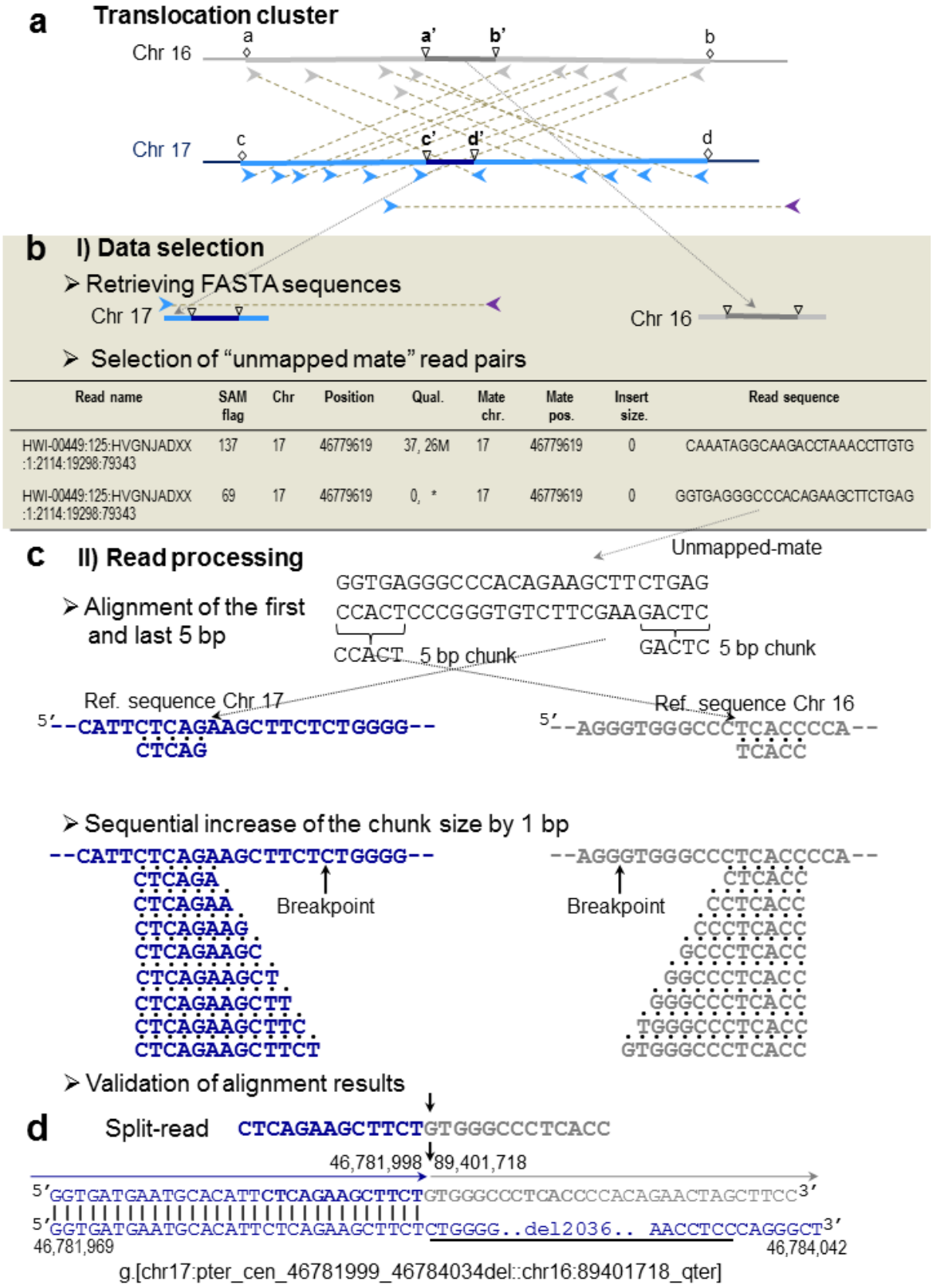


Fig. S1 Identification of split-reads
a Translocation cluster identified by liGS data analysis. Arrowheads depict reads. Reads mapped to chromosome 16 are in gray whereas those on chromosome 17 are

in blue. Arrowheads linked with dashed line are mate-pair read-pairs. Mate-pair reads in different colors are improper pairs. The unmapped-mate read is colored in violet. Open diamonds indicate the chromosome 16 and 17 translocation breakpoint regions defined between genomic positions a-b and c-d, respectively. Triangles indicate the narrowest breakpoint-spanning intervals between a'-b' and c'-d'.
b Data selection step. The retrieved narrowest chromosome 16 and 17 breakpoint-spanning FASTA sequences are in dark blue and gray, respectively. The additional flanking sequences on each side are in light blue and gray, respectively.
Below are shown lines of the SAM file corresponding to the unmapped-mate read-pair, where the first line corresponds to the mapped read and the second to the unmapped read. SAM flag defines read properties, including mapped and unmapped (Sequence Alignment/Map format specification). As default, BWA transcribes the chromosome and the genomic position of the mapped read for the unmapped-mate. The insert size is zero due to the unmapped read.
c Successive alignment steps of read chunks to the reference FASTA sequences. Arrows indicate the breakpoint positions.
d Nucleotide sequence of the der(17) junction fragment aligned against the identified split-read (also see Supplementary Fig. 3).

The script for the identification of split-reads is executed by the command line:

```
% python split_reads_V8.py [chr:a-a'-b'-b] [chr:c-c'-d'-d] [SAM
file] [reference genome][inv/trans/del/dup/ins]
```

The specific command line used for identification of the t(16;17) split-reads is:

```
% python split_reads_V8.py 16:89397524-89401663-89401740-
89406126 17:46778167-46781861-46784260-46788185
DGRC0016_complete_aligmenet.sam hg38 trans
```

Output results in tab-separated text format from the command-line interface and in table format are shown below (Table S1).

```
Original read name      Chunk Read Name Orientation     Size    Position        Sequence        Mate Position   Mate orientation
HWI-D00449:125:HVGNJADXX:1:2114:19298:79343     HWI-D00449:125:HVGNJADXX:1:2114:19298:79343_26.start    reverse 13      16:89401719-89401732    GTGGGCCCTCACC   17:46779619-46779645    forward
HWI-D00449:125:HVGNJADXX:1:2114:19298:79343     HWI-D00449:125:HVGNJADXX:1:2114:19298:79343_26.end      reverse 13      17:46781987-46782000    CTCAGAAGCTTCT   17:46779619-46779645    forward
```

Table S1 Output results of the t(16;17) split-read identification in table format

Original read name	Chunk read name	Strand	Chunk size	Alignment position	Sequence	Mate alignment position	Mate Strand
HWI-D00449:125:HVGNJADXX:1:2114:19298:79343	HWI-D00449:125:HVGNJADXX:1:2114:19298:79343.start	Reverse	13	16:89401719-89401732	GTGGGCCCTCACC	17:46779619-46779645	Forward
HWI-D00449:125:HVGNJADXX:1:2114:19298:79343	HWI-D00449:125:HVGNJADXX:1:2114:19298:79343.end	Reverse	13	17:46781987-46782000	CTCAGAAGCTTCT	17:46779619-46779645	Forward

The output includes the name of the original read and respective chunks (the chunk name is similar to the read name, with the suffix .start/.end according to the position of the aligned chunk within the unmapped read), the orientation in which the chunk is mapped, the size, position and sequence of the chunk, and information about its previously mapped mate.

Although the der(17) split-read is given as an example, split-reads defining del, dup, inv and ins breakpoints have been identified.

Four such instances are:
i) in DGRC0016 a 53 kb del at 3p24.1 shown at Supplementary Fig. 12 (chr3:27,354,65_AAGGAAGGGCAGTTC::chr3:27,408,191_ACTTATCTATA);
ii) in DGRC0006 a 16 kb inv at 4p13 (see Supplementary Table 15) (GCTGC_chr4:43,736,483::chr4:43,751,504_CTCCACTTTCTATACCTTAA);
iii) in DGRC0025 a polymorphic 193 kb dup at 4q13.1 (chr4:63,872,078_GTACAG::CA::chr4:63,674,807_GAAAAAGAACCCCAAAC); and
iv) in DGRC0006 an interchr inverted ins of a 29 kb fragment from 2q36.3 to 15q26.1 (chr15:93,296,516_ATAAGAAAAAAAATACG::GCCATGAA_chr2:227,358,389).

The TAD-Gene Content Tool (TAD-GConTool)

The TAD-GConTool was developed in python with a Common Gateway Interface (CGI) that allows easy, user-friendly applications through any internet browser. The source code of the tool is submitted to GitHub: https://github.com/DGRC-PT/TAD-GConTool and, can be accessed online at http://dgrctools-insa.min-saude.pt.

Based on genomic positions of rearrangement breakpoints the tool identifies breakpoint Topologically Associated Domains (brTADs) and adjacent TADs (upstream TAD-1 and downstream TAD+1), as well as protein coding and non-coding RNA genes, at exon/intron resolution, disrupted by the breakpoints. Genomic position of TAD boundaries in human embryonic stem cells (hESC) and IMR90 fibroblasts (IMR90) are according to Dixon et al. (2012) whereas in human lymphoblastoid cell line (LCL) GM12878 is according to Moore et al. (2015). Because the referred TAD boundary coordinates are only available in the GRCh37/hg19 genome version, these were converted to the GRCh38/hg38 genome assembly using the UCSC Batch Coordinate Conversion (liftOver) tool.

To run the TAD-GConTool, the necessary input data are shown in Fig. S2:
i) the reference genome assembly;
ii) the reference cell line (different cell types must be analyzed independently);
iii) the additional adjacent TADs to be included in the report table (by default, the three TADs are selected for the complete table whereas only the brTAD is included in the report table);
iv) the type of alteration; and
v) the breakpoint information (chromosome and genomic positions).

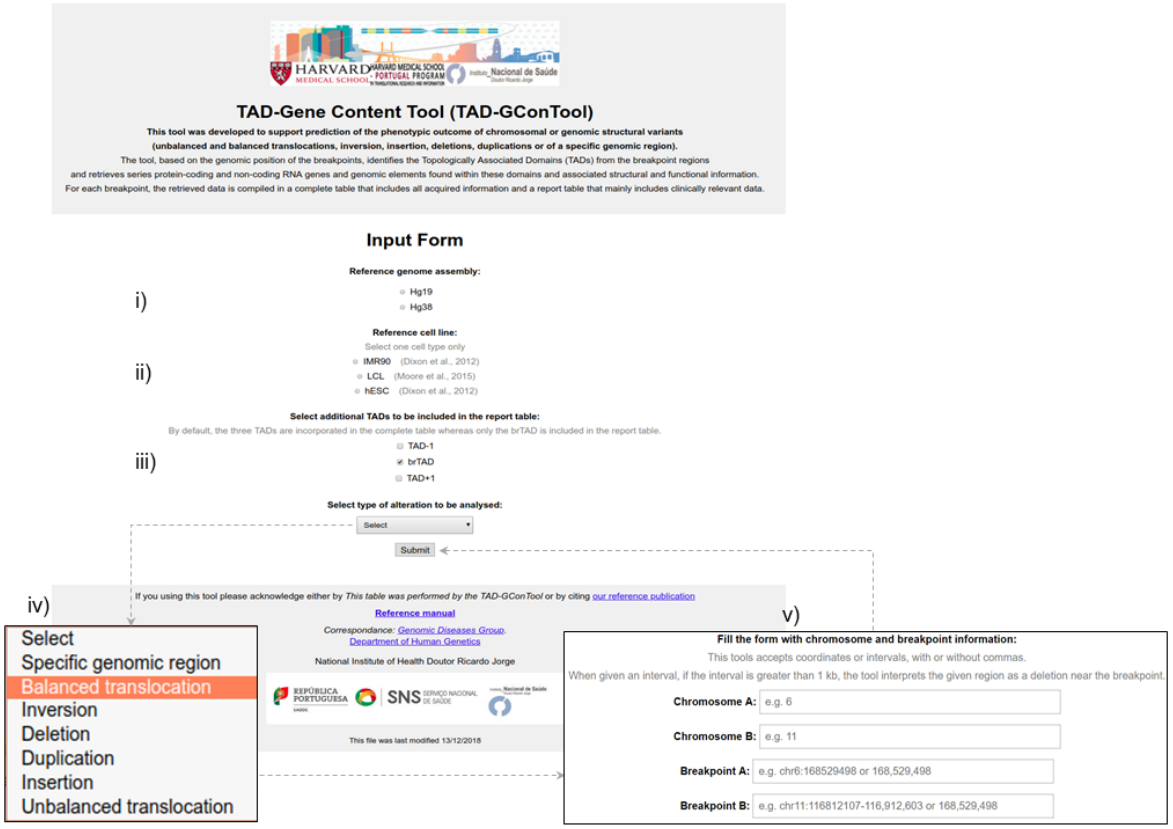


Fig. S2 Landing page of the TAD-Gene Content Tool
The necessary input parameters must be selected as follows:
i) genome version;
ii) reference cell line;
iii) additional adjacent TADs (TAD-1 and TAD+1) to be included in the report table;
iv) within the drop-down list, the alteration type to be analyzed; and
v) data defining the SV (chromosomes and breakpoints information).
Finally, pressing the “Submit” button will launch the tool.
By default, the complete table includes all three TADs (TAD-1, brTAD and TAD+1) whereas the reference table only the brTAD. In this case, unbalanced translocations are those resulting from unequal meiotic segregation of balanced translocations. The tool is unable to handle complex chromosome rearrangements; therefore, these must be subdivided into separate or simpler alterations.

Subsequently, the tool retrieves a series of protein-coding and non-coding RNA genes and genomic elements found within the selected TADs and associated structural and functional information, summarized in Table S2. In addition, the tool retrieves the clinical phenotypes associated with identified genes and highlights those causing major dominant developmental disorders (McKusick 1998; Wright et al. 2015), since these are the most important for prediction of the phenotypic outcome of *de novo* balanced chromosomal abnormalities.

Data are compiled in two distinct output tables per each breakpoint that can be downloaded through the output page shown in Fig. S3. A complete table (Appendix S1.xlsx) includes all acquired information and a report table (Supplementary Tables 8 and 9) mainly includes clinically relevant data. The latter may be included in patient reports.

Additionally, this tool, based on the genomic position of the breakpoints and affected chromosomes establishes the sequence based nomenclature of the rearrangement according to the International System for Human Cytogenomic Nomenclature 2016 (ISCN 2016) (McGowan-Jordan et al. 2016). The tool will be updated to comply with upcoming revisions to the ISCN.

Table S2 Data retrieved by the TAD-GConTool

Data description	Designation	WEB resource	Reference
Protein-coding and non-coding RNA genes (lincRNA, and miRNA) and associated structural information (genomic location, exons/introns, strand, biotype)	Ensembl	https://www.ensembl.org	Zerbino et al. (2018)
HGNC gene name and symbol and link to its integrated GeneCard database	GeneCards	https://www.genecards.org/	Stelzer et al. (2016)
OMIM genes and associated phenotype IDs and their inheritance	OMIM	https://omim.org/	McKusick (1998)
The probability of being haploinsufficiency (HI) sensitive expressed as HI index	DECIPHER	https://decipher.sanger.ac.uk/	Huang et al. (2010) Firth et al. (2009)
The loss of function (LoF) intolerance, expressed as a ratio between observed / expected (oe) number of LoF variants	gnomAD	http://gnomad.broadinstitute.org/	Lek et al. (2016)
Genetic, genomic, and biological data on the mouse ortholog	MGD-MGI	http://www.informatics.jax.org/	Bult et al. (2008)
Clinical phenotypes association based on the literature.	DDG2P	https://www.ebi.ac.uk/gene2phenotype	Wright et al. (2015)

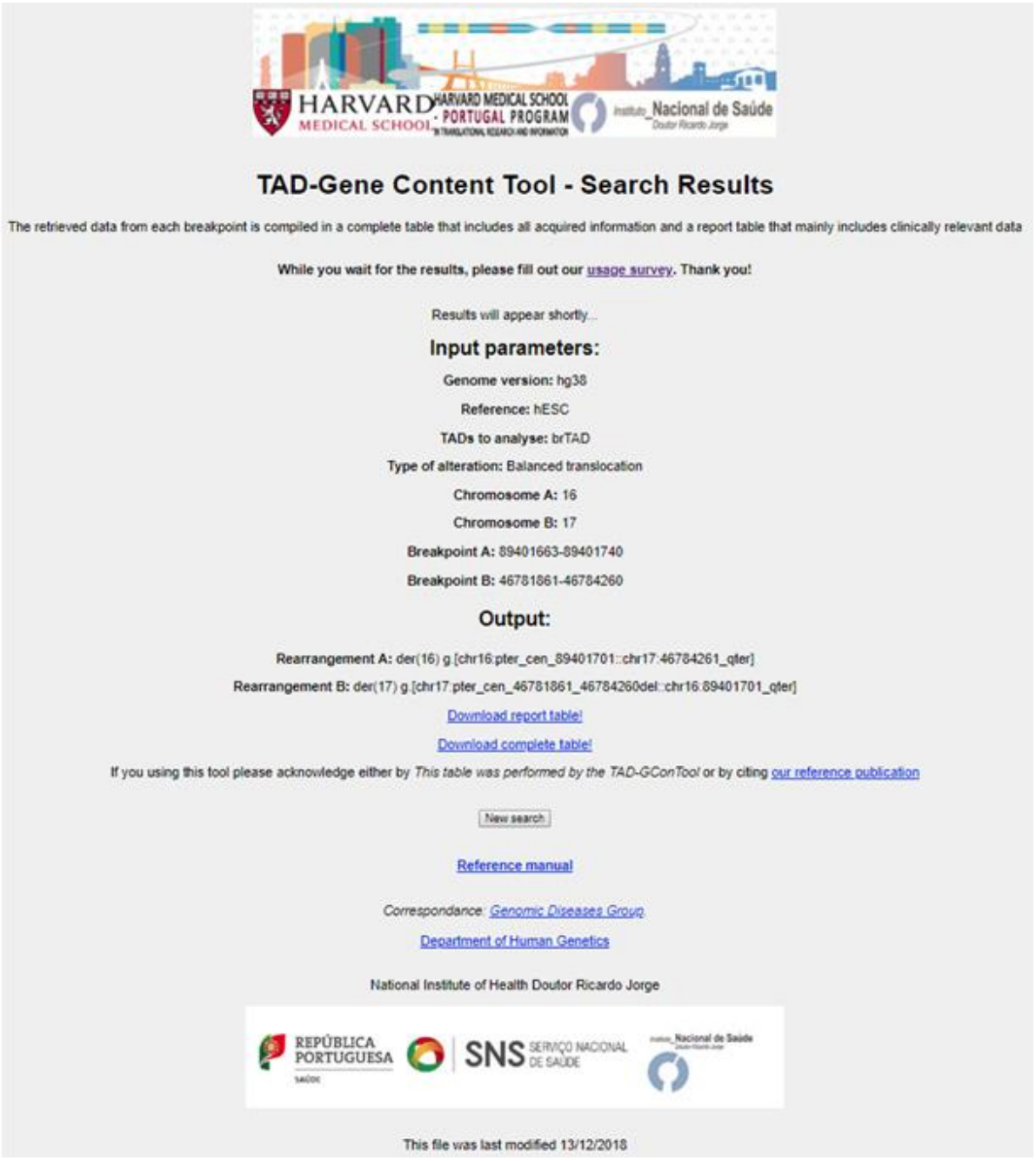


Fig. S3 Search results page of the TAD-Gene Content Tool
The page includes a usage survey, the summary of the input parameters, sequence-based nomenclature of the breakpoints according to ISCN 2016, and links to download the report and complete tables as well as the survey page.

The CNV-Content Tool (CNV-ConTool)

This tool was essentially developed to allow a faster and more informed evaluation of the SV identified in analyzed subjects. For that, CNV-ConTool is able to define the overlap between patient-specific breakpoints and CNVs and those reported in public databases. The reasoning behind this tool is summarized in Fig. S4. For breakpoint analysis, the tool identifies and reports all overlapping CNVs, whereas for unbalanced alterations, a mutual overlap or a query comprised by the reference search can be applied. Additionally, the tool identifies protein-coding and non-coding RNA genes

disrupted or deleted by these patient-specific alterations and performs a mutual overlap search.

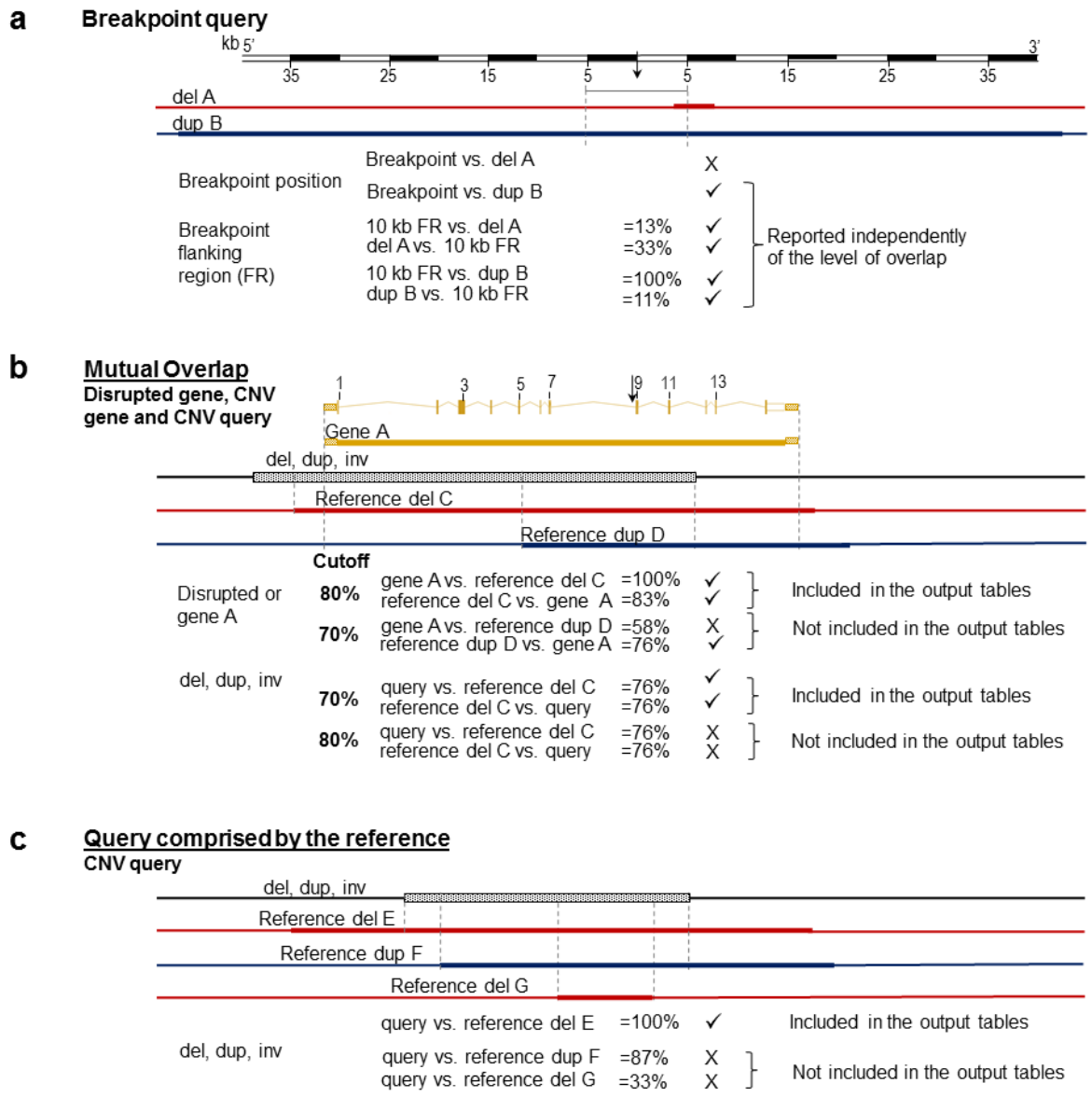


Fig. S4 Schematization of the mutual overlap and query comprised by the reference approaches and retrieved data

Vertical arrow indicates the breakpoint position. Reference dels are in red whereas dups are in blue. Gray dashed vertical lines outline the overlap between the query and reference alterations.

a Breakpoint genomic position and breakpoint-flanking region. All CNVs overlapping the query genomic position or breakpoint-flanking region are retrieved.

b Mutual overlap approach applied in CNVs (del, dup), inv and affected genes. Affected genes by either CNVs or breakpoints are depicted in yellow. Additional gene flanking genomic sequences included in the overlap search are depicted by shaded regions flanking the query genes. The query CNV is illustrated by the dashed back bar. In this approach, the percentage of overlap between query vs. reference and reference vs. query is computed, which ensures that, these show approximately the same size, and can be considered as similar. Only alterations with overlaps above the cutoff value (defined by the user), are included in the output tables.

c Query comprised by the reference search for del, dup and inv. In this approach, only the query vs. reference overlap percentage is taken into consideration: reference CNVs that cover 100% of the query are included in the output tables. This approach can be applied for CNVs and inv, but not for the affected genes, which by default are analyzed with the mutual overlap approach.

CNV-ConTool was developed in python with a user-friendly CGI interface, similar to TAD-GConTool. The source code of the tool is submitted to GitHub: https://github.com/DGRC-PT/CNV-ConTool and, can be accessed online at http://dgrctools-insa.min-saude.pt.To run the CNV-ConTool, the necessary input parameters must be selected and query data specified as shown in Fig. S5.

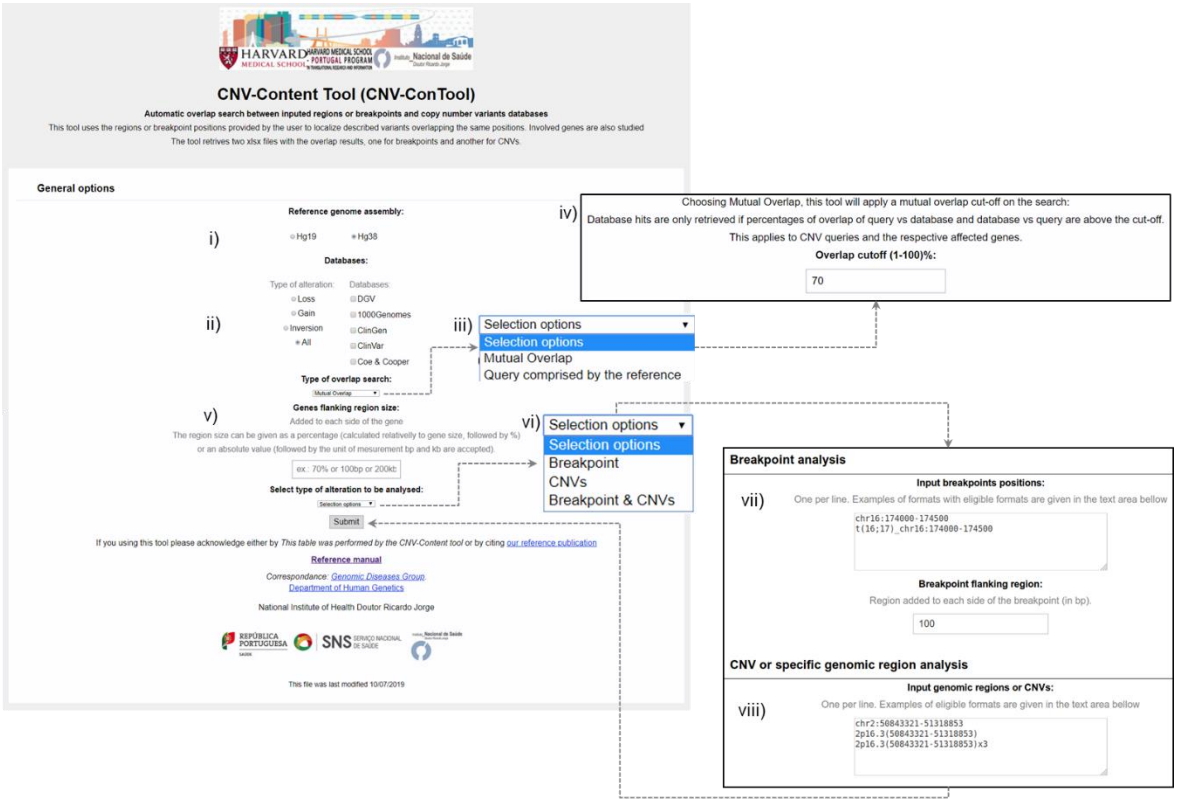


Fig. S5 Landing page of CNV-ConTool with input parameters and query data
Input parameters and query regions must be specified as follows:
i) genome version;
ii) type of alteration to search against and reference databases (see Table S3);
iii) type of overlap search: mutual overlap or query comprised by the reference (see Fig. S4).
iv) overlap cutoff, available when mutual overlap is chosen;
v) flanking gene regions to be included in the overlap search (the size of these can be given as an absolute value in bp or kb, or as a relative value, a percentage calculated relative to the size of each gene) – optional;
vi) select the type of analysis (breakpoints, CNVs or both);
vii) breakpoint analysis - input the breakpoint positions and the size of the breakpoint flanking region to be analyzed; definition of a region size to be added to each side of the breakpoint – optional; and
viii) CNV or specific genomic region analysis - input the genomic positions of CNVs or of specific genomic regions to be analyzed. The application starts by identifying coding and non-coding genes within the genomic regions to be analyzed. Then, the genomic positions of these genes are included in the overlap search outlined at Fig. S4.

Table S3 Databases available on CNV-Content tool

Data description	Designation	WEB resource	Reference
Database of Genomic Variants: a comprehensive database of structural variation in the human genome of healthy control samples	DGV	http://dgv.tcag.ca/dgv/app/home	MacDonald et al. (2014)
Comprehensive database of common human genetic variation in over 2500 subjects from 26 populations	1000 Genomes Project	http://www.internationalgenome.org/data	Auton et al. (2015)
Clinical Genome Resource Database defines the clinical relevance of genes and variants. Variant classification: benign, potentially benign, uncertain, potentially pathogenic, pathogenic	ClinGen	https://www.clinicalgenome.org/	Miller et al. (2010) Kaminsky et al. (2011)
ClinVar is a public archive of reported associations between human variations and phenotypes.	ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/	Landrum et al. (2016)
An expanded CNV morbidity map from almost 45,000 children with developmental delay.	Developmental Delay (Coe & Cooper)	https://www.ncbi.nlm.nih.gov/dbvar/studies/nstd100/ https://www.ncbi.nlm.nih.gov/dbvar/studies/nstd54/	Coe et al. (2014) Cooper et al. (2011)

The retrieved data, compiled in two xlsx files, is made available through the output page shown in Fig. S6. The first file contains the CNVs whereas the second the breakpoint data. Both include protein-coding and non-coding RNA genes affected by these regions. Each of these files are composed by a summary of the overlap results across user-defined databases, followed by one specific table per database, with the most complete information (Appendix S2).

The summary table shows for each query region and affected gene the number of CNV hits, frequencies, databases and the best hits. The following database specific tables consider the same results in a more extensive way, including all hits (one per line), respective ID (with link), size in bp, coverage region, type of alteration, support, bibliographic references and both calculated overlap percentages.

The source code for cluster-specific split-reads finder, TAD-GConTool and CNV-ConTool is available at https://github.com/DGRC-PT/.

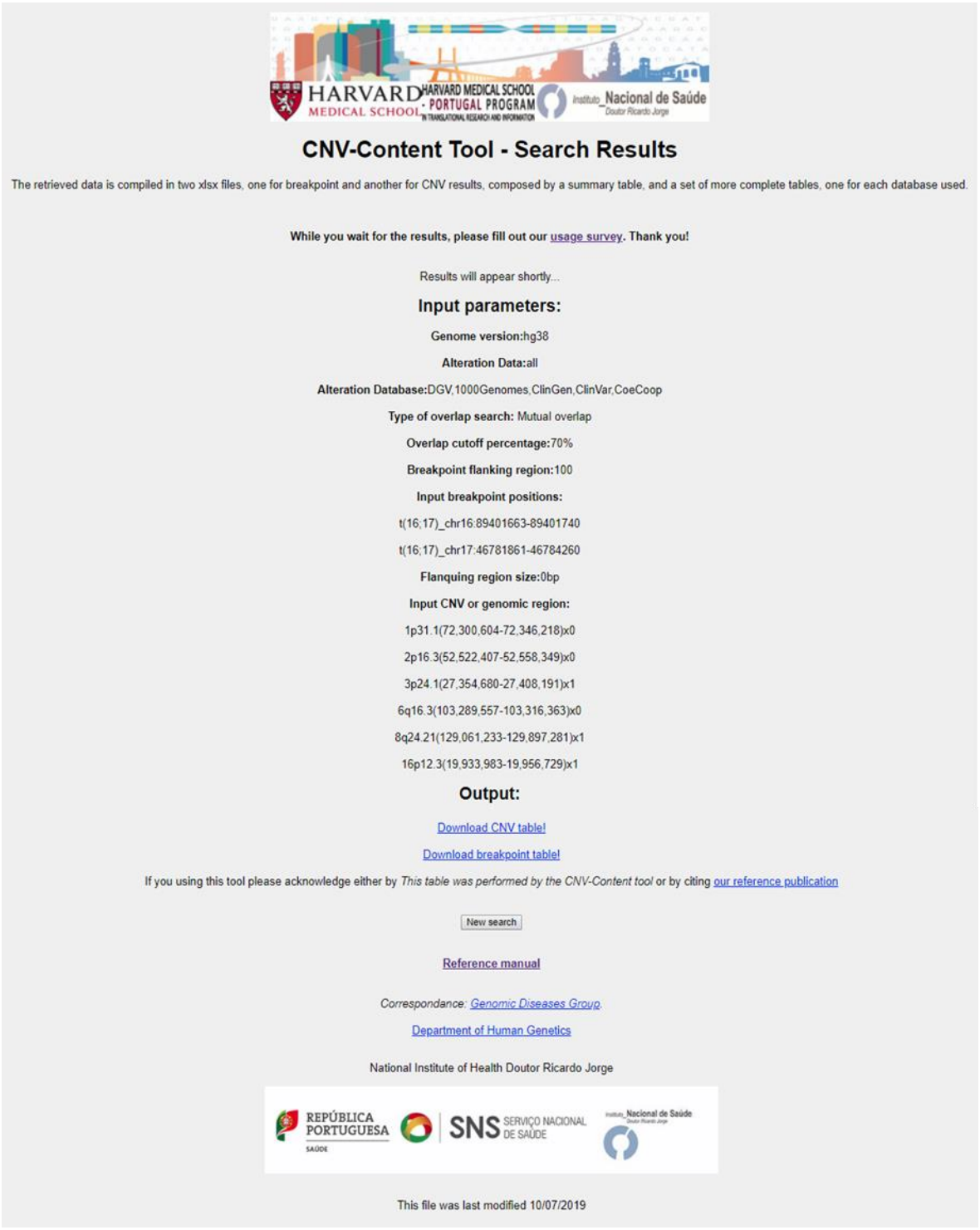


Fig. S6 Search results page of the CNV-Content Tool
The page includes the summary of the input parameters and regions and links to download the xlsx files.

Visualization of high-resolution chromosome conformation capture data

To evaluate genomic interactions, publicly available Hi-C data for IMR90 fibroblasts (Dixon et al. 2012) and GM12878 lymphoblastoid cell lines (Rao et al. 2014) were analyzed.

Hi-C interaction data, visualized as heatmaps, of the breakpoint regions from the available cell lines at different resolutions were assessed using Juicebox (Durand et al. 2016), a tool for exploring Hi-C contact map data.

Heatmaps of the selected genomic regions were combined with relevant genomic data:

i) TAD boundaries according to different publications: for hESC (Dixon et al. 2012), IMR90 fibroblasts (Dixon et al. 2012; Rao et al. 2014), and LCL GM12878 (Moore et al. 2015; Rao et al. 2014).

ii) chromatin loops for IMR90 fibroblasts and GM12878 lymphoblastoid cell lines (Rao et al. 2014);

iii) CCCTC-binding factor (CTCF) sites track (IMR90 CTCF IgG-rab ChIP-seq Signal from ENCODE/SYDH) plot;

iv) transcribed human enhancers color-coded according to their specific tissue/cell expression; and

v) the gene map with haploinsufficiency index, and with the LoF intolerance expressed as observed vs. expected (oe) number of LoF variants stated below the genes.

Supplementary Results

Retrospectively analyzed probands

DGRC0006 – t(8;14)

Proband DGRC0006 is a 12 year-old female with a *de novo* balanced reciprocal chromosomal translocation between the long arms of chromosomes 8 and 14, 46,XX,t(8;14)(q11.23;q24.3)dn identified during conventional prenatal diagnosis (PND) performed due to advanced maternal age and a previous child with psychomotor developmental delay.

At age 12 years abnormal clinical findings include moderate global developmental delay (HP:0011343) with speech problems as well as feeding difficulties (HP:0011968) and failure to thrive (HP:0001508) with decreased body weight (HP:0004325) (< 3rd percentile) and height at 15th percentile. Premature pubarche (HP:0012411) with pubic hair growth was present from seven years of age.

Mapping of the breakpoints by liGS followed by Sanger sequencing identified the 8q11.23 breakpoint at g.[chr8:pter_cen_64,209,134::chr14:83,126,596_83,126,598dup_qter] within IVS 1 of LINC01414 whereas the 14q24.3 breakpoint was identified at g.[chr14:pter_cen_83,126,598::chr8:64,209,135_qter] (Supplementary Fig. 7).

DGRC0013 – inv(13)

Proband DGRC0013 is a 13 year-old male with a *de novo* paracentric chromosomal 13 inversion 46,XY,inv(13)(q12.3q22)dn identified during conventional PND performed due to advanced maternal age.

Abnormal clinical findings include moderate global developmental delay (HP:0011343), especially at motor level, mild hypotonia (infantile muscular hypotonia, HP:0008947) and generalized joint laxity (HP:0002761).

Cardiac evaluation revealed dilated left ventricle, non-hypertrophied walls, with overall conservation of good systolic function and tricuspid valve ring thickening (abnormality of the tricuspid valve, HP:0001702).

Recurrent upper tract respiratory infections (HP:0002788) that led to amygdalectomy and adenoidectomy. Bilateral cryptorchidism (HP:0008689) and phimosis (HP:0001741) were surgically corrected by orchidopexy and preputioplasty, respectively.

Mapping of the breakpoints by liGS fallowed by Sanger sequencing identified the proximal 13q12.3 inversion breakpoint at position g.28,489,796 within IVS 1 of *FLT1* whereas the distal 13q22 breakpoint is at position g.74,831,804 (chr13:g.[pter-cen_28,489,793_28,489,795del;28,489,796_74,831,804inv;_74,831,805_74,831,817del_qter]) (Supplementary Fig. 8).

DGRC0025 – t(12;17)

Proband DGRC0025 is an 8 year-old male with a *de novo* balanced reciprocal chromosomal translocation between the long arms of chromosomes 12 and 17 46,XY,t(12;17)(q23;q22)dn identified during conventional PND performed due to

increased risk for aneuploidy following 1st trimester combined tests and increased nuchal translucency.

In the neonatal period, pulmonary artery stenosis (HP:0004415) and patent foramen ovale (HP:0001655) were identified by electrocardiography.

Physical examination did not reveal significant dysmorphism other than one glabellar (HP:0001076) and two lumbar hemangiomas (Supplementary Fig. 1). Additional facial features observed in this proband are thin upper lip vermilion, depressed nasal bridge, slightly anteverted nares noted throughout development, and synophrys (HP:0000664) evident from eight years of age (Supplementary Fig.1 and Supplementary Table 3).

Neurological assessment revealed delayed speech and language development HP:0000750, expressive language delay HP:0002474, and borderline fine motor skills and practical performance. At 48 months, his developmental age was equivalent on average to 39.5 months. Presently he has early intervention support with speech therapy.

Mapping of the breakpoints by liGS followed by Sanger sequencing identified the 12q23 breakpoint at g.[chr12:pter_cen_99,637,772_99,637,782del::chr17:51,565,697_qter] within IVS 9 of *ANKS1B* whereas the 17q22 breakpoint was identified at g.[chr17:pter_cen_51,565,696::chr12:99,637,783_qter] (Supplementary Fig. 9).

Clinical features of DGRC0025 overlap those of individuals with monogenic heterozygous microdeletions in *ANKS1B*, reported by Carbonell et al. (2019). Furthermore, we confirm that disruption of this gene results in *ANKS1B* haploinsufficiency syndrome.

DRGR0030 – t(1;3)

Proband DGRC0030, the third child of a non-consanguineous couple, is a 13 year-old male. Conventional PND performed due to advanced maternal age identified a *de novo* balanced reciprocal chromosomal translocation between the short arm of chromosome 1 and long arm of chromosome 3, 46,XY,t(1;3)(q42;p25)dn.

The proband's father and siblings are healthy, whereas his mother suffers from bilateral congenital cataracts (HP:0000519) and severe myopia (HP:0011003).

Delivery occurred at 37 weeks gestation and was uneventful. At birth, weight and height were 2.686 kg (7.5th percentile) and 49.5 cm (42nd percentile) according to WHO tables, respectively. Head circumference measured 33 cm (12th percentile) on WHO tables

The proband is characterized by dysmorphic facial features (Supplementary Fig. 2 and Supplementary Table 4) that include coarse face (HP:0000280) with prominent maxilla (HP:0430028) and upper lip (HP:0000215), wide mouth (HP:0000154), widely spaced teeth (HP:0000687), flat nasal bridge (HP:0005280) and a broad full nasal tip (HP:0000455) with anteverted nares (HP:0000463). His growth was stable with weight constantly around the 50th percentile and height around the 25th percentile. The head circumference was always around the 10th percentile.

Cardiovascular examination revealed a complex cardiopathy characterized by tetralogy of Fallot (HP:0001636) with a right aortic arch (HP:0012020) and subaortic interventricular communication.

Neurologic abnormalities include severe global developmental delay (HP:0011344) with absent speech (HP:0001344) and a severe behavioral clinical condition with hyperactivity (HP:0000752). Multiple stereotypies (HP:0000733), broad-based ataxic gait (HP:0002136), and an aberrant posture with tendency to hyperextension of the neck were also reported. A slight thoracic scoliosis (HP:0002943) developed with time.

Most likely due to severe developmental delay, the proband presented in addition to hypotonia and severe feeding difficulties (HP:0011968), with nocturnal and diurnal enuresis (HP:0000805) and encopresis (HP:0040183).

Additionally, several other congenital anomalies were noted such as bilateral talipes equinovarus (HP:0001776), supernumerary nipples (HP0002558), phimosis (HP:0001741), divergent strabismus (HP:0000486), joint hypermobility (HP:0001382), especially of the interphalangeal joints (HP:0005620). Unilateral hydrocele testis (HP:0000034) and an umbilical hernia (HP:0001537) were diagnosed later in infancy.

Mapping of the breakpoints by liGS fallowed by Sanger sequencing identified the 1q42.11 breakpoint at g.[chr1:pter_cen_224398162_224398174del::chr3:10670892_pter] within *WDR26* exon 12 whereas the 3p25.3 breakpoint was identified at g.[chr1:qter_224398174::chr3:10670893_10670894del_cen_qter] within IVS 1 of *ATP2B2* (Supplementary Fig. 10). *WDR26* encodes a WDR domain-containing protein presumably involved in multiple disease-associated signaling pathways.

In conclusion, disruption of *WDR26* by the 1q42.11 breakpoint most likely leads to its haploinsufficiency due to nonsense mediated RNA decay, resulting in a complex clinical phenotype matching both Skraban-Deardorff syndrome (SKDEAS OMIM #617616) (Skraban et al. 2017) and the 1q41q42 microdeletion syndrome. Therefore, DGRC0030 was diagnosed as having severe Skraban-Deardorff syndrome (SKDEAS OMIM #617616).

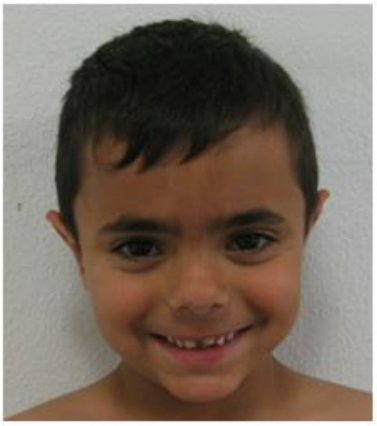
Moreover, the proband’s clinical features basically confirm the phenotypic overlaps between SKDEAS and the 1q41q42 microdeletion syndrome and affirm the causative role of *WDR26* in these phenocopy syndromes.

Supplementary Figures

4 months

4 years

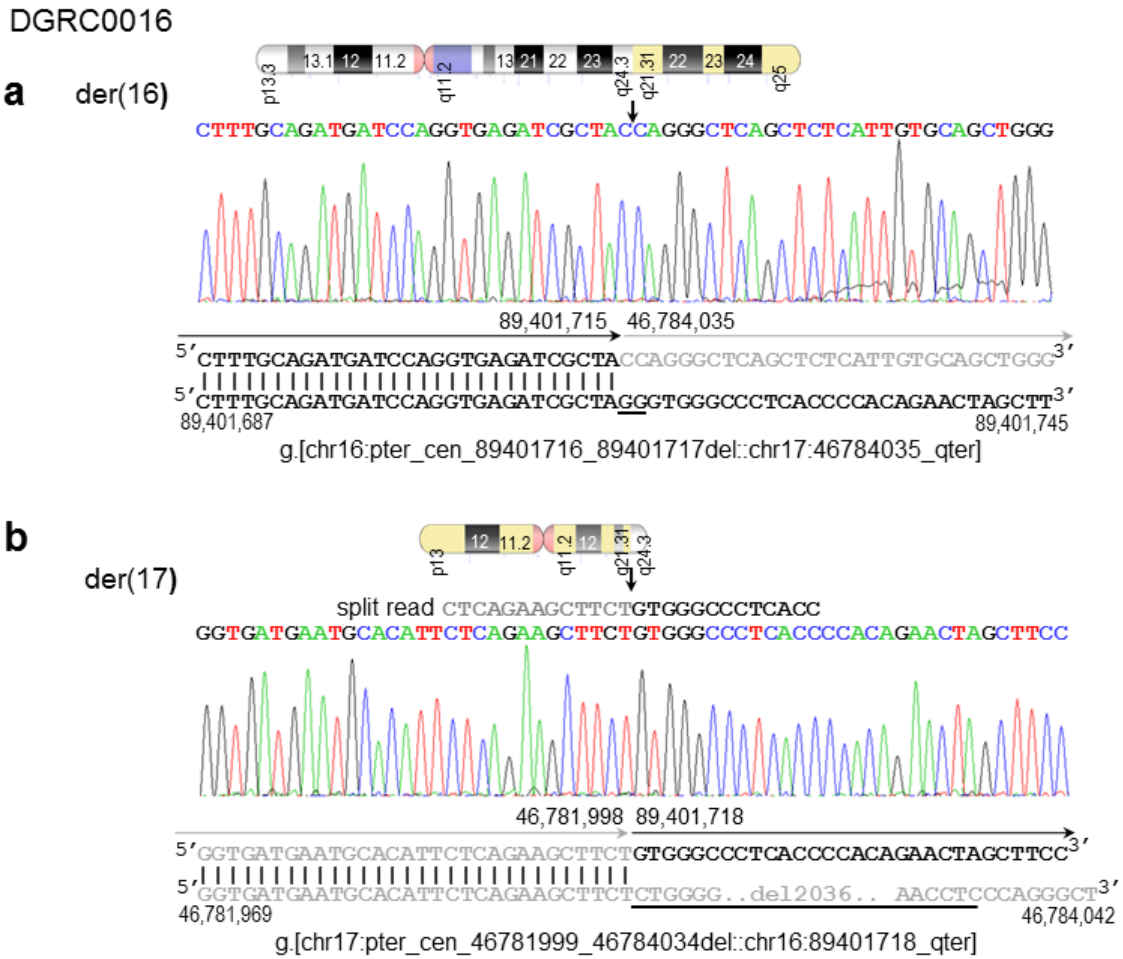
4 years



8 years

Supplementary Fig. 1 Evolution of DGRC0025 phenotypic facial features. Glabellar hemangioma, thin upper lip vermilion, depressed nasal bridge and slightly anteverted nares can be seen throughout time. Synophrys became evident from age 8.

Supplementary Fig. 2 Evolution of DGRC0030 phenotypic facial features from 3 months to 11 years of age. Coarse facial features, full/broad nasal tip, depressed nasal bridge and anteverted nares are evident throughout time. Tented, protruding upper lip and wide spaced teeth can be seen beginning at 4 years of age.



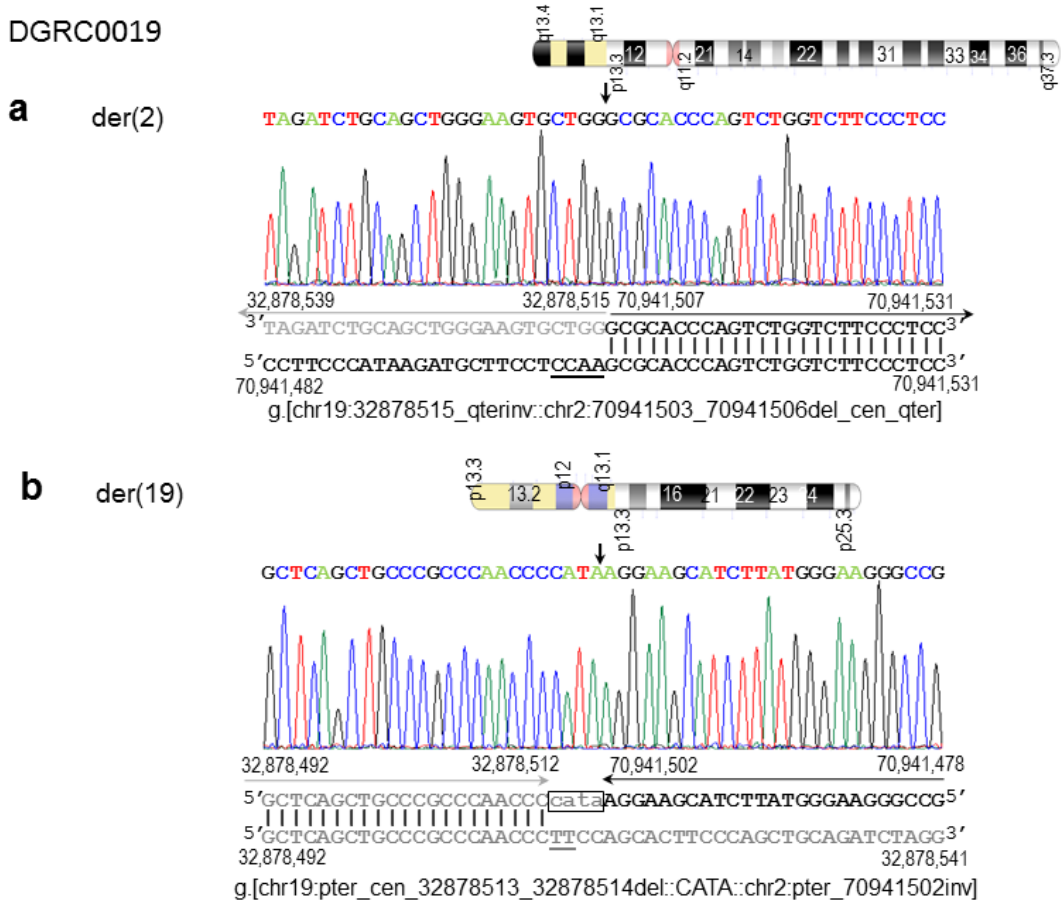
Supplementary Fig. 3 Nucleotide sequence of der(16) and der(17) breakpoints aligned against the GRCh38 reference human genome

a Chromosome 16q24.3 breakpoint. The chromosome 16 sequence is in black, whereas the chromosome 17 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. The GG dinucleotide deleted from the der(16) breakpoint is underlined.

b Chromosome 17q21.31 breakpoint. The chromosome 16 sequence is in black, whereas the chromosome 17 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes.

The 2,036 bp del (g.46781999_46784034del) identified at the der(17) breakpoint is underlined. Therefore, the translocation is classified as unbalanced. The translocation is revised and described as seq[GRCh38] 46,XX,t(16;17)(16pter→16q24.3::17q21.31→17qter;17pter→17q21.31::16q24.3→16qter)dn. According to next-gen cytogenetics nomenclature (Ordulu et al. 2014), the translocation is described as 46,XX,t(16;17)(q24;q21.3)dn.seq[GRCh38] t(16;17)(16pter→16q24.3(89,401,715)::17q21.31(46,784,035)→17qter;17pter→17p21.31(46,781,998::16q24.3(89,401,718)→16qter)dn.

NCBI-GenBank accession numbers of the der(16) and der(17) junction fragment sequences are MH843735 and MH843736, respectively.



Supplementary Fig. 4 Nucleotide sequence of der(2) and der(19) breakpoints aligned against the GRCh38 reference human genome

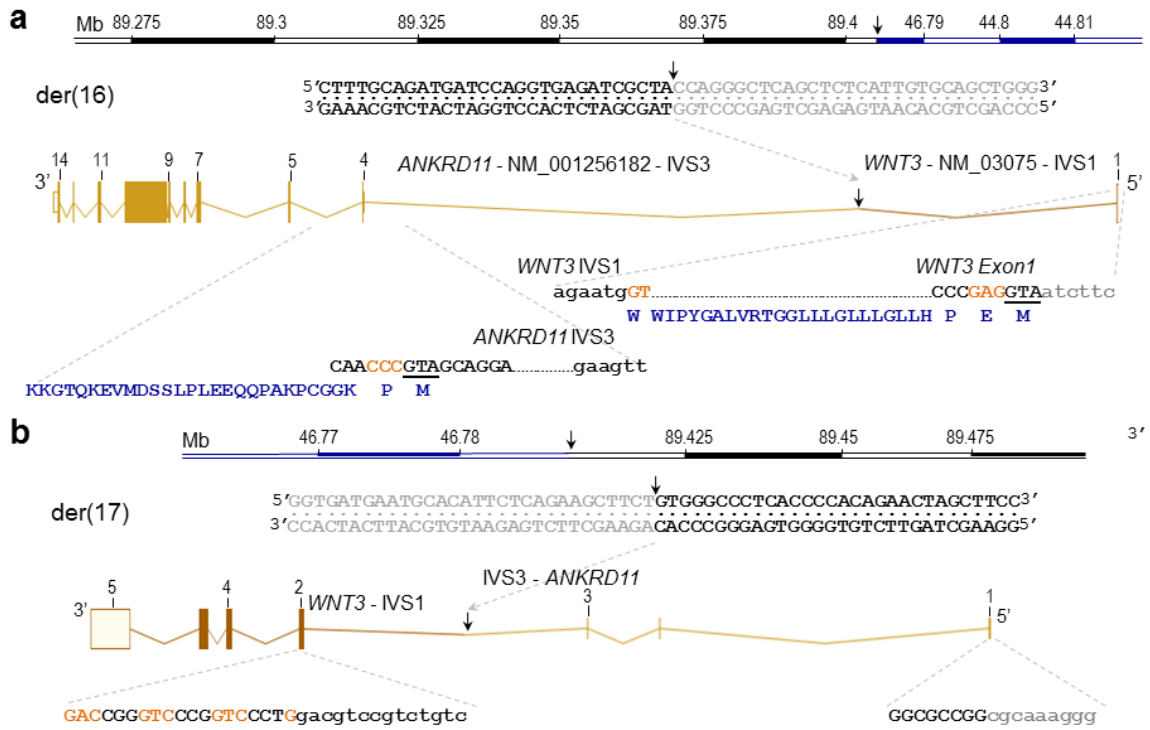
a Chromosome 2p13.3 breakpoint. The chromosome 2 sequence is in black, whereas the chromosome 19 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. The deleted tetranucleotide CCAA sequence from the der(2) breakpoint is underlined.

b Chromosome 19q13.11 breakpoint. The chromosome 2 sequence is in black, whereas the chromosome 19 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. The deleted TT dinucleotide from the der(19) breakpoint is underlined, whereas the inserted CATA tetranucleotide sequence is boxed and in lowercase.

The translocation is revised and described as seq[GRCh38] 46,XY,t(2;19)(19qter→19q13.11::2p13.3→2qter;19pter→19q13.11::2p13.3→2pter)dn.

According to next-gen cytogenetics nomenclature (Ordulu et al. 2014), the translocation can be defined as 46,XY,t(2;19)(p13;q13.11)dn.seq[GRCh38] t(2;19)(19qter(-)→19q13.11(32,878,515)::2p13.3(+)(70,941,507)→2qter;19pter→19q13.11(+)(32,878,512)::CATA::2p13.3(-)(70,941,502)→2pter)dn.

NCBI-GenBank accession numbers of the der(2) and der(19) junction fragment sequences are MH843737 and MH843738, respectively.

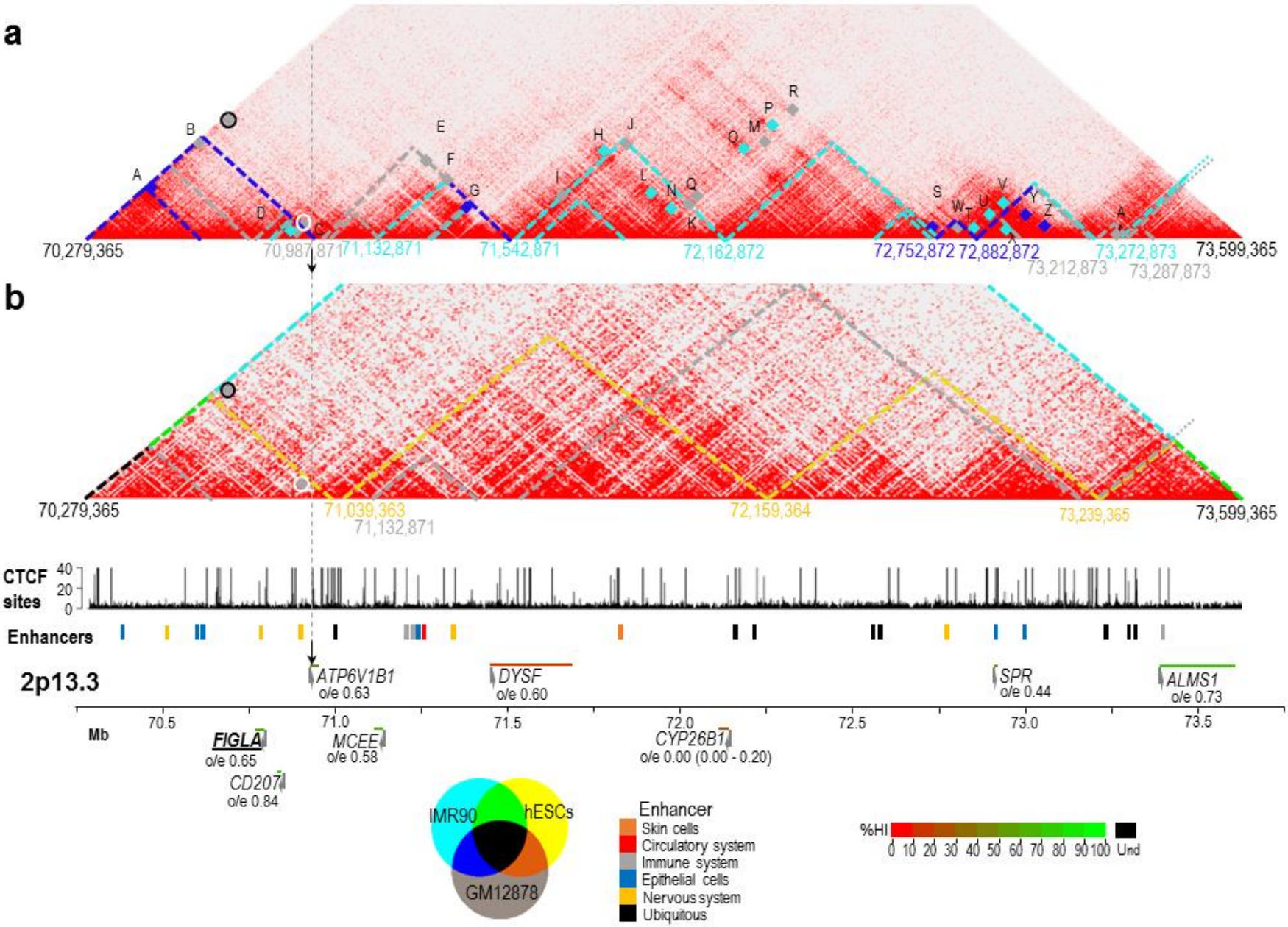


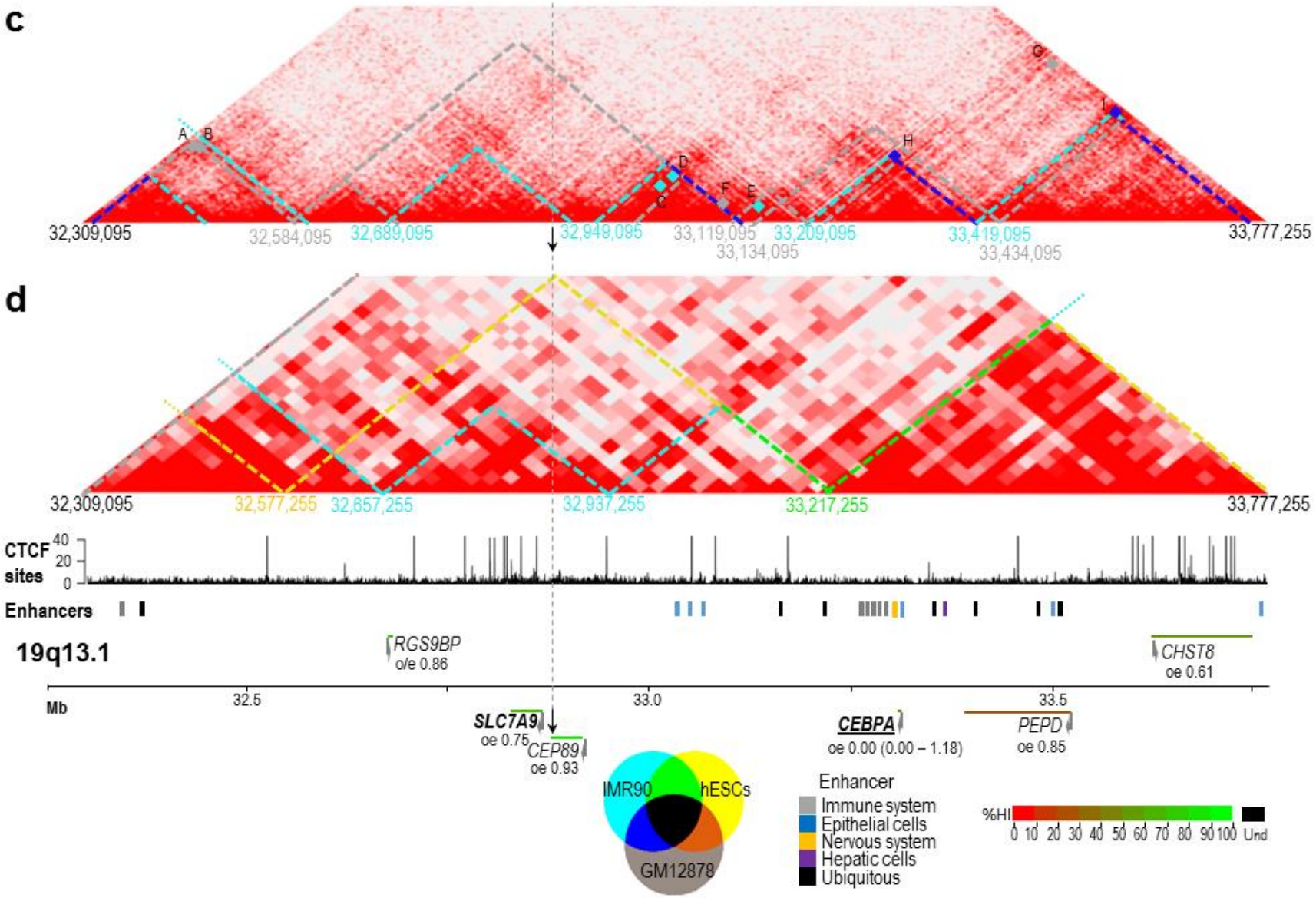
Supplementary Fig. 5 Fusion genes and hypothetical transcripts resulting from the t(16;17)(q24.3;q21.31)dn

a Diagram depicting the fusion gene at der(16) between *WNT3* IVS1 and *ANKRD11* IVS3. At the top, the physical map across the der(16) breakpoint region with the corresponding junction sequence is shown underneath (chromosome 16 sequence is in black, whereas the chromosome 17 sequence is in gray). Below, schematic representation of the fusion gene between *WNT3* exon 1 and *ANKRD11* exons 4-14 is shown. Theoretically, *WNT3* exon 1 may code for a truncated 26 amino acid polypeptide, while the translational initiation codon within exon 4 of *ANKRD11* is intact. Hypothetical translation products are indicated in blue.

b Diagram depicting the fusion gene at der(17) between *ANKRD11* IVS3 and *WNT3* IVS1. At the top, the physical map across the der(17) breakpoint region with the corresponding junction sequence is shown underneath. Below, schematic representation of the fusion gene between *ANKRD11* exons 1-3 and *WNT3* exons 2-5 is shown. The *ANKRD11* non-coding exons 1-3 are followed by coding exon 2 of *WNT3*.

Breakpoint positions are indicated by vertical arrows. Exons are numbered. Exonic and intronic sequences are in upper and lower-case letters, respectively. The 5' untranslated sequences are in lower case gray letters. Coding triplets are highlighted in brown and black and the translational initiation codons (ATG) are underlined.


a
b
70,279,365
70,987,871
71,132,871
71,542,871
72,162,872
72,752,872
72,882,872
73,212,873
73,272,873
73,287,873
73,599,365
71,039,363
72,159,364
73,239,365
CTCF sites
Enhancers
2p13.3
ATP6V1B1 o/e 0.63
DYSF o/e 0.60
SPR o/e 0.44
ALMS1 o/e 0.73
Mb
FIGLA o/e 0.65
MCEE o/e 0.58
CYP26B1 o/e 0.00 (0.00 - 0.20)
CD207 o/e 0.84
IMR90
hESCs
GM12878
Enhancer
Skin cells
Circulatory system
Immune system
Epithelial cells
Nervous system
Ubiquitous
%HI
Und


c
A
B
C
D
E
F
G
H
I
32,309,095
32,584,095
32,689,095
32,949,095
33,119,095
33,134,095
33,209,095
33,419,095
33,434,095
33,777,255
d
32,309,095
32,577,255
32,657,255
32,937,255
33,217,255
33,777,255
CTCF sites
40
20
0
Enhancers
19q13.1
RGS9BP
o/e 0.86
CHST8
oe 0.61
Mb
32.5
33.0
33.5
SLC7A9
oe 0.75
CEP89
oe 0.93
CEBPA
oe 0.00 (0.00 – 1.18)
PEPD
oe 0.85
IMR90
hESCs
GM12878
Enhancer
Immune system
Epithelial cells
Nervous system
Hepatic cells
Ubiquitous
%HI
0 10 20 30 40 50 60 70 80 90 100 Und

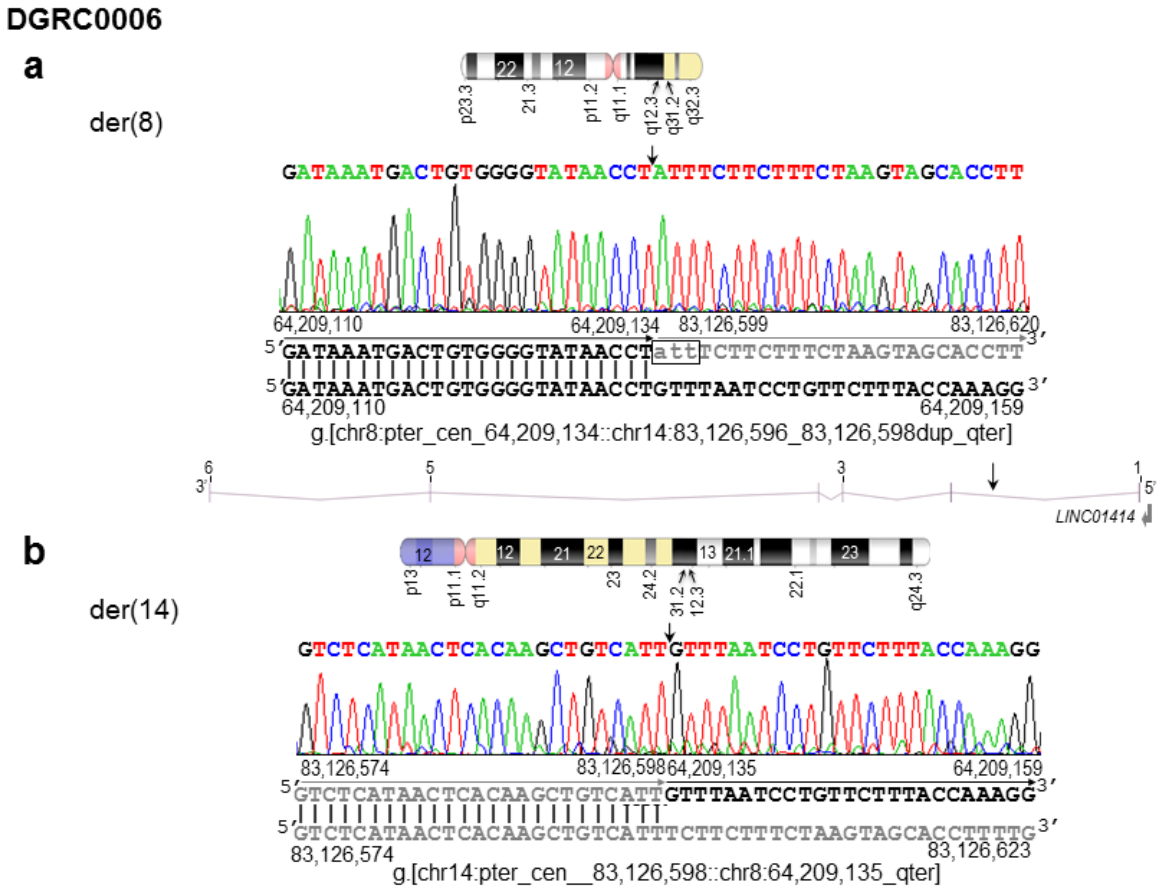
Supplementary Fig. 6 Chromatin interaction heatmaps of the t(2;19)(p13.3;q13.11)dn breakpoint regions for GM12878 lymphoblastoid cell lines and IMR90 fibroblasts, at different resolutions

a, c Interaction data from GM12878 lymphoblastoid cell lines (LCLs) at 5 kb resolution from Rao et al. (2014).

b, d Interaction data at 25 kb resolution from IMR90 fibroblasts from Dixon et al. (2012).

TADs and boundaries for GM12878 and IMR90 are indicated according to Rao et al. (2014). The dashed lines overlaid on the heatmaps are color-coded according to cell line and overlap between these lines. Chromatin loops are depicted by overlaid squares, color-coded by cell line and identified by letters. Loops circled in black and white are disrupted by the breakpoint (Rao et al. 2014). The one circled in black is between two 10 kb bins, at chr2:70,292,868-70302868, encompassing *FAM136A* and the promoter - exon 1 region of *SNRPG*, and, at chr2:71,062,870-71,072,870, encompassing the promoter - exons 1-5 region of *NAGK*. The second loop, circled in white, is between two 10 kb bins localized at chr2:70,862,870-70,872,870 and chr2:70,942,870-70,952,870, encompassing exon 2 of *ATP6V1B1*.None of these genes shows LoF sensitivity.

The black arrows indicate the breakpoint positions. Below is a track depicting CCCTC-binding factor (CTCF) sites from the analyzed region according to the Chip-seq track in IMR90 fibroblasts CTCF IgG-rab ChIP-seq Signal from ENCODE/SYDH). Further below are transcribed human enhancers from the reference genomic region, color-coded according to their tissue/cell specific expression. Beneath the enhancers is shown the gene map across the analyzed genomic region. Forward and reverse arrowheads indicate the position of genes in sense and antisense orientation. Genes are color-coded according to their HI scores. Genes presently associated with autosomal dominant disorders are underlined and bolded, whereas genes associated with both autosomal recessive and dominant disorders are only bolded. The LoF intolerance, expressed as observed / expected (oe) number of LoF variants, are stated below the genes.

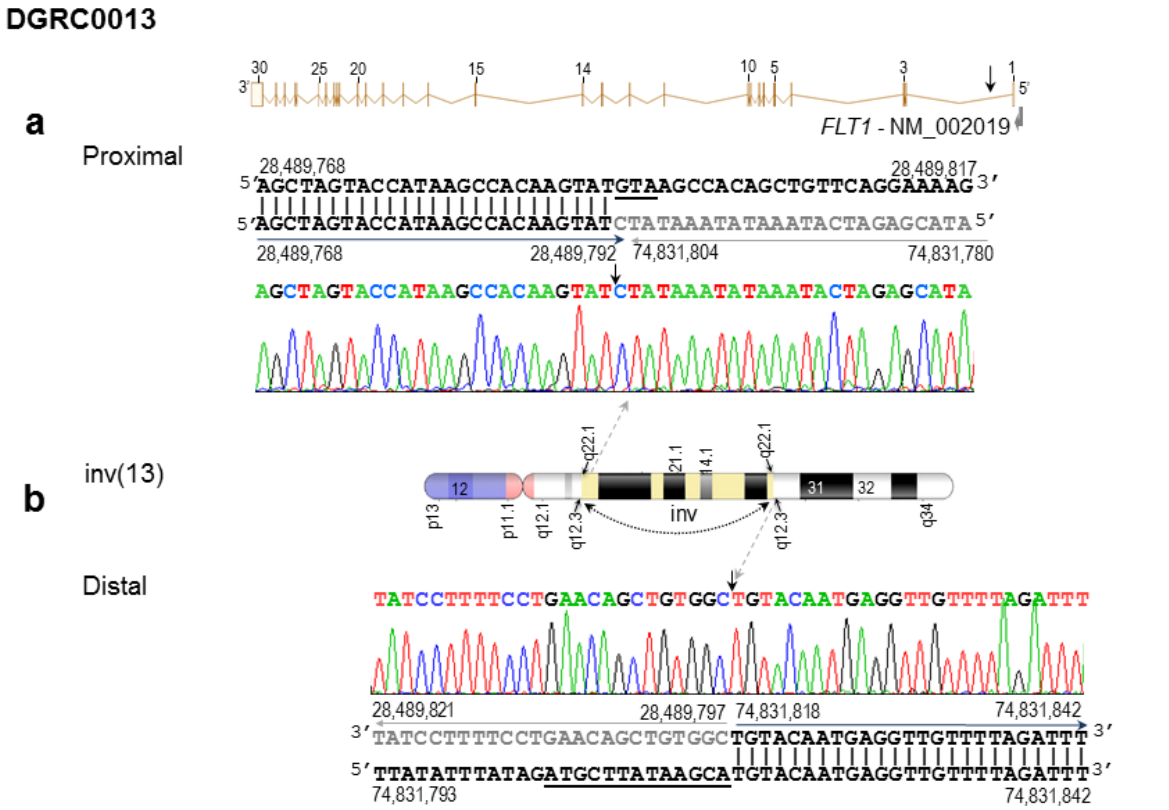


Supplementary Fig. 7 Nucleotide sequence of der(8) and der(14) breakpoints aligned against the GRCh38 reference human genome

a Chromosome 8q12.3 breakpoint. The chromosome 8 sequence is in black, whereas the chromosome 14 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. The duplicated trinucleotide ATT sequence is boxed and in lowercase. Below, arrow within IVS 1 of *LINC01414* indicates the position of the translocation breakpoint.

b Chromosome 14q31.2 breakpoint. The chromosome 8 sequence is in black, whereas the chromosome 14 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. A dashed line underlines the duplicated trinucleotide.

The translocation is revised and described as seq[GRCh38] 46,XX,t(8;14)(8pter→8q12.3::14q31.2→14qter;14pter→14q31.2::8q12.3→8qter)dn. According to next-gen cytogenetics nomenclature (Ordulu et al. 2014), the translocation can be defined as 46,XX,t(8;14)(q11.23;q24.3)dn.seq[GRCh38] t(8;14)(8pter->q12.3(64,209,134):: ATT::14q31.2(83,126,599)->14qter;14pter->14q31.2(83,126,598)::8q12.3(64,209,135)->8qter).



Supplementary Fig. 8 Nucleotide sequence of the paracentric chromosome 13 inversion breakpoints aligned against the GRCh38 reference human genome

a The proximal inversion breakpoint at 13q12.3.

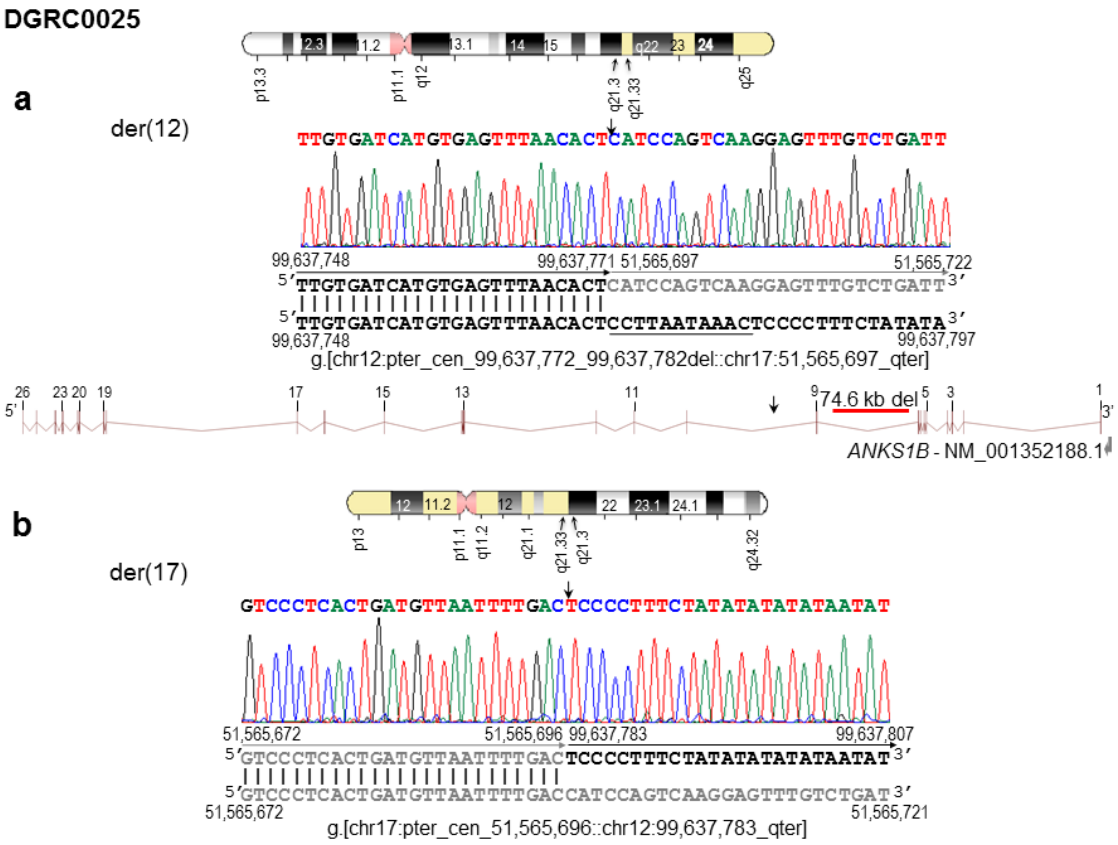
b The distal inversion breakpoint at 13q22.1.

Above, arrow within *FLT1* IVS 1 indicates the position of the proximal inversion breakpoint.

Reference chromosome 13 sequences are in black, whereas the inverted chromosome 13 sequence is in gray. Vertical lines indicate identical nucleotides between the derivative and reference chromosome. The deleted GTA trinucleotide and the tridecanucleotide ATGCTTATAAGCA sequence from proximal and distal breakpoints, respectively, are underlined.

The inversion is revised and described as seq[GRCh38] 46,XY,inv(13)(pter→q12.3::q22.1→q12.3::q22.1→qter)dn.

According to next-gen cytogenetics nomenclature (Ordulu et al. 2014), the inversion is defined as 46,XY,inv(13)(q12.3q23)dn.seq[GRCh38] inv(13)(pter->q12.3(28,489,792)::q22.1(-)(74,831,804)->q12.3(28,489,795)::q22.1(+)(74,831,818)->qter).



Supplementary Fig. 9 Nucleotide sequence of der(12) and der(17) breakpoints aligned against the GRCh38 reference human genome

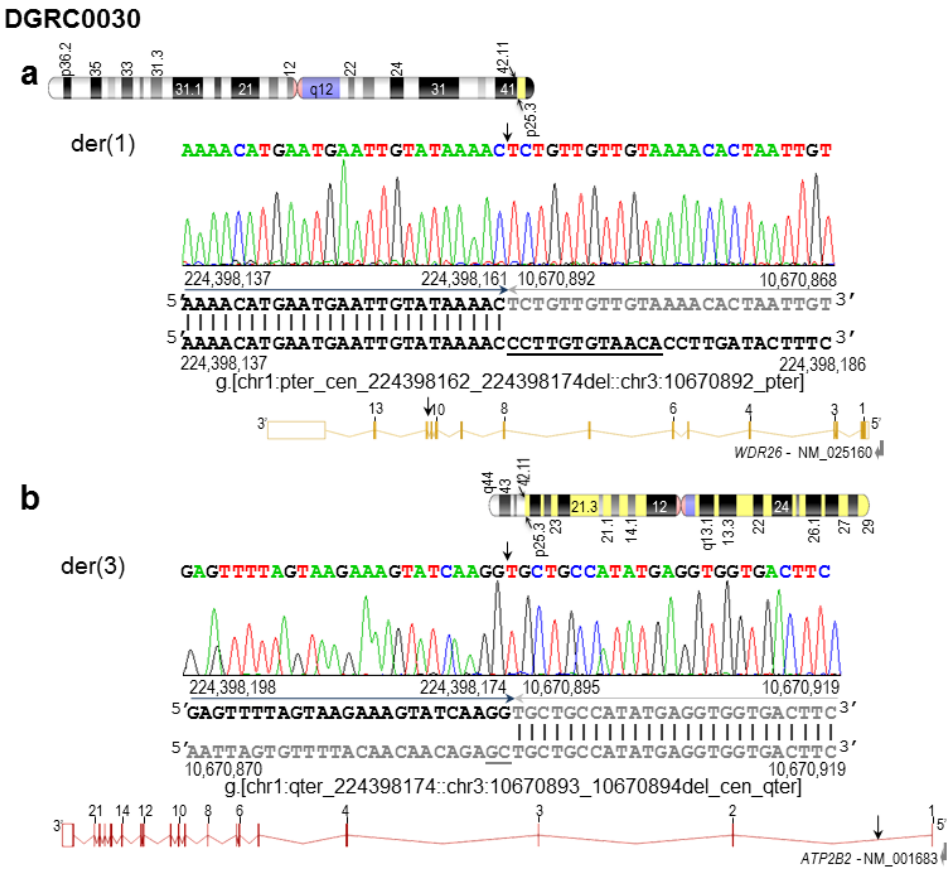
a Chromosome 12q23.1 breakpoint. The chromosome 12 sequence is in black, whereas the chromosome 17 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. The deleted hendecanucleotide CCTTAATAAAC sequence from the der(12) sequence is underlined.

Below, an arrow within IVS 9 and a red bar within IVS 8 of *ANKRD11* indicate the position of the translocation breakpoint and of a 74.6 Kb deletion, respectively.

b Chromosome 17q21.33 breakpoint. The chromosome 12 sequence is in black, whereas the chromosome 17 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes.

The translocation is revised and described as seq[GRCh38] 46,XY,t(12;17)(12pter→12q21.3::17q21.33→17qter;17pter→17q21.33::12q21.3→12qter)dn.

According to next-gen cytogenetics nomenclature (Ordulu et al. 2014), the translocation can be defined as 46,XY,t(12;17)(q23;q22)dn.seq[GRCh38] t(12;17)(12pter->q23.1(99,637,771)::q23.1(99,637,782)::17q21.33(51,565,697)->17qter;17pter->q21.33(51,565,696)::12q23.1(99,637,783-99,680,962::99,755,562)->12qter)

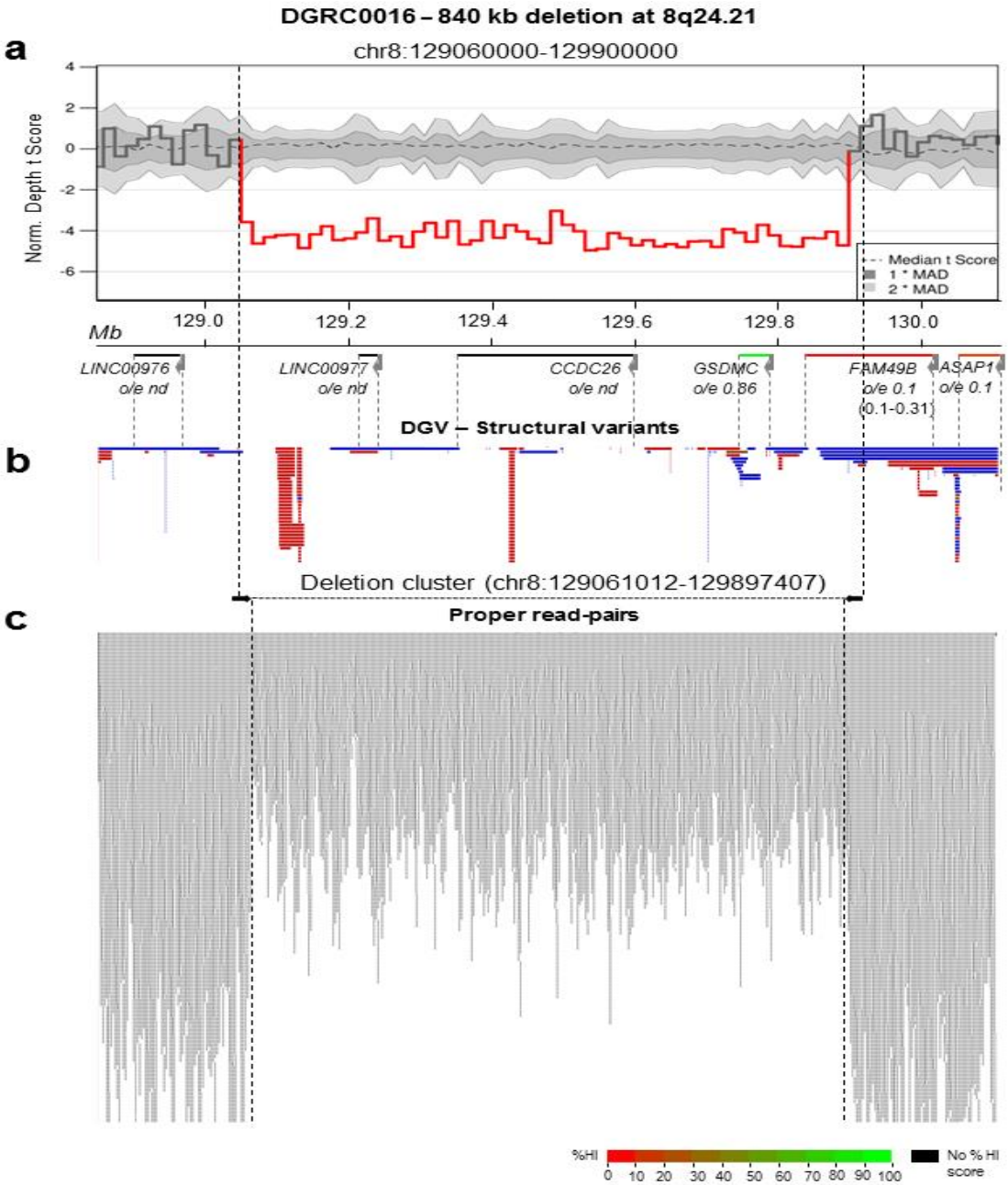


Supplementary Fig.10 Nucleotide sequence of der(1) and der(3) breakpoints aligned against the GRCh38 reference human genome

a Chromosome 1q42.11 breakpoint. The chromosome 1 sequence is in black, whereas the chromosome 3 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. The deleted dodecanucleotide CCTTGTGTAACA sequence from der(1) sequence is underlined. Below, arrow within exon 12 of *WDR12* indicates the translocation breakpoint.

b Chromosome 3p25.3 breakpoint. The chromosome 1 sequence is in black, whereas the chromosome 3 sequence is in gray. Vertical lines indicate identical nucleotides between derivative and reference chromosomes. The deleted GC dinucleotide from der(3) sequence is underlined. Below, arrow within IVS 1 *ATP2B2* indicates the translocation breakpoint.

The translocation is revised and described as seq[GRCh38] 46,XY,t(1;3)(1pter→1q42.11::3p25.3→3pter;1qter→1q42.11::3p25.3→3qter)dn. According to next-gen cytogenetics nomenclature (Ordulu et al. 2014), the translocation can be defined as 46,XY,t(1;3)(q42;p25)dn.seq[GRCh38] t(1;3)(1pter->q42.11(224,398,161)::3p25.3(-)(10,670,892)->3pter;1qter->q42.11(-)(224,398,174)::3p25.3(+)(10,670,895)->3qter)

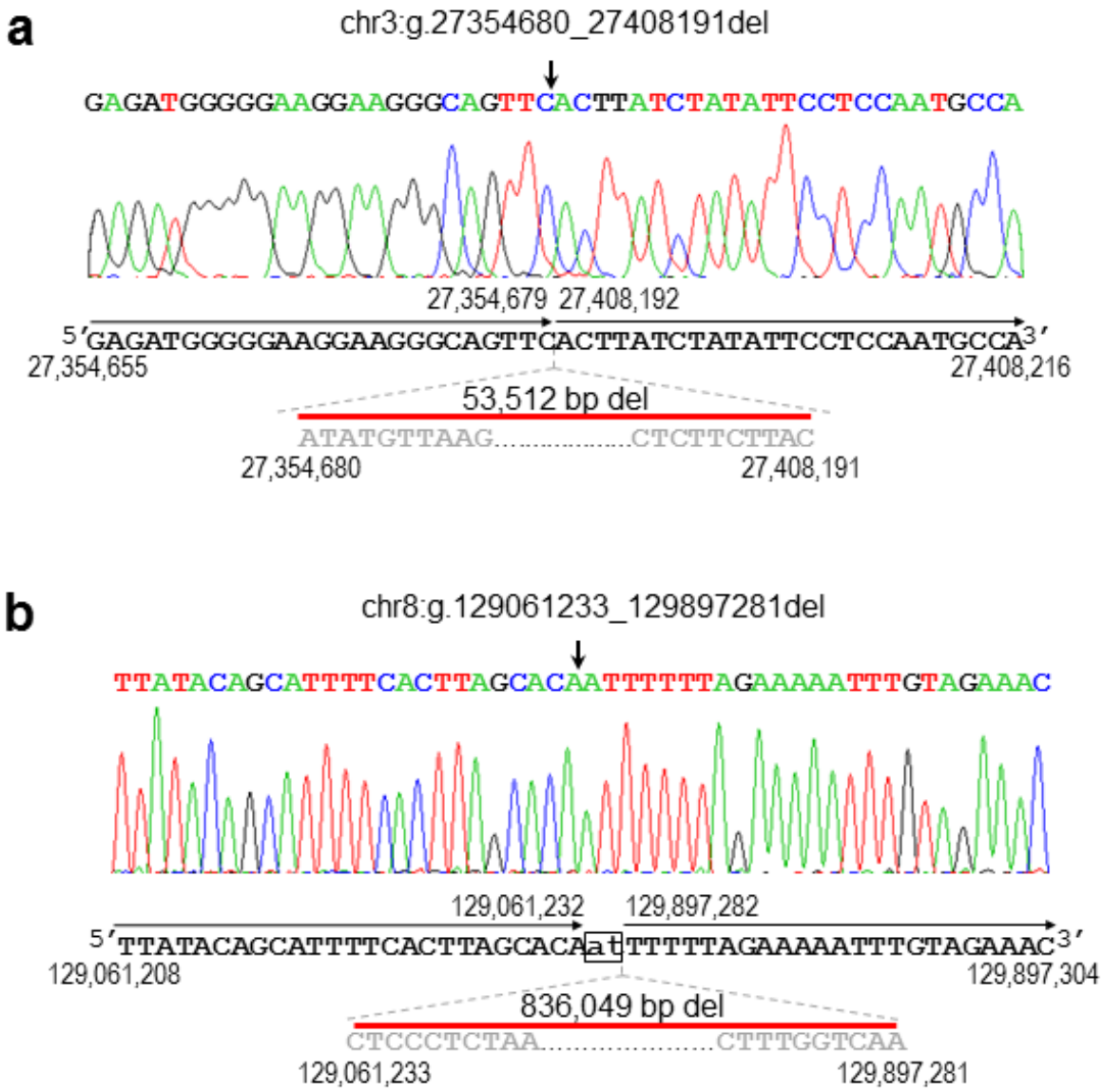


Supplementary Fig. 11 Illustration of DGRC0016-specific genomic imbalance identified at 8q24.21 by sequence coverage and read-pair cluster analysis

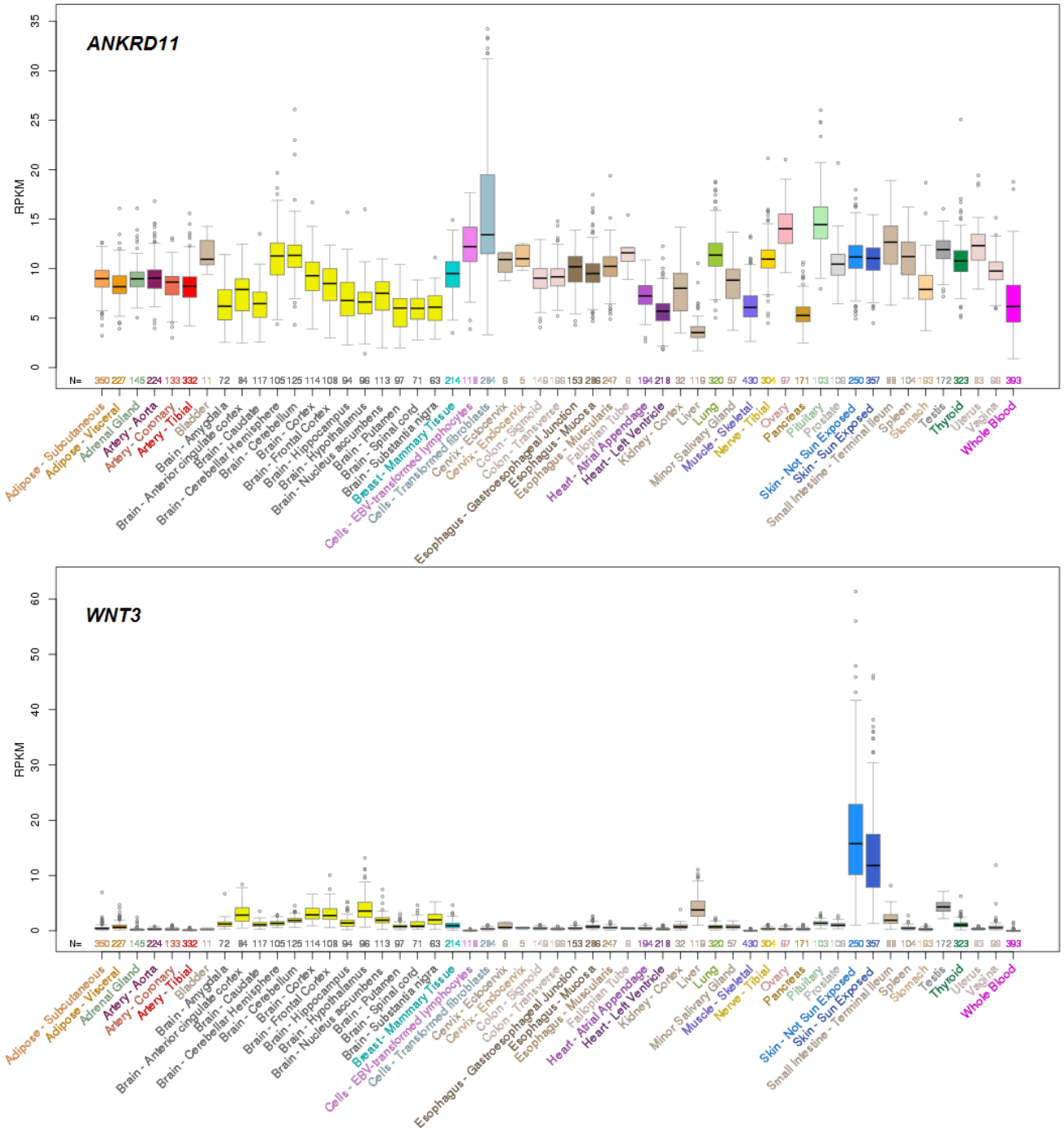
a Genomic imbalance plot generated by CNView (Collins et al. 2016) showing an 840 kb del. A group of 32 cases, analyzed using the same type of liGS library, was used as control. The horizontal black dashed line with darker and lighter gray shading indicate median coverage and deviation, respectively. Regions with a statistically significant decrease in sequence coverage (α=0.05, Bonferroni correction) indicating dels are depicted in red. Gene map across this region is shown below. Forward and reverse folded arrows indicate the position of the genes in sense and antisense orientation, respectively. Genes are color coded according to their haploinsufficiency index (HI) and LoF intolerance, expressed as observed / expected (oe) number of LoF variants, are stated below the genes. Confidence interval of *FAM49B* oe is indicated in parentheses.

b CNVs from the affected genomic region reported in the DGV database (MacDonald et al. 2014). Blue and red bars represent gains and losses, respectively. Vertical dashed black lines delimitate the overlap between the del with gene and CNVs.

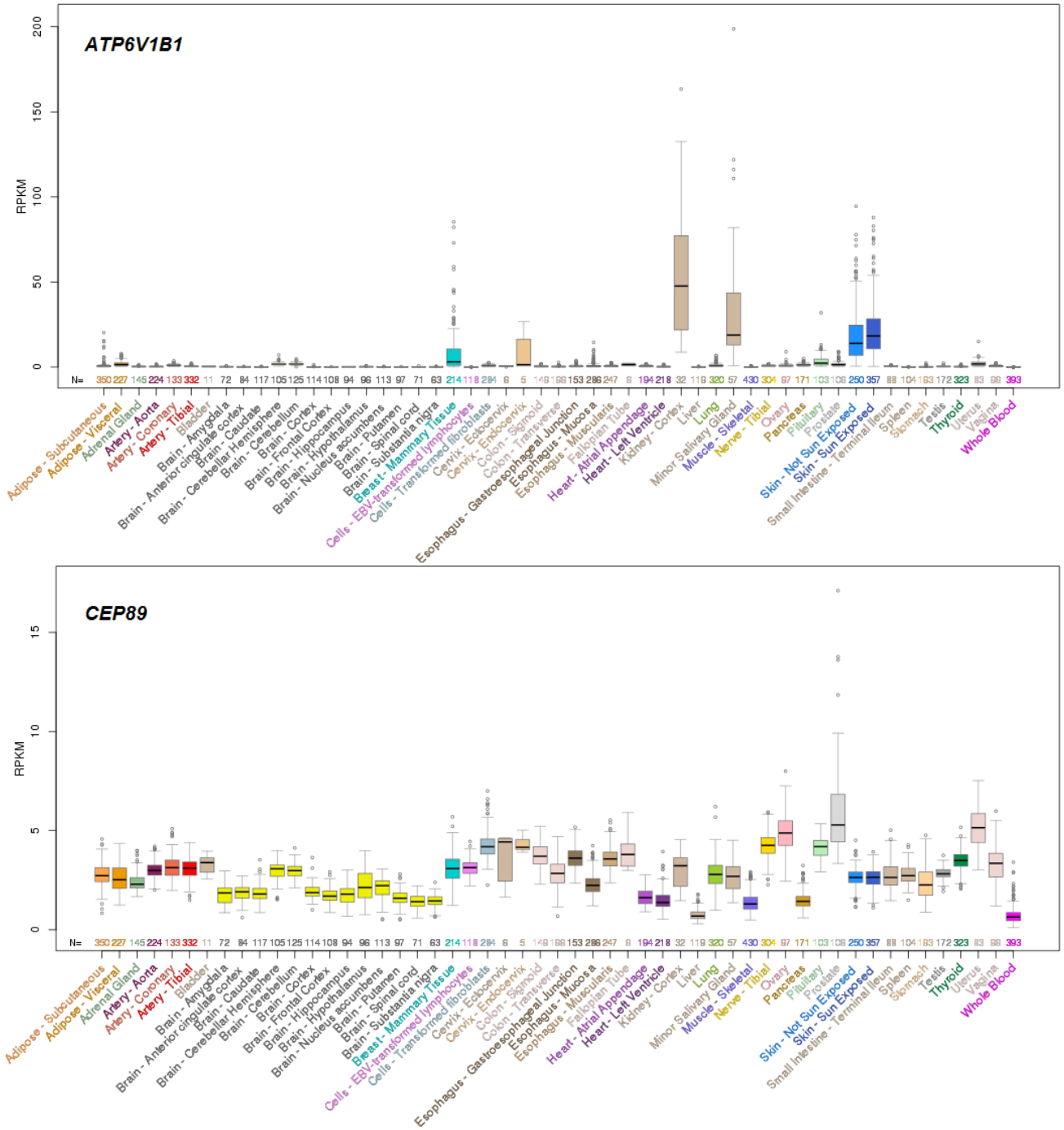
c Read-pair cluster delimiting the del and proper read-pairs aligned within the involved genomic region. Black arrows depict inward-facing read-pair cluster delimiting the del. Below, small gray vertical bars denote proper read-pairs mapped within this region, illustrating the decrease in coverage within the deleted region.



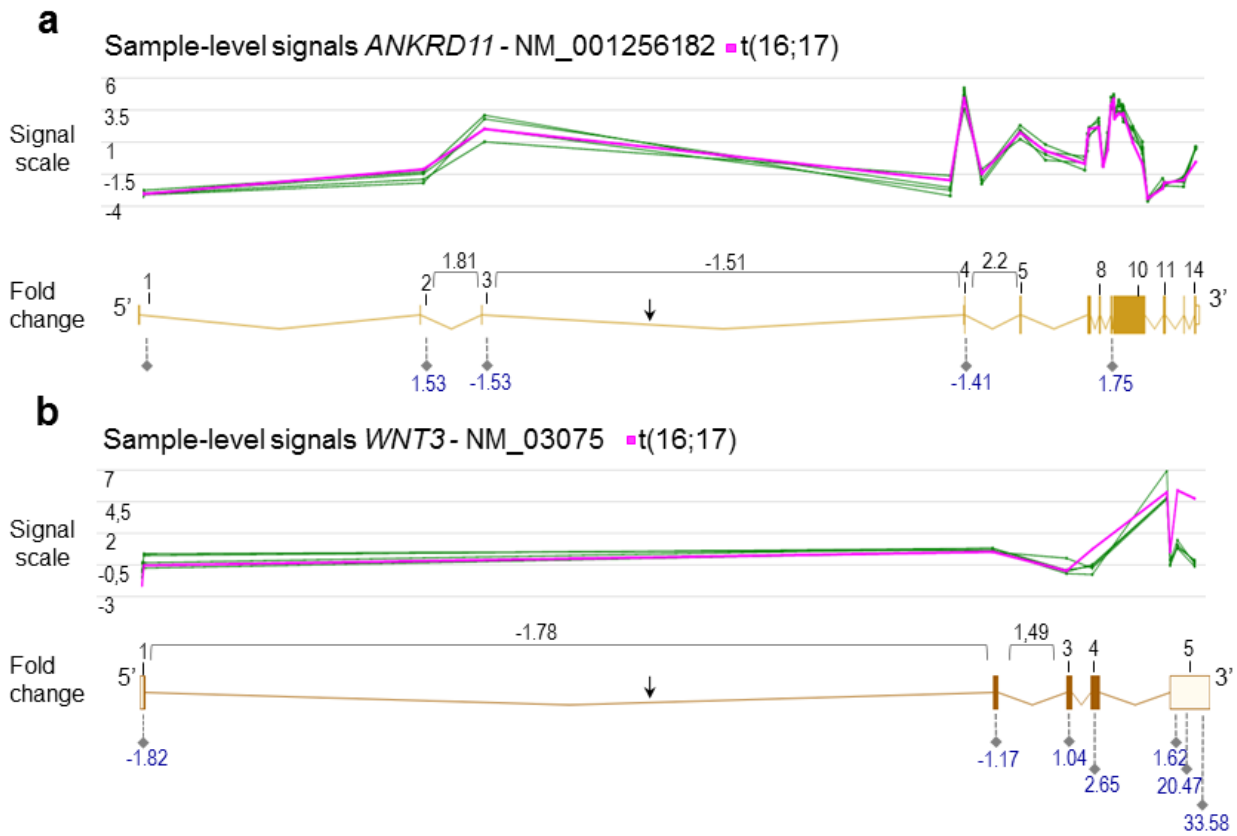
Supplementary Fig. 12 Nucleotide sequence of genomic imbalance breakpoints at 3p24.1 and 8q24.21 identified in DGRC0016
a The 53,512 bp del breakpoint at 3p24.1.
b The 836,049 bp del breakpoint at 8q24.21.
Vertical arrows indicate positions of del breakpoints, whereas horizontal arrows denote orientation of the genomic sequence. Deleted sequences at breakpoint junctions are in gray and dels are depicted by red bars. The inserted AT dinucleotide at the breakpoint is boxed and in lowercase.



Supplementary Fig. 13 Tissue-specific expression profiles of *ANKRD11* and *WNT3* disrupted by the 16q24.3 and 17q21.31 breakpoints, respectively (from Genotype-Tissue Expression, GTEx)



Supplementary Fig. 14 Tissue-specific expression profiles of *ATP6V1B1* and *CEP89* (from Genotype-Tissue Expression, GTEx)
The 2p13.3 breakpoint is within IVS1 of *ATP6V1B1* whereas that of 19q13.11 is within the 3' UTR of *CEP89*.



Supplementary Fig. 15 Differential expression at exon and exon junction levels of the t(16;17) disrupted genes
a Gene structure-based display of sample-level signal intensities of the disrupted *ANKRD11*. Below, schematic gene map with exons and exon junction fold changes.
b Gene structure-based display of sample-level signal intensities of the disrupted *WNT3*. Below, schematic gene map with exons and exon junction fold changes.
The detection threshold of HTA 2.0 microarray is a two-fold change. Control LCL samples are in green and the proband's LCL sample is in purple. Exons are numbered. Under the gene maps, dashed lines with filled diamonds indicate exons and the corresponding probe set fold change between the proband and the control samples. Above the gene map, horizontal square brackets show exon junction fold changes between the designated exons. Arrows indicate the breakpoints.

Supplementary Tables

Supplementary Table 1 Primers used for validation of structural chromosomal abnormalities and CNVs

Fragments	Designation	Primer sequence (5'-3')	Primer localization [a]	Annealing T (C°)	Amplicon length (bp)
		DGRC0016			
t(16;17)(q24.3;q21.31)					
Control fragment chr16	ANKRD11_IVS3-1F	CCACCTCCCATCCACACCT	Chr16:89,401,547-89,401,566	61	295
	ANKRD11_IVS3-2R	GTACCCAGAGAGGGGTCAGC	Chr16:89,401,821-89,401,840		
Control fragment chr17	WNT3_IVS1-3F	TAGCTGGGCTTTGGAATATCAT	Chr17:46,783,899-46,783,920	61	520
	WNT3_IVS1-2R	GCAACACAAACCTCTACCCCTA	Chr17:46,784,399-46,784,418		
Junction fragment der(16)	ANKRD11_IVS3-1F	CCACCTCCCATCCACACCT	Chr16:89,401,547-89,401,566	61	559
	WNT3_IVS1-2R	GCAACACAAACCTCTACCCCTA	Chr17:46,784,399-46,784,418		
Junction fragment der(17)	WNT3_IVS1-1F	CAGCCTACCCCTTACTTTTCAC	Chr17:46,781,765-46,781,785	61	357
	ANKRD11_IVS3-2R	GTACCCAGAGAGGGGTCAGC	Chr16:89,401,821-89,401,840		
seq[GRCh38] chr3:g.27354680_27408191del					
Control fragment proximal	SLC4A7_IVS13-1F	ACACACTCTGGGACCGTATCT	Chr3:27,407,917-27,407,937	63	964
	SLC4A7_IVS13-2R	CTGTGGCTTTTTGAAGTGGCAT	Chr3:27,408,859-27,408,880		
Control fragment distal	NEK10_IVS2-1F	CCAAGTCTCTCAGCACAGGAG	Chr3:27,353,627-27,353,647	63	606
	NEK10_IVS2-2R	GCAGAAGAAGCCTGTGAGCTT	Chr3:27,354,222-27,354,242		
Deletion junction fragment	NEK10_IVS2-1F	CCAAGTCTCTCAGCACAGGAG	Chr3:27,353,627-27,353,647	63	1,741
	SLC4A7_IVS13-2R	CTGTGGCTTTTTGAAGTGGCAT	Chr3:27,408,859-27,408,880		
seq[GRCh38] chr8:g.129061233_129897281del					
Control fragment proximal	AC104256-3F	TAGGGAAACCAAGTTCAGGCTC	Chr8:129,060,775-129,060,796	60	681
	AC104256-4R	GGTGAAACACAGATGATGCTCC	Chr8:129,061,434-129,061,455		
Control fragment distal	FAM49B-1F	ACCTCCCCAGAGTGAGAGTC	Chr8:129,896,671-129,896,689	60	1,010
	FAM49B-3R	TATAGGCATTAGCCGCCCAC	Chr8:129,897,660-129,897,679		
Deletion junction fragment	AC104256-3F	TAGGGAAACCAAGTTCAGGCTC	Chr8:129,060,775-129,060,796	60	858
	FAM49B-3R	TATAGGCATTAGCCGCCCAC	Chr8:129,897,660-129,897,679		
		DGRC0019			
t(2;19)(p13.3;q13.11)					
Control fragment chr2	ATP6V1B1-1F	AGGTGTGAGCCACTGTAGCTG	Chr2:70,941,481-70,941,502	61	635
	ATP6V1B1-2R	GACTCACACTCCTCGCTCTCAG	Chr2:70,941,807-70,941,828		
Control fragment chr19	A008805-1F	AGCAGCACCTTGTTTTTTTTG	Chr19:32,878,365-32,878,385	61	287
	A008805-2R	TCTGTTTTAGCCAGGGCATG	Chr19:32,878,632-32,878,651		
Junction fragment der(2)	ATP6V1B1-2R	GACTCACACTCCTCGCTCTCAG	Chr2:70,941,807-70,941,828	61	459
	A008805-2R	TCTGTTTTAGCCAGGGCATG	Chr19:32,878,632-32,878,651		
	A008805-1F	AGCAGCACCTTGTTTTTTTTG	Chr19:32,878,365-32,878,385		

Junction fragment	ATP6V1B1-1F	AGGTGTGAGCCACTGTAGCTG	Chr2:70,941,481-70,941,502	61	461
		DGRC0006			
t(8;14)(q12.3 ;q31.2)					
Control fragment chr8	AC023533-1F	TGGAGAGCAAGGCAAACAGT	Chr8:64,208,781-64,208,800	58	570
	AC023533-2R	CCCTTGAAACTCCCCAGACT	Chr8:64,209,331-64,209,350		
Control fragment chr14	AL162872-1F	AGATGTTTGGCGTAGGCTTG	Chr14:83,126,183-83,126,202	58	614
	AL162872-3R	CACAGGGATTTGGTTTATCATTCCT	Chr14:83,126,626-83,126,650		
Junction fragment der(8)	AC023533-1F	TGGAGAGCAAGGCAAACAGT	Chr8:64,208,781-64,208,800	58	555
	AL162872-3R	CACAGGGATTTGGTTTATCATTCCT	Chr14:83,126,626-83,126,650		
Junction fragment der(14)	AL162872-1F	AGATGTTTGGCGTAGGCTTG	Chr14:83,126,183-83,126,202	58	632
	AC023533-2R	CCCTTGAAACTCCCCAGACT	Chr8:64,209,331-64,209,350		
		DGRC0013			
inv(13)(q12.3q22.1)					
Control fragment proximal chr13	FLT1-IVS1-1F	AGAGATGGTGATTACTGGGCTG	Chr13:28,489,561-28,489,582	60	1,154
	FLT1-IVS1-3R	ACTGGTGTATTTCCTCTTCCC	Chr13:28,490,028-28,490,048		
Control fragment distal chr13	AL590063-1F	TGAATCCACAGCATCTCCAAG	Chr13:74,831,722-74,831,742	60	487
	AL590063-2R	TGTTTCAGGGTTGCTTTCTTG	Chr13:74,832,041-74,832,061		
Junction fragment proximal chr13	FLT1-IVS1-1F	AGAGATGGTGATTACTGGGCTG	Chr13:28,489,561-28,489,582	60	315
	AL590063-1F	TGAATCCACAGCATCTCCAAG	Chr13:74,831,722-74,831,742		
Junction fragment distal chr13	AL590063-2R	TGTTTCAGGGTTGCTTTCTTG	Chr13:74,832,041-74,832,061	60	496
	FLT1-IVS1-3R	ACTGGTGTATTTCCTCTTCCC	Chr13:28,490,028-28,490,048		
		DGRC0025			
t(12;17)(q23.1q21.33)					
Control fragment chr12	AC141554-3F	TGGGCAGGCGCAATCTAAT	Chr12:99,637,540-99,637,561	64	443
	AC141554-2R	GGGTAACCATCCCCGTGATT	Chr12:99,637,964-99,637,983		
Control fragment chr17	AC005883-1F	TGTCACGTCTCTTTAGTCTCCTT	Chr17:51,565,437-51,565,459	64	771
	AC005883-2R	TAGCCCAGTTTCCGCTCATT	Chr17:51,566,189-51,566,208		
Junction fragment der(12)	AC141554-3F	TGGGCAGGCGCAATCTAAT	Chr12:99,637,540-99,637,561	63	744
	AC005883-2R	TAGCCCAGTTTCCGCTCATT	Chr17:51,566,189-51,566,208		
Junction fragment der(17)	AC005883-1F	TGTCACGTCTCTTTAGTCTCCTT	Chr17:51,565,437-51,565,459	63	461
	AC141554-2R	GGGTAACCATCCCCGTGATT	Chr12:99,637,964-99,637,983		
		DGRC0030			
t(1;3)(q42.11p25.3)					
Control fragment chr1	WDR26-1F	AGCACTGCTAATCCGTGACA	Chr1:224,397,833-224,397,852	62	418
	WDR26-2R	ATCTGTATTCTCCTTTGCCTGC	Chr1:224,398,229-224,398,250		
Control fragment chr3	ATPB2B2-1F	TCTAAAGGATGTTCGTGTGGAAG	Chr3:10,670,741-10,670,763	62	511
	ATPB2B2-2R	ACTCCACCCCTTTTACCACC	Chr3:10,671,232-10,671,251		
Junction fragment der(1)	WDR26-1F	AGCACTGCTAATCCGTGACA	Chr1:224,397,833-224,397,852	62	481
	ATPB2B2-2R	ACTCCACCCCTTTTACCACC	Chr3:10,671,232-10,671,251		

Junction fragment der(3)	ATPB2B2-1F	TCTAAAGGATGTTCGTGTGGAAG	Chr3:10,670,741-10,670,763	62	434
	WDR26-2R	ATCTGTATTCTCCTTTGCCTGC	Chr1:224,398,229-224,398,250		

[a]Reference genome assembly GRCh38/hg38.

Supplementary Table 2 Comparison of the proband's clinical features with those of patients with pathogenic SNVs, deletions and intragenic duplications of *ANKRD11* and 16q24.3 microdeletions encompassing *ANKRD11*

Clinical features	HPO	Proband's phenotype t(16;17)	Pathogenic variants /deletion/intragenic dup	16q24.3 microdeletions encompassing *ANKRD11* N=36	Prenatal case of KBG by deletion 16q24.2q24.3
Neurologic					
Global developmental delay	HP:0001263	Yes	16/40 (40%)	30/36 (~83%)	NA
Seizures	HP:0001250	No	8/40 (20%)	10/36 (~28%)	NA
Abnormality of brain morphology	HP:0012443	Unknown	1/40 (2.5%)	10/36 (~28%)	Not reported
Craniofacial					
Microcephaly	HP:0000252	No			
Low posterior hairline	HP:0002162	Yes			
Round face early in life	HP:0000311	Yes			
Long philtrum	HP:0000343	No			
Anteverted nares	HP:0000463	No	Characteristic facial anomalies	Characteristic facial anomalies	Triangular face Mildly low-seat ears
Underdeveloped nasal alae	HP:0000430	No			
Protruding ear	HP:0000411	Yes			
Hypertelorism	HP:0000316	Yes	40/40 (100%)	16/36 (~44%)	
Telecanthus	HP:0000506	No			
Long palpebral fissures	HP:0000637	No			
Thick eyebrows	HP:0000574	No			
Growth					
Postnatal growth retardation	HP:0008897	Yes	27/40 (675%)	12/36 (~33%)	Yes
Hand anomalies					
Single transverse palmar crease	HP:0000954	Yes			
Finger clinodactyly	HP:0040019	Yes	Hand anomalies 35/40 (87.5%)	Hand anomalies 10/36 (~28%)	Not reported
Finger syndactyly	HP:0006101	No			
Skeletal					
Delayed skeletal maturation	HP:0002750	Unknown	12/40 (30%)	10/36 (~28%)	
Thoracic kyphosis	HP:0002942	No			Cervical ribs Irregular C1-C4 vertebral bodies
Vertebral fusion	HP:0002948	Unknown	Costovertebral anomalies	Costovertebral anomalies	
Vertebral arch anomaly	HP:0008438	Unknown	17/40 (42.5%)	0/36 (0%)	
Accessory cervical ribs	HP:0000891	Unknown			
Eyes					
Astigmatism	HP:0000483	Unknown	0/40 (0%)	10/36 (~28%)	Not reported
Cardiovascular					
Congenital heart defect	HP:0001627	Yes	5/40 (12.5%)	12/36 (~33%)	Not reported
Hematopoietic system					
Thrombocytopenia	HP:0001873	Unknown	0/40 (0%)	8/36 (~22%)	Not reported

Pathogenic variants of *ANKRD11* and 16q24.3 microdeletions are according to Novara et al. (2017). Prospective antenatal diagnosis of KBG syndrome by a 1.86 Mb deletion, arr[GRCh37] 16q24.2q24.3(87614996_89479537)x1 dn, containing 26 genes including *ANKRD11* reported in Hodgetts et al. (2017). Pregnancy was terminated at 27 weeks; additional features at postmortem examination included lobation of the left lung, lobulated spleen and portal tracts calcification of the liver.

Supplementary Table 3 Comparison of the DGRC0025 – t(12;17) proband's clinical features with those of patients with *ANKS1B* haploinsufficiency syndrome

Clinical Features	HPO nomenclature	Proband's phenotype t(12;17)(q23;q22)	*ANKS1B* microdeletion n=9 (%)
Craniofacial			
Brachycephaly	HP:0000248	No	3/9 (33%)
Positional plagiocephaly	HP:0001357	No	1/9 (11%)
Round face	HP:0000311	No	4/9 (44%)
Prominent brow	HP:0000336	No	1/9 (11%)
Prominent metopic ridge	HP:0005487	No	1/9 (11%)
Congenital glabellar hemangioma	HP:0001076	Yes	-
Midface hypoplasia	HP:0011800	No	3/9 (33%)
Hypertelorism	HP:0000316	No	1/9 (11%)
Synophrys	HP:0000664	No	1/9 (11%)
Downslanted palpebral fissure	HP:0000494	No	1/9 (11%)
Upslanted palpebral fissure	HP:0000582	No	1/9 (11%)
Short nose	HP:0003196	No	1/9 (11%)
Depressed nasal bridge	HP:0005280	Yes	-
Bulbous nasal tip	HP:0000414	Yes	-
Anteverted nares	HP:0000463	Yes	1/9 (11%)
Thin upper lip vermilion	HP:0000219	Yes	-
Macroglossia	HP:0000158	No	1/9 (11%)
Preauricular pit	HP:0004467	No	1/9 (11%)
Dysplastic ears	HP:0000377	Yes	-
Microtia	HP:0008551	No	1/9 (11%)
Neurologic			
Intellectual disability	HP:0001249	No	3/9 (33%)
Developmental delay	HP:0001263	Yes	6/9 (67%)
Speech delay	HP:0000750	Yes	8/9(89%)
Speech apraxia	HP:0011098	Yes	3/9 (33%)
Motor delay	HP:0001270	No	6/9 (67%)
Motor dyspraxia	HP:0004302	Yes	3/9 (33%)
Autism Spectrum Disorder (ASD)	HP:0000729	No	5/9 (56%)
Attention Deficit Hyperactivity Disorder (ADHD)	HP:0007018	Yes	4/9 (44%)
Encopresis	HP:0040183	No	1/9 (11%)
Hypertonicity	HP:0001276	No	1/9 (11%)
Hypersensitive sensory processing deficit	HP:0003474	No	1/9 (11%)
Congenital torticollis	HP:0005988	No	1/9 (11%)
Myoclonic jerks	HP:0001336	No	1/9 (11%)
Sleep disturbance	HP:0002360	No	1/9 (11%)
Eyes			
Astigmatism	HP:0000483	No	2/9 (22%)
Otolaryngology			
Choanal atresia	HP:0000453	No	1/9 (11%)
High palate	HP:0000218	No	3/9 (33%)
Cardiovascular			
Atrial septal defect	HP:0001631	No	1/9 (11%)
Pulmonary artery stenosis	HP:0004415	Yes	1/9 (11%)
Patent Foramen Ovale (PFO)	HP:0001655	Yes	-
Respiratory			
Asthma	HP:0002099	No	1/9 (11%)
Hands/Toes Anomalies			
Brachydactyly	HP:0001156	No	1/9 (11%)
Short phalanx of finger	HP:0009803	No	1/9 (11%)
Short toe	HP:0001831	No	1/9 (11%)
Other			

Specialized schooling required		Yes	1/9 (11%)
Speech therapy		Yes	-
MRI			
Enlarged ventricles		Unknown	1/9 (11%)
Thin corpus callosum	HP:0002079	Unknown	3/9 (33%)
Dysgenesis of corpus callosum	HP:0006989	Unknown	1/9 (11%)
Hyperintensity in left periventricular white matter		Unknown	1/9 (11%)
Absent splenium		Unknown	1/9 (11%)

Clinical phenotypes of patients with *ANKS1B* haploinsufficiency syndrome are according to Carbonell et al. (2019).

Supplementary Table 4 Comparison of clinical features between the DGRC0030 – t(1;3) proband and those with pathogenic SNVs and 1q41q42 microdeletions either encompassing *WDR26* or not encompassing *WDR26*

DGRC0030 - t(1;3) proband's phenotype	HPO nomenclature	Pathogenic *WDR26* SNVs n=15 (%)	1q41q42 microdeletions		
			encompassing *WDR26*		proximal to *WDR26*
			<4Mb n=12 (%)	>4Mb n=13 (%)	n=8 (%)
Neurologic		**15 (100)**	**12 (100)**	**11 (85)**	**7 (88)**
Central Nervous System					
Severe global developmental delay with enuresis and encopresis	HP:0011344	15 (100)	11 (92)	9 (69)	6 (75)
Delayed speech and language development	HP:0000750	11 (73)	7 (58)	1 (8)	1 (13)
Absent speech	HP:0001344	2 (13)	1 (8)	1 (8)	0
Broad based ataxic gait	HP:0002136	9 (60)	1 (8)	1 (8)	1 (13)
Hypotonia	HP:0001290	9 (60)	4 (33)	3 (23)	3 (38)
Feeding problems	HP:0011968	4 (27)	2 (17)	2 (15)	1 (13)
Seizures[a]	(HP:0001250)	15 (100)	9 (75)	6 (46)	2 (25)
Abnormality of brain morphology[b]	(HP:0012443)	10 (67)	8 (67)	9 (69)	1 (13)
Behavioral psychiatric manifestations					
Behavioral problems	HP:0000708	4 (27)	2 (17)	0	2 (25)
Autistic features	HP:0000729	5 (33)	1 (8)	0	0
Happy and/or friendly personality	HP:0040082	10 (67)	3 (25)	0	0
Craniofacial		**15 (100)**	**12 (100)**	**11 (85)**	**4 (50)**
Mild microcephaly	HP:0040196	11 (73)	2 (17)	4 (31)	2 (25)
Full cheeks as a child	-	11 (73)	1 (8)	1 (8)	0
Coarse facial features	HP:0000280	12 (80)	6 (50)	4 (31)	0
Prominent maxilla	HP:0430028	13 (87)	2 (17)	0	0
Depressed nasal bridge	HP:0005280	5 (33)	5 (38)	5 (45)	2 (25)
Anteverted nares	HP:0000463	8 (53)	2 (17)	2 (15)	0
Full/broad nasal tip	HP:0000453	11 (73)	5 (42)	3 (23)	1 (13)
Wide mouth	HP:0000154	4 (17)	2 (17)	1 (8)	0
Decreased cupid´s bow	HP:0010800	11 (73)	3 (25)	1 (8)	0
Tented upper lip vermilion	HP:0010804	0	4 (33)	2 (15)	2 (25)
Protruding upper lip	HP:0000215	13 (87)	5 (42)	3 (23)	3 (38)
Widely spaced teeth	HP:0000687	13 (87)	5 (42)	2 (15)	0
Growth		**12 (80)**	**6 (50)**	**4 (31)**	**3 (38)**
Postnatal growth retardation	HP:0008897	11 (73)	6 (50)	4 (31)	3 (38)
Eyes		**1 (7)**	**2 (17)**	**4 (31)**	**2 (25)**
Strabismus	HP:0000486	1 (7)	1 (8)	2 (15)	2 (25)
Cardiovascular (*Congenital heart disease*)	**HP:0030680**	**2 (13)**	**3 (25)**	**2 (15)**	**0 (0)**
Ventricular septal defect – Subaortic interventricular communication	HP:0001629 - HP:0011681	1 (7)	0	2 (15)	0
Tetralogy of Fallot	HP:0001636	0	0	0	0
Right Aortic Arch	HP:0012020	1 (7)	0	0	0

Genitourinary (*Male genital anomalies*)	**HP:0000032**	**1 (7)**	**3 (25)**	**3 (23)**	**1 (13)**
Phimosis	HP:0001741	0	0	3 (23)	0
Hydrocele testis	HP:0000034	0	0	0	0
Midline defect		**1 (7)**	**6 (50)**	**7 (54)**	**0 (0)**
Umbilical hernia	HP:0001537	0	0	0	0
Chest		**0 (0)**	**1 (8)**	**0 (0)**	**0 (0)**
Accessory or supernumerary nipple	HP:0002558	0	1 (8)	0	0
Skeletal		**4 (27)**	**3 (25)**	**4 (31)**	**0 (0)**
Thoracic scoliosis	HP:0002943	2 (13)	1 (8)	0	0
Clubfoot or bilateral *talipes equinovarus*	HP:0001776	0	2 (17)	4 (31)	0
Posture		**1 (7)**	**0 (0)**	**1 (8)**	**0 (0)**
Neck hyperextension		1 (7)	0	1 (8)	0

Shaded in light gray are clinical features of the DGRC0025 – t(1;3) proband predominantly shared with subjects with pathogenic SNVs and 1q41q42 microdeletions either encompassing *WDR26* or not, whereas those in dark gray are predominantly shared by subjects with affected *WDR26*. Pathogenic *WDR26* SNVs as well as 1q41q42 microdeletions are mainly from Skraban et al. (2017).
[a]Have not been reported yet; [b]Have not been identified by magnetic resonance imaging.

Supplementary Table 5 liGS libraries and mapping metrics of the analyzed cases

Metrics	DGRC0016 t(16;17)		DGRC0019 t(2;19)		DGRC0006 t(8;14)		DGRC0013 inv(13)		DGRC0025 t(12;17)		DGRC0030 t(1;3)	
	#	%	#	%	#	%	#	%	#	%	#	%
liGS libraries												
Read size (bp)	26	—	25	—	25	—	25	—	25	—	25	—
Reads per library	146.11 M	—	196.26 M	—	165.53 M	—	129.11M	—	212.96 M	—	208.62 M	—
Read-pairs per library	73.07 M	—	98.13 M	—	82.77 M	—	64.56 M	—	106.48 M	—	104.31 M	—
Median insert size (bp)	2,877	—	3,110	—	3,326	—	3,429	—	3,188	—	3,136	—
MAD[a] insert size (bp)	740	—	777	—	670	—	691	—	656	—	615	—
Physical coverage	52	—	42	—	72	—	58	—	88	—	72	—
Sequence coverage	1.02	—	1.26	—	1.09	—	0.84	—	1.5	—	1.16	—
Mapping												
Total mapped reads	142.22 M	97.34	182.89 M	93.70	162.55 M	98.2	127.06 M	98.41	207.66 M	97.51	202.08 M	96.86
Singletons[b]	2.92 M	2.17	2.45 M	2.19	2.48 M	1.5	1.88 M	1.46	4.49 M	2.11	5.51 M	2.64
Total mapped read-pairs	69.42 M	95.02	89.53 M	91.24	80.04 M	96.7	62.59 M	96.95	101.58 M	95.40	98.29 M	94.22
Duplicated mapped read-pairs	5.63 M	8.24	42.09 M	47.10	4.37M	5.27	3.23 M	5.00	6.44M	6.05	7.07M	6.78
Distinctly mapped read-pairs	63.79 M	94.61	47.45 M	84.66	75.93 M	91.74	59.54 M	92.24	95.65 M	89.83	80.16 M	76.84
Distinct proper pairs	61.35 M	90.99	43.05 M	76.82	71.11 M	85.91	56.63 M	87.72	88.14 M	82.78	73.23 M	70.21
Distinct chimeric and improper pairs	2.44 M	3.62	4.39 M	7.84	4.83 M	5.83	2.92M	4.52	7.51 M	7.05	6.92 M	6.64

[a]Median Absolute Deviation; [b]Reads mapped without a pair

Supplementary Table 6 Summary of interchr chimeric and intrachr inv type structural variants identified at different resolution

Analysis	DGRC0016 t(16;17)		DGRC0019 t(2;19)		DGRC0006 t(8;14)		DGRC0013 inv(13)		DGRC0025 t(12;17)		DGRC0030 t(1;3)	
	Chimeric	Improper	Chimeric	Improper	Chimeric	Improper	Chimeric	Improper	Chimeric	Improper	Chimeric	Improper
Identification method												
Number of read-pair clusters[a]	375	135	481	111	550	186	378	130	515	182	477	176
Blacklist filtering[b]	269	108	312	84	309	149	214	112	292	148	290	139
Sub-clustering[c] (Cplt Clu; Incplt Clu)	62; 207	38; 70	78; 234	26; 58	124; 185	76; 73	76; 138	59; 53	138; 154	95; 33	109; 181	76; 63
< 3 read-pairs filter[d] (Cplt Clu; Incplt Clu)	22; 93	—	16; 74	—	26; 59	—	18; 54	—	25; 44	—	21; 57	—
Structural variants – fully;partly resolved[e]												
Translocation	1[f]; 0	—	1[f]; 0	—	1[f]; 0	—	0; 0	—	1[f]; 0	—	1[f]; 0	—
Interchromosomal insertion	11; 8	—	6; 8	—	18; 8	—	10; 9	—	13; 10	—	8; 10	—
Inversion	—	29; 17	—	20; 19	—	49; 9	—	42[f]; 10	—	61; 7	—	49; 7
Intrachromosomal insertion	—	6; 11	—	3; 10	—	15; 11	—	12; 11	—	25; 3	—	21; 6
Complex SV	—	5; 4	—	5; 0	—	8; 2	—	8; 1	—	5; 2	—	8; 1
Total	**52**; **40**		**35**; **37**		**91**; **30**		**72**; **31**		**105**; **22**		**87**; **24**	
At clinical resolution of ≥ 10 kb – fully;partly resolved[g]												
Interchromosomal insertion	1; 0	—	1; 0	—	3; 0	—	1(1); 0	—	0; 0	—	0; 0	—
Inversion	—	2; 3	—	1; 3	—	4(1); 2	—	7[f](1); 2	—	7; 0	—	3(1); 1
Intrachromosomal insertion	—	0; 1	—	0; 2	—	1; 0	—	1; 0	—	3; 0	—	0; 0
Complex SV	—	4(2); 2	—	4(4); 0	—	3(3); 2(1)	—	3(2); 1(1)	—	3(1); 2(1)	—	3(2); 1(1)
Total	**7(2)**; **6**		**6(4)**; **5**		**11(4)**; **4(1)**		**12(4)**; **3(1)**		**13(1)**; **2(1)**		**6(3)**; **2(1)**	
At liGS resolution of < 10 kb – fully;partly resolved[g]												
Interchromosomal insertion	10; 8	—	5; 8	—	15(2); 8(1)	—	9; 9(1)	—	13(2); 10	—	8; 10	—
Inversion	—	27; 14	—	19; 16	—	45(2); 7	—	35; 8	—	55(10); 7	—	46(3); 6
Intrachromosomal insertion	—	6; 10	—	3; 8	—	14; 11	—	11(1); 11	—	22(4); 3	—	21(3); 6
Complex SV	—	1; 2	—	1(1); 0	—	5(1); 0	—	5(2); 0	—	2; 0	—	5(1); 0
Total	**44**; **34**		**28(1)**; **32**		**79(5)**; **26(1)**		**60(3)**; **28(1)**		**92(16)**; **20**		**80(7)**; **22**	

Read-pairs are considered interchr or chimeric if both reads of the pair map to different chromosomes, allowing identification of translocations and interchr ins. On the other hand, those considered intrachr inv type improper read-pairs map on the same chromosome and in the same orientation, allowing identification of inv, intrachr ins and cx variants. Clusters of different categories (complete and incomplete) as well as fully and partly resolved alterations are separated by semicolon (;). [a]Read-pair clusters denoting interchr chimeric and intrachr inv type improper read-pair clusters were identified by an in-house developed Python script improperCLAS.py and clustered by readPairCluster (Talkowski et al. 2011). [b]Genomic regions delimited by different type of clusters were filtered against the regions of the "blacklist", with an overlap cutoff of ≥ 30%. [c]If possible, according to orientation of each read of each cluster specific read-pair, a cluster is divided into two breakpoint specific sub-clusters

constituting a complete cluster (Cplt Clu) or otherwise they are considered as incomplete clusters (Incplt Clu). [d]Sub-clusters with less than 3 read-pairs are filtered out. [e]Number of fully and partly resolved structural variants are indicated. [f]Cytogenetically reported rearrangement. The mean number of read-pairs that define breakpoints of cytogenetically reported rearrangement was 24 (ranging 14 to 43) whereas breakpoint resolution was about 200 bp (ranging from zero to 919). [g]Proband-specific alterations and those with a frequency lower than 1% in the SVref dataset are presented in parentheses. The SVref dataset is according to Collins et al. (2017). Patient specific alterations are also presented in Supplementary Table 15.

Supplementary Table 7 del and dup identified by different methods, cross validations, at different resolutions and frequencies

Analysis	DGRC0016 t(16;17)		DGRC0019 t(2;19)		DGRC0006 t(8;14)		DGRC0013 inv(13)		DGRC0025 t(12;17)		DGRC0030 t(1;3)	
	del	dup	del	dup	del	dup	del	dup	del	dup	del	dup
Identification method												
Improper read-pair clusters[a]	39,781	43	23,989	23,714	17,838	74	1,403	51	60,855	67	69,394	70
Overlap with blacklist regions[b]	517	40	78	18	371	68	293	49	538	63	573	66
Altered DoC regions[c]	556	1,449	707	3,007	488	383	326	608	253	425	294	1,046
Overlap with blacklist regions[b]	333	367	341	979	305	215	244	391	202	311	227	690
Cross-validation												
Cluster + DoC[d]	46	3	19	2	80	8	75	2	62	5	29	7
Cluster + DoC - SVref dataset[e]	7	0	4	0	15	6	8	1	6	2	3	2
Cluster + DoC + SVref dataset[f]	39	3	15	2	65	2	67	1	56	3	26	5
Cluster + SVref dataset[g]	91	3	18	0	75	6	109	4	56	8	36	6
DoC+ SVref dataset[h]	29	24	44	41	32	42	22	31	44	34	55	66
Total	**166**	**30**	**81**	**43**	**187**	**56**	**206**	**37**	**162**	**47**	**120**	**79**
At clinical resolution of ≥ 30 kb												
Cluster + DoC[d]	7	1	2	1	8	1	7	0	7	2	2	5
Cluster + DoC - SVref dataset[e]	2	0	0	0	4	1	1	0	1	2	0	1
Cluster + DoC + SVref dataset[f]	5	1	2 (1)	1	4	0	6 (1)	0	6 (1)	0	2	4 (1)
Cluster + SVref dataset[g]	9	1 (1)	5 (1)	0	4	2	6	3	9 (2)	4	3	3
DoC+ SVref dataset[h]	8 (4)	9 (5)	7 (1)	15 (7)	6 (2)	14 (4)	9 (1)	12 (1)	3 (1)	11 (5)	4	22 (10)
Total	**24 (4)**	**11 (6)**	**14 (3)**	**16 (7)**	**18 (2)**	**17 (4)**	**22 (2)**	**15 (1)**	**19 (4)**	**17 (5)**	**9**	**30 (11)**
At liGS resolution < 30 kb												
Cluster + DoC[d]	39	2	17	1	72	7	68	2	55	3	27	2
Cluster + DoC - SVref dataset[e]	5	0	4	0	11	5	7	1	5	0	3	1
Cluster + DoC + SVref dataset[f]	34 (3)	2	13 (2)	0	61 (14)	2	61 (1)	1	50	3	24 (2)	1
Cluster + SVref dataset[g]	82 (2)	2	13	0	71 (3)	4 (2)	103 (1)	1	47 (2)	4	33	3 (1)
DoC+ SVref dataset[h]	21 (3)	15 (11)	37 (4)	26(15)	26 (5)	28 (15)	13 (1)	19 (10)	41 (1)	23 (8)	51 (2)	44 (27)
Total	**142 (8)**	**19 (11)**	**67 (6)**	**27 (15)**	**169 (22)**	**39 (17)**	**184 (3)**	**22 (10)**	**143 (3)**	**30 (8)**	**111 (4)**	**49 (28)**

Total number of CNVs identified by each analysis are shown. Identified CNVs validated in the SVref dataset with a frequency <1% are in parentheses. The underlined numbers indicate CNVs analyzed in more detail using the CNV-ConTool.

[a]Improper read-pairs denoting del and tandem dup were identified by an in-house developed Python script designated improperCLAS.py and clustered together by readPairCluster (Talkowski et al. 2011). Inward and outward facing read-pairs with insert size (IS) larger than the median IS+3* IS SD denote improper del and dup read-pairs, respectively. [b]Genomic regions delimited by read-pair clustering and DoC analysis were filtered against ≥30% overlap with the genomic regions defined in the "blacklist". [c]Identification of del and dup through DoC analysis, performed by CN.MOPS. [d]Total number of del and dup identified by read-pair clustering and DoC analysis. [e]del and dup identified by read-pair clustering

and DoC analysis, not reported in the SVref dataset. [f]del and dup identified by read-pair clustering and DoC analysis, that are reported in the SVref dataset. [g]del and dup identified only by read-pairs clustering, also present in SVref dataset. [h]del and dup identified only by DoC analysis, also present in SVref dataset.

Supplementary Table 8 Protein coding and non-coding RNA genes localized within hESC TADs disrupted by DGRC0016 t(16;17) breakpoints and associated phenotypes

Genes				OMIM and DD2GP based pathologies					
GeneCard	**OMIM**	**oe (90% CI)**	**%HI**	**Clinical phenotype**	**OMIM**	**Inheritance**	**DD2GP category**	**PMID reference**	
16q24.3 breakpoint within hESC TAD at 16:89246091-89686091									
ANKRD11 - IVS2	611192	0.05 (0.02-0.11)	75.55	KBG syndrome	148050	AD	confirmed	15378538	21782149
der(16) g.[chr16:pter_cen_89401716_89401717del::chr17:46784035_qter]									
SPG7	602783	1.26	61.82	Spastic paraplegia 7, AR	607259	AD; AR[a]	nd	na	na
RPL13	113703	0.00 (0.00-0.32)	58.62	—	na	nd	nd	na	na
CPNE7	605689	1.05	75.88	—	na	nd	nd	na	na
DPEP1	179780	0.85	77.20	—	na	nd	nd	na	na
CHMP1A	164010	0.61	33.54	Pontocerebellar hypoplasia, type 8	614961	AR	probable	23023333	na
SPATA33	615409	1.51	98.71	—	na	nd	nd	na	na
				Al Kaissi syndrome	617694	AR	nd	na	na
CDK10	603464	1.01	51.33	Severe Growth Retardation Spine Malformations and Developmental Delays			possible	28886341	29130579
17q21.31 breakpoint within hESC TAD at 17:46687454-47647635									
NSF	601633	0.33 (0.18-0.61)	4.09	—	na	nd	nd	na	na
WNT3 - IVS3	165330	0.13 (0.05-0.39)	13.89	Tetra-amelia syndrome 1	273395	AR	confirmed	na	na
der(17) g.[chr17:pter_cen_46781999_46784034del::chr16:89401718_qter]									
WNT9B	602864	0.20 (0.08-0.62)	34.84	—	na	nd	nd	na	na
GOSR2	604027	0.82	32.66	Epilepsy, progressive myoclonic 6	614018	AR	nd	na	na
CDC27	116946	0.12 (0.06-0.25)	9.78	—	na	nd	nd	na	na
MYL4	160770	0.74	34.00	Atrial fibrillation, familial, 18	617280	AD	nd	na	na
				Bleeding disorder, platelet-type, 16, AD	187800	AD	nd	na	na
				Glanzmann thrombasthenia	273800	AR	nd	na	na
ITGB3	173470	0.32 (0.20-0.51)	35.78	Purpura, posttransfusion	na	nd	nd	na	na
				Thrombocytopenia, neonatal alloimmune	na	nd	nd	na	na
				Myocardial infarction, susceptibility to	608446	nd	nd	na	na
NPEPPS	606793	0.00 (0.00-0.06)	11.57	—	na	nd	nd	na	na

Analysis performed with bioinformatics tool TAD-GConTool. Reference genome assembly is GRCh38/hg38, and human embryonic stem cells (hESC) TADs are according to Dixon et al. (2012). Genes showing high probability of being disease causing are in bold (McKusick 1998; Wright et al. 2015). LoF intolerance, is expressed as observed / expected number (oe) of LoF variants. For genes with oe <0.35 the oe at 90% CI is also stated. HI, Haploinsufficiency index; na, not assigned; nd, not determined; AR, Autosomal Recessive, AD, Autosomal Dominant. [a]A dominant effect for some *SPG7* mutations has been reported (Sánchez-Ferrero et al. 2013).

Supplementary Table 9 Protein coding and non-coding RNA genes localized within hESC TADs disrupted by DGRC0019 t(2;19) breakpoints and associated phenotypes

Genes				OMIM and DD2GP based pathologies					
GeneCard	OMIM	oe (90% CI)	%HI	Clinical phenotype	OMIM	Inheritance	DD2GP category	PMID reference	
2p13.3 breakpoint within hESC TAD at 2:70279364-70999362									
PCYOX1	610995	1.00	55.49	—	na	nd	nd	na	na
SNRPG	603542	0.00 (0.00-0.55)	7.48	—	na	nd	nd	na	na
FAM136A	616275	0.93	74.63	—	na	nd	nd	na	na
TGFA	190170	0.11 (0.04-0.55)	2.83	—	na	nd	nd	na	na
ADD2	102681	0.13 (0.07-0.27)	37.61	—	na	nd	nd	na	na
FIGLA	608697	0.65	65.57	Premature ovarian failure 6	612310	AD	nd	na	na
CLEC4F	na	0.91	90.63	—	na	nd	nd	na	na
CD207	604862	0.84	86.68	Birbeck granule deficiency	613393	nd	nd	na	na
LINC01143	na	nd	nd	—	na	nd	nd	na	na
VAX2	604295	0.33 (0.14-1.04)	47.75	—	na	nd	nd	na	na
ATP6V1B1 - IVS1	192132	0.63 (0.43-0.93)	46.96	Renal tubular acidosis with deafness	267300	AR	confirmed	12566520	9916796
AC007040 - Exon 5	na	nd	nd	—	na	nd	nd	na	na
der(2) g.[chr19:32878515_qterinv::chr2:70941503_70941506del_cen_qter]									
ANKRD53	617009	0.83	89.63	—	na	nd	nd	na	na
TEX261	na	0.59	28.75	—	na	nd	nd	na	na
19q13.11 breakpoint within hESC TAD at 19:32577254-33217254									
PDCD5	604583	0.56	14.97	—	na	nd	nd	na	na
ANKRD27	na	0.83	48.32	—	na	nd	nd	na	na
RGS9BP	607814	0.86	74.66	Bradyopsia	608415	nd	nd	na	na
AC008736	na	nd	nd	—	na	nd	nd	na	na
NUDT19	na	1.96	87.63	—	na	nd	nd	na	na
TDRD12	na	0.98	59.89	—	na	nd	nd	na	na
SLC7A9	604144	0.75	61.03	Cystinuria	220100	AD;AR	nd	na	na
CEP89 - Exon 18	615470	0.93 (0.72-1.21)	82.86	—	na	nd	nd	na	na
der(19) g.[chr19:pter_cen_32878513_32878514del::CATA::chr2:pter_70941502inv]									
FAAP24	610884	nd	nd	—	na	nd	nd	na	na
RHPN2	617932	0.49	64.24	—	na	nd	nd	na	na
GPATCH1	na	0.48	58.94	—	na	nd	nd	na	na

WDR88	na	0.73	83.65	—	na	nd	nd	na	na
LRP3	603159	0.47	60.40	—	na	nd	nd	na	na
SLC7A10	607959	0.47	40.99	—	na	nd	nd	na	na

Analysis performed with bioinformatics tool TAD-GConTool. Reference genome assembly is GRCh38/hg38, and human embryonic stem cells (hESC) TADs are according to Dixon et al. (2012). oe, LoF intolerance, expressed as observed / expected number of LoF variants. For genes with oe <0.35 the oe at 90% CI is also stated. HI, Haploinsufficiency index; na, not assigned; nd, not determined; AR, Autosomal Recessive; AD, Autosomal Dominant.

Supplementary Table 10 Protein coding and non-coding RNA genes localized within hESC TADs disrupted by DGRC0019 t(2;19) breakpoints with their associated phenotypes and GWAS data from the referred regions

Genes and intergenic regions			Clinical Phenotype	GWAS SNP data
GeneCard	OMIM	oe (CI 95%)	Phenotype;#OMIM_Inheritance; DDG2P_category	Genetic traits
2p13.3 breakpoint within hESC TAD at 2:70279364-70999362				
PCYOX1	610995	1.00	None	None
SNRPG	603542	0.00 (0.00-0.55)	None	None
FAM136A	616275	0.93	None	None
TGFA	190170	0.11 (0.04-0.55)	None	None
ADD2	102681	0.13 (0.07-0.27)	None	None
FIGLA	608697	0.65	Premature ovarian failure 6; #612310_**AD**;DDG2P_NA	None
CLEC4F	na	0.91	None	None
CD207	604862	0.84	Birbeck granule deficiency; #613393_nd;DDG2P_NA	# 1 SNPs - HP:0001047_Atopic dermatitis [1.11E-8)
Intergenic - chr2:70835822-70887871			None	# 19 SNPs - HP:0001047_Atopic dermatitis [1.11E-8]
LINC01143	na	nd	None	# 1 SNPs - HP:0001047_Atopic dermatitis [1.11E-8];
VAX2	604295	0.33 (0.14-1.04)	None	None
ATP6V1B1 - IVS1	192132	0.63 (0.43-0.93)	Renal tubular acidosis with deafness; #267300_AR;DDG2P_confirmed	# 1 SNPs - EFO:0004612_High density lipoprotein cholesterol measurement [3.61E-7]; DOID:1461_Cholesterol embolism [3.61E-7]; MESH:D008076_Cholesterol hdl [3.61E-7]
AC007040 - Exon 5	na	nd	None	None
der(2) g.[chr19:32878515_qterinv::chr2:70941503_70941506del_cen_qter]				
Intergenic - chr2:70941504-70978380			None	# 1 SNPs - MESH:D008076_Cholesterol hdl [3.61E-7]; DOID:1461_Cholesterol embolism [3.61E-7]; EFO:0004612_High density lipoprotein cholesterol measurement [3.61E-7]
ANKRD53	617009	0.83	None	None
TEX261	na	0.59	None	None
19q13.11 breakpoint within hESC TAD at 19:32577254-33217254				
PDCD5	604583	0.56	None	None
ANKRD27	na	0.83	None	None
RGS9BP	607814	0.86	Bradyopsia; #608415_nd;DDG2P_NA	# 1 SNPs - EFO:0004518_Serum creatinine measurement [5E-11]; DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11]; HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10]
Intergenic – chr19:32678300-32687089			None	# 5 SNPs - HP:0000077_Abnormality of the kidney [5E-11]; DOID:784_Chronic kidney failure [5E-11]; EFO:0004518_Serum creatinine measurement [5E-11] # 2 SNPs - HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10]

AC008736	na	nd	None	# 3 SNPs - HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10] # 1 SNPs - EFO:0004518_Serum creatinine measurement [5E-11]; DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11]
Intergenic - chr19:32691750-32691961			None	# 1 SNPs - HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10]
NUDT19	na	1.96	None	# 10 SNPs - HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10] # 3 SNPs - EFO:0004518_Serum creatinine measurement [5E-11]; DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11]
Intergenic - chr19:32713796-32719753			None	# 4 SNPs - HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10] # 1 SNPs - HP:0000077_Abnormality of the kidney [5E-11]; DOID:784_Chronic kidney failure [5E-11]; EFO:0004518_Serum creatinine measurement [5E-11]
TDRD12	na	0.98	None	# 23 SNPs - EFO:0004518_Serum creatinine measurement [5E-11]; DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11] # 15 SNPs - HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10]
Intergenic - chr19:32829580-32830509			None	# 1 SNPs - DOID:614_Lymphopenia [5.85E-8]; EFO:0004518_Serum creatinine measurement [5E-11]; DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11]
SLC7A9	604144	0.75	Cystinuria; #220100_ **AD**,AR;DDG2P_NA	# 32 SNPs - DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11] # 7 SNPs - DOID:655_Inherited metabolic disorder [1E-16]; HP:0001939_Abnormality of metabolism homeostasis [1E-16] # 4 SNPs - HP:0001877_Abnormality of erythrocytes [1E-10]; MESH:D012805_Sickle cell trait [1E-10]
Intergenic - chr19:32869766-32875925			None	# 7 SNPs - DOID:655_Inherited metabolic disorder [1E-16]; HP:0001939_Abnormality of metabolism homeostasis [1E-16] # 5 SNPs - DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11] # 3 SNPs - EFO:0004518_Serum creatinine measurement [5E-11]
CEP89 - Exon 18	615470	0.93 (0.72-1.21)	None	# 109 SNPs - DOID:2841_Asthma [5.85E-8]; DOID:1240_Leukemia [5.85E-8]; DOID:614_Lymphopenia [5.85E-8]; HP:0002665_Lymphoma [5.85E-8] # 66 SNPs - DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11] # 62 SNPs - EFO:0004518_Serum creatinine measurement [5E-11] # 7 SNPs - DOID:655_Inherited metabolic disorder [1E-16]; HP:0001939_Abnormality of metabolism homeostasis [1E-16] # 1 SNPs - HP:0001370_Rheumatoid arthritis [1.76E-7]
der(19) g.[chr19:pter_cen_32878513_32878514del::CATA::chr2:pter_70941502inv]				
Intergenic - chr19:32878514-32972209			None	# 106 SNPs - DOID:614_Lymphopenia [5.85E-8]; DOID:2841_Asthma [5.85E-8]; HP:0002665_Lymphoma [5.85E-8]; DOID:1240_Leukemia [5.85E-8] # 63 SNPs - DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11] # 60 SNPs - EFO:0004518_Serum creatinine measurement [5E-11] # 6 SNPs - DOID:655_Inherited metabolic disorder [1E-16]; HP:0001939_Abnormality of metabolism homeostasis [1E-16] # 1 SNPs - HP:0001370_Rheumatoid arthritis [1.76E-7]
FAAP24	610884	nd	None	# 3 SNPs - EFO:0004518_Serum creatinine measurement [5E-11]; DOID:784_Chronic kidney failure [5E-11]; HP:0000077_Abnormality of the kidney [5E-11]

RHPN2	617932	0.49	None	# 24 SNPs - HP:0004348_Abnormality of bone mineral density [7E-11]; DOID:0080011_Bone resorption disease [7E-11] # 12 SNPs - MESH:D015519_Bone density [7E-11] # 7 SNPs - DOID:11476_Osteoporosis [4E-12]
Intergenic - chr19:33064888-33080880			None	# 14 SNPs - DOID:0080011_Bone resorption disease [7E-11]; HP:0004348_Abnormality of bone mineral density [7E-11] # 8 SNPs - MESH:D015519_Bone density [7E-11] # 6 SNPs - DOID:11476_Osteoporosis [4E-12] # 1 SNPs - DOID:9261_Nasopharynx carcinoma [5E-9]
GPATCH1	na	0.48	None	# 56 SNPs - HP:0004348_Abnormality of bone mineral density [7E-11]; DOID:0080011_Bone resorption disease [7E-11] # 39 SNPs - DOID:11476_Osteoporosis [4E-12] # 37 SNPs - MESH:D015519_Bone density [7E-11]
WDR88	na	0.73	None	None
LRP3	603159	0.47	None	None
SLC7A10	607959	0.47	None	None

Table prepared with TAD-GConTool that includes GWAS SNP data from Functional Annotation of the Mammalian Genome CAGE Associated Transcriptome (FANTOM CAT) database ((Hon et al. 2017). The SNP associated genetic traits are presented according to their ontology term ID. Although the conventional GWAS significance threshold is ≤ 5.00E-8, for disrupted genes, adjacent genomic regions and for genetic traits identified in GWAS overlapping the proband's clinical phenotype, a threshold of 9.99E-6 is used. As an example, genetic traits clearly associated with this locus are highlighted in different colors. Clinical phenotype data are from OMIM and Developmental Disorders Genotype-Phenotype (DDG2P) databases. The clinical phenotype data are configured as follows: i) gene-associated phenotypes in humans from either database; ii) #OMIM phenotype number underscore type of inheritance (AR, AD or AR/AD) or #NA in the absence of OMIM phenotype number; and DDG2P underscore phenotype classification (confirmed, probable and possible) or DDG2P_NA in the absence of DDG2P data.

Supplementary Table 11 del and dup identified in DGRC0016 t(16;17) by read-pair clustering and DoC analysis and cross validated using these data and an SVref dataset

Cluster/Coverage/ Array/Collins2017_ID[a]	Genomic variants [GRCh38][b]	Size (kb)	Ref.. group[c]	OMIM gene or phenotype[d]	Genomic variants <1% Frequency[e,f,g]	Genomic variants PMID reference	ACMG classif.[h]
	At clinical resolution of ≥30 kb (<1% frequency)						
Coverage Collins2017_DUP_3	DoC[GRCh38]1p36.33(50,000-85,000)x3 Ref[GRCh38]1p36.33(45,558-85,889)x3	35.00 40.33	—	—	nsv544867 (0.01)[f]	21841781	VUS
Cluster (29) Coverage Sanger validated	Clu[GRCh38] 3p24.1(27,354,585-27,408,246) DoC[GRCh38] 3p24.1(27,360,001-27,410,000)x1 SanS[GRCh38] chr3:g.27,354,680_27,408,191del	53.66 50.00 53.51	—	*SLC4A7*	Familial	—	VUS
Cluster (21) Coverage/Array Sanger validated	Clu[GRCh38] 8q24.21(129,061,012-129,897,302) DoC[GRCh38] 8q24.21(129,060,001-129,900,000)x1 SanS[GRCh38] chr8:g.129061233_129897281del	836.29 840.00 836.05	—	*CCDC26*, *GSDMC*, *FAM49B*	*de novo*	—	VUS
Coverage Collins2017_DEL_3501	DoC[GRCh38]9q13(62,400,000-62,580,000)x1 Ref[GRCh38]9q13(62,368,699-62,550,699)x1	180.00 182.00	—	—	(0.87)[g]	28260531	VUS
Coverage Collins2017_DEL_3519	DoC[GRCh38] 9q13q21.1(64,940,000-65,010,000)x1 Ref[GRCh38] 9q13q21.1(64,937,582-64,998,124)x1	70.00 60.54	—	—	(0.29)[g]	28260531	VUS
Coverage Collins2017_DEL_4027	DoC[GRCh38]11p11.12(48,810,000-48,840,000)x1 Ref[GRCh38]11p11.12(48,806,448-48,845,448)x1	30.00 39.00	—	—	esv2677500 (0.09)[e]	23128226	VUS
Coverage Collins2017_DEL_6280	DoC[GRCh38]Xq21.31(90,195,000-90,240,000)x1 Ref[GRCh38]Xq21.31(90,192,814-90,241,972)x1	45.00 49.16	—	—	esv2676290 (0.17)[e]	23128226	VUS

Analysis performed with CNV-ConTool with minimum reciprocal overlap of 70%. [a]Structural variants ID form the SVref dataset according to Collins et al. (2017). Values in parentheses indicate number of read-pairs per sequence cluster. [b]Clu, sequence cluster; DoC, depth-of-coverage; SanS, Sanger sequencing. [c]Number of subjects with similar genomic alteration in the internal reference group (N=32). [d]Genes with OMIM number, with or without an associated phenotype. [e]Frequency of variant from 1000 Genomes Project (Auton et al. 2015); [f]Frequency of variant from DGV public database (MacDonald et al. 2014) and [g]Frequency of variant from the SVref dataset (Collins et al. 2017). [h]CNV classification according to American College of Medical Genetics and Genomics and the Clinical Genome Resource (Riggs et al. 2019). VUS – Variant of unknown significance.

Supplementary Table 12 Protein coding and non-coding RNA genes localized within the 836.05 kb del at 8q24.21, and GWAS data from the referred region

Genes and intergenic regions			Clinical phenotype	GWAS SNP data
GeneCard	OMIM	oe (CI 95%)	OMIM# Phenotype_ Inheritance; (DDG2P)	Genetic traits
8q24.21 breakpoint within hESC TAD at 8:126838573-129758572				
chr8: g.129061233_129897281del				
LINC00977	na	nd	None	# 7 SNPs - HP:0000202_Oral cleft [7.03E-10]; EFO:0003959_Progranulin measurement [7.03E-10]; MESH:D002971_Cleft lip [7.03E-10]; DOID:0050567_Orofacial cleft [7.03E-10]
CCDC26	na	nd	None	# 43 SNPs - HP:0003002_Breast carcinoma [4.79E-10]; HP:0100615_Ovarian neoplasm [4.79E-10] # 42 SNPs - EFO:0005090_Monocyte count [3.00E-20] # 29 SNPs - MESH:D007962_Leukocytes [2.00E-10] # 22 SNPs - HP:0004386_Gastrointestinal inflammation [2.00E-9]; DOID:0050589_Inflammatory bowel disease [2.00E-9]
Intergenic - chr8:129061233-129685401			None	# 130 SNPs - HP:0000202_Oral cleft [7.03E-10]; EFO:0003959_Progranulin measurement [7.03E-10]; MESH:D002971_Cleft lip [7.03E-10]; DOID:0050567_Orofacial cleft [7.03E-10] # 109 SNPs - MESH:D005910_Glioma [5E-21]; DOID:0060108_Brain glioma [5E-21]; EFO:0000326_Central nervous system cancer [5E-21] # 42 SNPs - EFO:0005090_Monocyte count [3E-20] # 34 SNPs - HP:0003002_Breast carcinoma [4.79E-10]; HP:0100615_Ovarian neoplasm [4.79E-10] # 29 SNPs - MESH:D007962_Leukocytes [2E-10] # 25 SNPs - DOID:9261_Nasopharynx carcinoma [2E-18] # 22 SNPs - DOID:0050589_Inflammatory bowel disease [2E-9]; HP:0004386_Gastrointestinal inflammation [2E-9]
GSDMC	608384	0.86	None	# 1 SNPs - MESH:D005910_Glioma [5.00E-21]; EFO:0000326_Central nervous system cancer [5.00E-21]; DOID:0060108_Brain glioma [5.00E-21]; DOID:9261_Nasopharynx carcinoma [2.00E-18];
8q24.21 breakpoint within hESC TAD at 8:129798572-130718572				
Intergenic - chr8:129798572-129839593			None	None
FAM49B - IVS1	na	0.10 (0.04-0.31)	None	None

Table performed by TAD-GConTool, that includes GWAS SNP data from FANTOM CAT database (Hon et al. 2017), with GWAS significance threshold at ≤ 5.00E−8. The GWAS associated traits are presented according to their ontology term ID. Clinical phenotypes associated with genes were obtain from both OMIM and DDG2P databases.

Supplementary Table 13 del and dup identified in DGRC0019 t(2;19) by read-pair clustering and DoC analysis and cross validated using these data and an SVref dataset

Cluster/Coverage/ Array/Collins2017_ID[a]	Genomic variants [GRCh38][b]	Size (kb)	Ref. group[c]	OMIM gene or phenotype[d]	Genomic variants		ACMG classif.[g]
					<1% Frequency[e,f]	PMID reference	
	At clinical resolution of ≥30 kb (<1% frequency)						
Coverage Collins2017_DUP_1748	DoC[GRCh38]9q34.3(138,170,000-138,210,000)x3 Ref[GRCh38]9q34.3(138,174,809-138,219,550)x3	40.00 44.74	1	—	nsv1161919 (0.27)[e]	26073780	VUS
Coverage Collins2017_DUP_1803	DoC[GRCh38]10q11.22(47,560,000-47,600,000)x3 Ref[GRCh38]10q11.22(47,560,011-47,592,990)x3	40.00 32.98	—	—	(0.15)[f]	28260531	VUS
Coverage Collins2017_DUP_1917	DoC[GRCh38]10q26.3(133,700,000-133,797,422)x3 Ref[GRCh38]10q26.3(133,664,165-133,773,008)x3	97.42 108.84	—	—	(0.73)[f]	28260531	VUS
Coverage Collins2017_DUP_2593	DoC[GRCh38]16p11.2(32,850,000-32,880,000)x3 Ref[GRCh38]16p11.2(32,859,125-32,883,735)x3	30.00 24.61	—	—	(0.15)[f]	28260531	VUS
Cluster (32) Coverage/Array Collins2017_DEL_5923	Clu[GRCh38] 20q12(42,549,659-42,621,186)x1 DoC[GRCh38] 20q12(42,550,000-4,2620,000)x1 Ref[GRCh38] 20q12(42,549,279-42,621,903)x1	71.53 70.00 72.62	—	*PTPRT*	nsv519654 (0.25)[e]	19592680	VUS
	At liGS resolution <30 kb (<1% frequency)						
Coverage Collins2017_DUP_124	DoC[GRCh38]1p21.1(105,468,000-105,483,000)x3 Ref[GRCh38]1p21.1(105,471,438-105,483,407)x3	15.00 11.97	—	—	(0.44)[f]	28260531	VUS
Coverage Collins2017_DUP_862	DoC[GRCh38]5p15.2(12,810,000-12,822,000)x3 Ref[GRCh38]5p15.2(12,809,754-12,822,081)x3	12.00 12.33	7	—	(0.29)[f]	28260531	VUS
Coverage Collins2017_DUP_1350	DoC[GRCh38]7q36.1(150,945,000-150,960,000)x3 Ref[GRCh38]7q36.1(150,946,081-150,959,579)x3	15.00 13.49	—	*KCNH2*	(0.15)[f]	28260531	VUS
Cluster(5) Coverage Collins2017_DEL_2724	Clu[GRCh38]7p12.2(49875020-49885550)x1 DoC[GRCh38]7p12.2(49,875,000-49,888,000)x1 Ref[GRCh38]7p12.2(49,873,862-49,886,430)x1	10.53 13.00 12.57	—	*VWC2*	(0.29)[f]	28260531	VUS
Coverage Collins2017_DUP_1318	DoC[GRCh38]7q31.1(109,797,000-109,812,000)x3 Ref[GRCh38]7q31.1(109,794,416-109,815,260)x3	15.00 20.85	1	—	(0.87)[f]	28260531	VUS
Coverage Collins2017_DUP_1972	DoC[GRCh38]11p14.1(28,983,000-28,995,000)x3 Ref[GRCh38]11p14.1(28,982,519-28,994,389)x3	12.00 11.87	3	—	(0.15)[f]	28260531	VUS
Coverage Collins2017_DUP_1984	DoC[GRCh38]11p11.2(47,034,000-47,046,000)x3 Ref[GRCh38]11p11.2(47,033,616-47,045,317)x3	12.00 11.70	7	—	(0.44)[f]	28260531	VUS
Cluster (8) Coverage	Clu[GRCh38]11q12.1(58,336,732-58,348,764)x1 DoC[GRCh38]11q12.1(58,337,000-58,350,000)x1	12.03 13.00	—	—	novel	—	VUS

Coverage Collins2017_DEL_5835	DoC[GRCh38]19q13.43(56,960,000-56,970,000)x1 Ref[GRCh38]19q13.43(56,959,256-56,969,428)x1	10.00 10.17	—	—	esv3556681 (0.13)[e]	23714750	VUS

Analysis performed with CNV-ConTool with minimum reciprocal overlap of 70%. [a]Structural variants ID form the SVref dataset is according to Collins et al. (2017); Values in parentheses indicate number of read-pairs per sequence cluster. [b]Clu, sequence cluster; DoC, depth-of-coverage. [c]Number of subjects with similar genomic alteration in the internal reference group (N=32). [d]Genes with OMIM number, with or without an associated phenotype. [e]Frequency of variant from DGV public database (MacDonald et al. 2014) and [f]Frequency of variant from SVref dataset (Collins et al, 2017).[g] CNV classification according to American College of Medical Genetics and Genomics and the Clinical Genome Resource (Riggs et al. 2019). VUS – Variant of unknown significance.

Supplementary Table 14 Cross-validated, probands'-specific del and dup identified in the retrospectively analyzed subjects at both resolutions

Cluster/ Coverage/Array [a]	Genomic variants [GRCh38][b]	Size (kb)	OMIM gene or phenotype[c]	ACMG Classification[d]
	At clinical resolution of ≥30 kb			
	DGRC0006 - t(8;14)			
Cluster (5)	Clu[GRCh38]12q23.3(105,626,312-105,679,417)x1	53.11	—	VUS
Coverage	DoC[GRCh38]12q23.3(105,627,000-105,680,000)x1	53.00		
	DGRC0013 - inv(13)			
Cluster (24)	Clu[GRCh38]7q36.1(148,576,804-148,659,035)x1	82.23	—	VUS
Coverage/Array	DoC[GRCh38]7q36.1(148,577,000-148,660,000)x1	83.00		
	DGRC0030 – t(1;3)			
Cluster (31)	Clu[GRCh38]1p36.22(10,478,274-10,524,390)x3	46.12	*PEX14*	VUS
Coverage	DoC[GRCh38]1p36.22(10,480,000-10,521,000)x3	41.00		
	At liGS resolution <30 kb			
	DGRC0006 - t(8;14)			
Cluster (32)	Clu[GRCh38]2p23.1(30,166,431-30,183,788)x1	17.36	—	VUS
Coverage	DoC[GRCh38]2p23.1(30,166,000-30,184,000)x1	18.00		
Cluster (26)	Clu[GRCh38]Xq27.2(141,880,897-141,895,782)x3	14.89	*MAGEC3*	VUS
Coverage	DoC[GRCh38]Xq27.2 (141,884,000-141,894,000)x3	10.00		
	DGRC0013 - inv(13)			
Cluster(9)	Clu[GRCh38]4q28.2(128,771,573-128,784,988)x3	13.42	—	VUS
Coverage	DoC[GRCh38]4q28.2(128,772,000-128,784,000)x3	12.00		

Analysis performed with the reported workflow and CNV-ConTool with minimum reciprocal overlap of 70%. [a]Values in parentheses indicate number of read-pairs per sequence cluster. [b]Clu, sequence cluster; DoC, Depth-of-Coverage. [c]Genes with OMIM number, with or without an associated phenotype. [d] CNV classification according to American College of Medical Genetics and Genomics and the Clinical Genome Resource (Riggs et al. 2019). VUS – Variant of unknown significance.

Supplementary Table 15 Probands'-specific inv, ins and cx variants identified in retrospectively analyzed subjects

Alteration	Cluster[a]	Genomic variants [GRCh38][b]	Size (kb)	OMIM gene or phenotype[c]	Other genes and non coding RNAs
	DGRC0006 - t(8;14)				
inv Fully Rslv[d]	Cluster (30)	Clu[GRCh38]4p13(43,736,343-43,752,624)inv	16.28	—	—
inv Fully Rslv	Cluster (20)	Clu[GRCh38]1p13.2(112,873,541-112,878,455)inv	4.91	—	*RP3-522D1.1*
interchr ins Fully Rslv	Cluster (6)	Excision: Clu[GRCh38]8q22.3(100,715,571-100,717,804) Insertion: Clu[GRCh38]3q25.31(155,308,345-155,310,600)	2.23	*PABPC1* —	—
dellNVdel Fully Rslv	Cluster (6)	Clu[GRCh38]6q16.3(100,113,153-100,115,762)del Clu[GRCh38]6q16.3(100,115,763-100,115,785)inv Clu[GRCh38]6q16.3(100,115,786-100,119,075)del	5.92	—	—
	DGRC0013 - inv(13)				
interchr ins Fully Rslv	Cluster (13)	Excision: Clu[GRCh38]9q34.11(129,424,605-129,444,016) Insertion: Clu[GRCh38]22q11.23(23,510,309-23,510,596)	19.41	—	*RP3-65J3*
interchr. ins Partly Rslv	Cluster (24)	Excision: Clu[GRCh38]2q31.2(177,984,262-?) Insertion: Clu[GRCh38]12q13.2(55,390,573-?)	—	*PDE11A*	—
dellNVdel Fully Rslv	Cluster (3)	Clu[GRCh38]12q13.2(49,784,688-49,788,877)del Clu[GRCh38]12q13.2(49,788,877-49,788,985)inv Clu[GRCh38]12q13.2(49,788,986-49,792,873)del	8.18	*NCKAP5L*	—
dellNVdel Fully Rslv	Cluster (4)	Clu[GRCh38]13q12.11(22,446,323-22,449,632)del Clu[GRCh38]13q12.11(22,449,633-22,449,661)inv Clu[GRCh38]13q12.11(22,449,662-22,452,615)del	6.29	—	—
	DGRC0025 – t(12;17)				
inv Fully Rslv	Cluster (7)	Clu[GRCh38]3p21.31(50,130,322-50,135,115)inv	4.80	—	*SEMA3F-AS1*
inv Fully Rslv	Cluster (3)	Clu[GRCh38]3p12.3(79,609,499-79,616,630)inv	7.13	*ROBO1*	—
inv Fully Rslv	Cluster (5)	Clu[GRCh38]3q25.32(159,075,698-159,081,455)inv	5.76	*IQCJ*	—
inv Fully Rslv	Cluster (5)	Clu[GRCh38]7q31.32(121,520,333-121,526,336)inv	6.00	—	—
inv Fully Rslv	Cluster (4)	Clu[GRCh38]10q21.3(65,664,415-65,671,542)inv	7.13	—	*LINC01515*
inv Fully Rslv	Cluster (5)	Clu[GRCh38]15q26.2(97,353,880-97,359,017)inv	5.14	—	*LINC02253*
interchr ins Fully Rslv	Cluster (7)	Excision: Clu[GRCh38]3p11.1(89,459,946-89,460,378) Insertion: Clu[GRCh38]1p12(119,252,479-119,255,488)	0.432	*EPHA3*	*RP11-418J17*
interchr ins Fully Rslv	Cluster (8)	Excision: Clu[GRCh38]10p12.33(17,548,358-17,548,457) Insertion: Clu[GRCh38]3p21.31(49,401,798-49,406,995)	0.099	— *RHOA*	—

	DGRC0030 – t(1;3)				
inv Fully Rslv	Cluster (3)	Clu[GRCh38]5q32(146,449,266-146,454,922)inv	5.66	*TCERG1*	—
inv Fully Rslv	Cluster (4)	Clu[GRCh38]7p12.1(52,318,497-52,324,132)inv	5.64	—	—
delINVdel Fully Rslv	Cluster (4)	Clu[GRCh38]11p15.5(1,351,552-1,354,130)del Clu[GRCh38]11p15.5(1,354,130-1,354,199)inv Clu[GRCh38]11p15.5(1,354,199-1,356,719)del	5.17	—	*AC136297.1*

Analysis performed with the reported workflow and CNV-ConTool with minimum reciprocal overlap of 70%. Individual SVs within a cx variant are designated in lowercase and uppercase letters. [a]Values in parentheses indicate number of read-pairs per sequence cluster. [b]Clu, sequence cluster. [c]Genes with OMIM number, without or without an associated phenotype. [d]Inversion breakpoint is validated by a split-read. Rslv, Resolved.

Supplementary Table 16 Expression of genes within disrupted hESC TADs in LCLs with DGRC0016 t(16;17)

Transcript Cluster ID	Gene Symbol	Description	Group	Strand	Chr.	Genomic Position	Bi-weight Avg Signal (log2)		LCLs SD
							Proband's LCL	LCLs (n=4)	
The 16q24.3 breakpoint within hESC TAD at chr16:89246091-89686091									
TC16001350.hg.1	***ANKRD11***	**Ankyrin repeat domain 11**	**Coding**	**-**	**16q24.3**	**89,267,619-89,490,561**	**7.77**	**7.85**	**0.09**
TC16000695.hg.1	*LOC100287036*	Uncharacterized LOC100287036	Coding	+	16q24.3	89,321,133-89,325,110	4.03	4.03	0,05
TC16001711.hg.1	*LOC101927817*	Uncharacterized LOC101927817	NonCoding	+	16q24.3	89,430,918-89,454,494	5.15	5.23	0.15
TC16000696.hg.1	*LOC101927817*	Uncharacterized LOC101927817	Coding	+	16q24.3	89,430,918-89,454,494	4.81	4.85	0,12
TC16000698.hg.1	*SPG7; LOC101930112*	Spastic paraplegia 7 (pure and complicated autosomal recessive); uncharacterized LOC101930112	Coding	+	16q24.3	89,490,719-89,557,768	7.13	7.12	0
TC16001713.hg.1	*RPL13*	Ribosomal protein L13	NonCoding	+	16q24.3	89,560,657-89,566,828	14.27	15.06	0.77
TC16000699.hg.1	*SNORD68*	Ribosomal protein L13; small nucleolar RNA, C/D box 68	Coding	+	16q24.3	89,561,434-89561517	7.13	7.11	0.13
TC16000700.hg.1	*CPNE7*	Copine VII	Coding	+	16q24.3	89,575,768-89,597,246	6.86	6.92	0.380
TC16000701.hg.1	*DPEP1*	Dipeptidase 1 (renal)	Coding	+	16q24.3	89,613,308-89,638,456	5.4	5.4	0.02
TC16001353.hg.1	*CHMP1A*	Charged multivesicular body protein 1A	Coding	-	16q24.3	89,644,431-89,657,845	7.78	7.73	0.24
TC16000702.hg.1	*SPATA33*	Spermatogenesis associated 33; chromosome 16 open reading frame 55	Coding	+	16q24.3	89,657,802-89,671,272	5.17	5.17	0
TC16000703.hg.1	*CDK10*	Cyclin-dependent kinase 10	Coding	+	16q24.3	89,680,737-89,696,364	6.86	6.9	0.24
17q21.31 breakpoint within hESC TAD at chr17:46687454-47647635									
TC17000600.hg.1	*NSF; NSFP1; LOC101930324*	N-ethylmaleimide-sensitive factor; N-ethylmaleimide-sensitive factor pseudogene 1	Coding	+	17q21.31	46,590,669-46,757,464	9.53	9.56	0.27
TC17001615.hg.1	***WNT3; LOC101929777***	**Wingless-type MMTV integration site family, member 3; uncharacterized LOC101929777**	**Coding**	**-**	**17q21.31**	**46,762,506-46,833,154**	**5.86**	**4.6**	**0.01**
TC17000601.hg.1	*WNT9B*	Wingless-type MMTV integration site family, member 9B	Coding	+	17q21.32	46,833,201-46,886,730	4.69	4.69	0.05
TC17000603.hg.1	*GOSR2*	Golgi SNAP receptor complex member 2	Coding	+	17q21.32	46,923,075-46,975,524	9.26	9.25	0.12
TC17001616.hg.1	*RPRML*	Reprimo-like	Coding	-	17q21.32	46,978,156-46,979,248	4.83	4.81	0.14
TC17001621.hg.1	*CDC27*	Cell division cycle 27	Coding	-	17q21.32	47,117,703-47,189,422	11.78	11.72	0.24
TC17000606.hg.1	*MYL4*	Myosin, light chain 4, alkali; atrial, embryonic	Coding	+	17q21.32	47,200,446-47,223,679	4.57	4.57	0.06
TC17002878.hg.1	*ITGB3*	Integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	Coding	+	17q21.32	47,253,846-47,311,816	7.45	7.59	1.5
TC17002879.hg.1	*EFCAB13; C17orf57*	EF-hand calcium binding domain 13; chromosome 17 open reading frame 57	Coding	+	17q21.32	47,323,290-47,441,312	6.51	6.37	0.37
TC17001625.hg.1	MRPL45P2	Mitochondrial ribosomal protein L45 pseudogene	Coding	-	17q21.32	47,450,568-47,492,492	8.27	8.34	0.2
TC17000609.hg.1	*NPEPPS; LOC100653042*	Aminopeptidase puromycin sensitive; uncharacterized LOC100653042	Coding	+	17q21.32	47,522,942-47,623,276	11.49	11.49	0.12

Genes disrupted by breakpoints are in bold. A statistically significant linear fold change of 2.6 of the disrupted *WNT3* was identified.

Supplementary Table 17 Protein coding and non-coding RNA genes localized within hESC TADs disrupted by DGRC0006 t(8;14) breakpoints, their associated phenotypes and GWAS data from the referred regions

Genes and intergenic regions			Clinical Phenotype	GWAS SNP data
GeneCard	OMIM	oe (CI 95%)	Phenotype;#OMIM_Inheritance; DDG2P_category	Genetic traits
8q12.3 breakpoint within hESC TAD at 8:63204888-64444889				
YTHDF3	na	0.05 (0.02-0.23)	None	None
AC011978	na	nd	None	None
AC011124	na	nd	None	None
AC018953	na	nd	None	None
Intergenic - chr8:63589408-63687179			None	# 1 SNPs - EFO:0004342_Waist circumference [6.92E-6]; DOID:9970_Obesity [6.92E-6]
AC069133	na	nd	None	None
LINC01289	na	nd	None	# 2 SNPs - EFO:0004342_Waist circumference [6.92E-6]; DOID:9970_Obesity [6.92E-6]
RP11-32K4.1-IVS1	na	nd	None	# 2 SNPs - EFO:0004342_Waist circumference [6.92E-6]; DOID:9970_Obesity [6.92E-6]
der(8)g.[chr8:pter_cen_64209134::chr14:83126596_83126598dup_qter]				
Intergenic - chr8:64209111-64373328			None	# 49 SNPs - EFO:0004342_Waist circumference [6.92E-6]; DOID:9970_Obesity [6.92E-6]
MIR124-2HG	na	nd	None	# 1 SNPs - EFO:0004342_Waist circumference [6.92E-6]; DOID:9970_Obesity [6.92E-6]
14q31.2 breakpoint within hESC interTAD region at chr14:83103903-83503903				
der(14) g.[chr14:pter_cen_83126598::chr8:64209135_qter]				None

Table prepared with TAD-GConTool that includes GWAS SNP data FANTOM CAT database (Hon et al. 2017). SNP associated genetic traits are designated according to their ontology term ID. All p-values are below GWAS significance threshold of ≤ 5.00E−8. Clinical phenotype data are from OMIM and DDG2P databases. The clinical phenotype data are configured as follows: i) gene-associated phenotypes in humans from either database; ii) #OMIM phenotype number underscore type of inheritance (AR, AD or AR/AD) or #NA in the absence of OMIM phenotype number; and DDG2P underscore phenotype classification (confirmed, probable and possible) or DDG2P_NA in the absence of DDG2P data.

Supplementary Table 18 Protein coding and non-coding RNA genes localized within hESC TADs disrupted by DGRC0013 inv(13) breakpoints with their associated phenotypes and GWAS data from the referred regions

Genes and intergenic regions			Clinical Phenotype	GWAS SNP data
GeneCard	**OMIM**	**oe (CI 95%)**	**Phenotype;#OMIM_Inheritance; DDG2P_category**	**Genetic traits**
13q12.3 breakpoint within hESC TAD at 13:28007863-28807863				
FLT3	136351	0.22 (0.14-0.35)	Leukemia, acute lymphoblastic, somatic; #613065_nd; DDG2P_NA Leukemia, acute myeloid, somatic; #601626_nd; DDG2P_NA	None # 1 SNPs - HP:0000707_Abnormality of the nervous system [9.64E-13]; EFO:0004329_Alcohol drinking [9.64E-13]; DOID:1574_Alcohol abuse [9.64E-13]; DOID:0050741_Alcohol dependence [9.64E-13]
Intergenic - chr13:28100592-28138506			None	
PAN3	617448	0.05 (0.02-0.15)	None	None
FLT1 - IVS1	165070	0.14 (0.09-0.24)	None	**# 28 SNPs - MESH:D003327_Coronary disease [6E-07]** # 24 SNPs - EFO:0004731_Eye measurement [9.75E-7]; DOID:2462_Retinal vascular disease [9.75E-7] **# 3 SNPs - HP:0000707_Abnormality of the nervous system [9.64E-13]**; EFO:0004329_Alcohol drinking [9.64E-13]; DOID:1574_Alcohol abuse [9.64E-13]; DOID:0050741_Alcohol dependence [9.64E-13] # 1 SNPs - MESH:D008076_Cholesterol hdl [1.24E-6]; EFO:0004612_High density lipoprotein cholesterol measurement [1.24E-6]; DOID:1461_Cholesterol embolism [1.24E-6]
g.[pter_cen_28489792_74831812inv;74831813_74831818del_qter]				
Intergenic - chr13:28489792-28659104			None	# 18 SNPs - HP:0003002_Breast carcinoma [7.90E-6] # 2 SNPs - DOID:1574_Alcohol abuse [9.64E-13]; DOID:0050741_Alcohol dependence [9.64E-13]; HP:0000707_Abnormality of the nervous system [9.64E-13]; EFO:0004329_Alcohol drinking [9.64E-13] # 1 SNPs - DOID:12603_Acute leukemia [6E-6]; DOID:9952_Acute lymphocytic leukemia [6E-6]
POMP	613386	0.00 (0.00-0.39)	Keratosis linearis with ichthyosis congenita and sclerosing keratoderma; #601952_AR; DDG2P_ possible Proteasome-associated autoinflammatory syndrome 2; #618048_**AD**; DDG2P_NA	None
Intergenic - chr13:28678925-28700064			None	None
SLC46A3	616764	0.75	None	None
Intergenic - chr13:28718970-28807863			None	None
13q22.1 breakpoint within hESC TAD at 13:73727862-75327863				
KLF12	607531	0.06 (0.02-0.28)	None	**# 9 SNPs - HP:0005344_Abnormality of the carotid arteries [7.90E-6]** # 6 SNPs - HP:0001699_Sudden death [1E-8]; DOID:10273_Heart conduction disease [1E-8] # 1 SNPs - HP:0001645_Sudden cardiac death [5E-20]
Intergenic - chr13:73995056-74231457			None	# 49 SNPs - HP:0001645_Sudden cardiac death [5E-20] # 1 SNPs - DOID:8986_Narcolepsy [3.59E-25]; HP:0100786_Hypersomnia [3.59E-25]
LINC00402	na	nd	None	None

Intergenic - chr13:74259976-74412957			None	None
AL355390	na	nd	None	None
LINC00381	na	nd	None	None
Intergenic - chr13:74435159-74552503			None	**# 68 SNPs - HP:0005368_Abnormality of humoral immunity [3E-7]; DOID:8736_Smallpox [3E-7]; EFO:0004873_Response to vaccine [3E-7]** # 11 SNPs - DOID:9352_Type 2 diabetes mellitus [1E-8]; EFO:0004639_Phospholipid measurement [1E-08]
LINC00347	na	nd	None	# 10 SNPs - EFO:0004639_Phospholipid measurement [1E-08]; DOID:9352_Type 2 diabetes mellitus [1E-8]
Intergenic - chr13:74565445-74831722			None	# 2 SNPs - EFO:0004639_Phospholipid measurement [1E-8]; DOID:9352_Type 2 diabetes mellitus [1E-8]
g.[pter_cen_28489792_74831812inv;74831813_74831818del_qter]				
Intergenic - chr13:74831722-75250480			None	**# 6 SNPs - DOID:6000_Congestive heart failure [6E-7]; MESH:D006333_Heart failure [6E-7]**
LINC01078	na	nd	None	None
Intergenic - chr13:75252012-75284665			None	None
TBC1D4	612465	0.71	Diabetes mellitus, noninsulin-dependent, 5; #616087_nd; DDG2P_NA	None # 1 SNPs - DOID:9744_Type 1 diabetes mellitus [6E-11]

Table prepared with TAD-GConTool that includes GWAS SNP data from FANTOM CAT database (Hon et al. 2017). SNP associated genetic traits are presented according to their ontology term ID. Although conventional GWAS significance threshold is ≤ 5.00E-8, for disrupted genes, adjacent genomic regions and for genetic traits identified in GWAS (in bold), overlapping the proband's clinical phenotype, a threshold of 9.99E-6 is used. Clinical phenotype data are from OMIM and DDG2P databases. The clinical phenotype data are configured as follows i) gene-associated phenotypes in humans from either database; ii) #OMIM phenotype number underscore type of inheritance (AR, AD or AR/AD) or #NA in the absence of OMIM phenotype number; and DDG2P underscore phenotype classification (confirmed, probable and possible) or DDG2P_NA in the absence of DDG2P data.

Supplementary Table 19 Protein coding and non-coding RNA genes localized within hESC TADs disrupted by DGRC0025 t(12;17) breakpoints with their associated phenotypes and GWAS data from the referred regions

Genes and intergenic regions			Clinical Phenotype	GWAS SNP data
GeneCard	OMIM	oe (CI 95%)	Phenotype;#OMIM_Inheritance; DDG2P_category	Genetic traits
12q23.1 breakpoint within hESC TAD at 12:99042091-99802091				
ANKS1B – IVS9	607815	0.10 (0.05-0.20)	None	# 119 SNPs - DOID:5419_Schizophrenia [4E-7] # 90 SNPs - DOID:1094_Attention deficit hyperactivity disorder [2E-6]; HP:0007302_Bipolar affective disorder [2E-6]; HP:0000729_Autistic behavior [2E-6]; DOID:1470_Major depressive disorder [2E-6] # 33 SNPs - DOID:9970_Obesity [7E-6] # 6 SNPs - DOID:332_Amyotrophic lateral sclerosis [8E-6]
der(12) g.[chr12:pter_cen_99637772_99637781del::chr17:51565697_qter]				
FAM71C	na	0.31 (0.13-0.99)	None	None
Intergenic - chr12:99650046-99802091			None	None
17q21.33 breakpoint within hESC TAD at 17:51287640-54967640				
UTP18	612816	0.37	None	None
LINC02071	na	nd	None	None
AC005823	na	nd	None	None
LINC02072	na	nd	None	None
LINC02073	na	nd	None	None
Intergenic - chr17:51445802-51565697			None	# 4 SNPs - DOID:9352_Type 2 diabetes mellitus [7E-13]; HP:0000140_Abnormality of the menstrual cycle [7E-13]; MESH:D011293_Premenstrual syndrome [7E-13]; DOID:10652_Alzheimer s disease [7E-13]; DOID:11476_Osteoporosis [7E-13]; DOID:9970_Obesity [7E-13]; DOID:1612_Breast cancer [7E-13]
der(17) g.[chr17:pter_cen_51565696::chr12:99637782_qter]				
CA10	604642	0.21 (0.10-0.47)	None	None
Intergenic - chr17:52160017-52390515			None	None
LINC01982	na	nd	None	None
LINC02089	na	nd	None	None
C17orf112	na	0.28 (0.10-1.30)	None	None
Intergenic - chr17:52987652-53822901			None	None
KIF2B	615142	0.89	None	None
Intergenic - chr17:53825213-54899387			None	# 24 SNPs - DOID:1574_Alcohol abuse [7.87E-20]; DOID:0050741_Alcohol dependence [7.87E-20]; HP:0000707_Abnormality of the nervous system [7.87E-20]; EFO:0004329_Alcohol drinking [7.87E-20]
TOM1L1	604701	0.68	None	# 25 SNPs - HP:0003002_Breast carcinoma [2E-13]
COX11	603648	0.40	None	# 11 SNPs - HP:0003002_Breast carcinoma [2E-13]

Table prepared with TAD-GConTool that includes GWAS SNP data FANTOM CAT database (Hon et al. 2017). SNP associated genetic traits are presented according to their ontology term ID. Although conventional GWAS significance threshold is ≤ 5.00E-8, for disrupted genes and adjacent genomic regions a threshold of 9.99E-6 is used. Clinical phenotype data are from OMIM and DDG2P databases.

Supplementary Table S20 Protein coding and non-coding RNA genes localized within hESC TADs disrupted by DGRC0030 t(1;3) breakpoints with their associated phenotypes and GWAS data from the referred regions

Genes and intergenic regions			Clinical Phenotype	GWAS SNP data
GeneCard	OMIM	oe (CI 95%)	Phenotype; #OMIM_Inheritance; DDG2P_category	Genetic traits
1q42.11 breakpoint within hESC TAD at 1:224025675-224425675				
Intergenic - chr1:224025675-224114087			None	# 10 SNPs - DOID:614_Lymphopenia [3.66E-8]; DOID:2841_Asthma [3.66E-8]; HP:0002665_Lymphoma [3.66E-8]; DOID:1240_Leukemia [3.66E-8]
FBXO28	609100	0.06 (0.02-0.29)	None	# 43 SNPs - DOID:1240_Leukemia [3.66E-8]; HP:0002665_Lymphoma [3.66E-8]; DOID:614_Lymphopenia [3.66E-8]; DOID:2841_Asthma [3.66E-8]
Intergenic - chr1:224162047-224175756			None	# 7 SNPs - DOID:1240_Leukemia [3.66E-8]; DOID:2841_Asthma [3.66E-8]; HP:0002665_Lymphoma [3.66E-8]; DOID:614_Lymphopenia [3.66E-8]
DEGS1	615843	0.30 (0.14-0.77)	Leukodystrophy, hypomyelinating, 18; #618404_AR;DDG2P_NA	# 1 SNPs - DOID:1240_Leukemia [3.66E-8]; HP:0002665_Lymphoma [3.66E-8]; DOID:614_Lymphopenia [3.66E-8]; DOID:2841_Asthma [3.66E-8]
NVL	602426	0.70	None	# 74 SNPs - DOID:9255_Frontotemporal dementia [5.30E-26]; MESH:D009203_Myocardial infarction [5.30E-26]
Intergenic - chr1:224330387-224356850			None	# 33 SNPs - DOID:9255_Frontotemporal dementia [5.30E-26]; MESH:D009203_Myocardial infarction [5.30E-26]
CNIH4	617483	0.43	None	None
WDR26 - IVS5	617424	0.00 (0.00-0.08)	**Skraban-Deardorff syndrome; #617616_AD;DDG2P_confirmed**	# 1 SNPs - DOID:3083_Chronic obstructive pulmonary disease [4.05E-6]; MESH:D008171_Lung diseases [4.05E-6]; DOID:2841_Asthma [4.05E-6]; HP:0002795_Functional respiratory abnormality [4.05E-6]; EFO:0004713_Forced expiratory volume [4.05E-6]
der(1) g.[chr1:pter_cen_224398161_224398171del::chr3:pter_10670893inv]				
3p25.3 breakpoint within hESC TAD at 3:10343316-10743315				
ATP2B2 - IVS1	108733	0.06 (0.03-0.15)	Deafness, autosomal recessive 12, modifier of; #601386_AR;DDG2P_NA	# 73 SNPs - HP:0100786_Hypersomnia [2.51E-9]; DOID:8986_Narcolepsy [2.51E-9] # 8 SNPs - MESH:D015992_Body mass index [2.82E-6]; HP:0001507_Growth abnormality [2.82E-6]; DOID:9970_Obesity [2.82E-6]] # 7 SNPs - DOID:332_Amyotrophic lateral sclerosis [7E-6] # 2 SNPs - DOID:9975_Cocaine dependence [3.30E-6]; HP:0000707_Abnormality of the nervous system [3.30E-6]
der(3) g.[chr1:pter_224398172inv::chr3: 10670894_10670895del_cen_qter]				
Intergenic - chr3:10670895-10743315			None	# 12 SNPs - DOID:1561_Cognitive disorder [1.02E-6]

Table prepared with TAD-GConTool that includes GWAS SNP data from FANTOM CAT database (Hon et al. 2017). Genes showing high probability of being AD disease causing are in bold (McKusick 1998; Wright et al. 2015). SNP associated genetic traits are presented according to their ontology term ID. Although conventional GWAS significance threshold is ≤ 5.00E-8, for disrupted genes and adjacent genomic regions a threshold of 9.99E-6 is used. Clinical phenotype data are from OMIM and DDG2P databases. The clinical phenotype data are configured as follows: i) gene-associated phenotypes in humans from either database; ii) #OMIM phenotype number underscore type of inheritance (AR, AD or AR/AD) or #NA in the absence of OMIM phenotype number; and DDG2P underscore phenotype classification (confirmed, probable and possible) or DDG2P_NA in the absence of DDG2P data.

Supplementary References

Auton A, Abecasis G, Altshuler D, et al. (2015) A global reference for human genetic variation. Nature 526:68-74.

Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA, Mouse Genome Database Group. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. Nucleic Acids Res 36:D724-728.

Carbonell AU, Cho CH, Tindi JO, Counts PA, Bates JC, Erdjument-Bromage H, Cvejic S, Iaboni A, Kvint I, Rosensaft J, Banne E, Anagnostou E, Neubert TA, Scherer SW, Molholm S, Jordan BA. (2019) Haploinsufficiency in the ANKS1B gene encoding AIDA-1 leads to a neurodevelopmental syndrome. Nat Commun 10:3529.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422-1423.

Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout ATV, Bosco P, Friend KL, Baker C, Buono S, Vissers LELM, Schuurs-Hoeijmakers JH, Hoischen A, Pfundt R, Krumm N, Carvill GL, Li D, Amaral D, Brown N, Lockhart PJ, Scheffer IE, Alberti A, Shaw M, Pettinato R, Tervo R, de Leeuw N, Reijnders MRF, Torchia BS, Peeters H,Thompson E, O'Roak BJ, Fichera M, Hehir-Kwa JY, Shendure J, Mefford HC, Haan E, Gécz J, de Vries BBA, Romano C, Eichler EE. (2014) Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet 46:1063-1071.

Collins RL, Stone MR, Brand H, Glessner JT, Talkowski ME. (2016) CNView: a visualization and annotation tool for copy number variation from whole-genome sequencing. *bioRxiv* 049536; (CNView, https://github.com/RCollins13/CNView).

Collins RL, Brand H, Redin CE, Hanscom C, Antolik C, Stone MR, Glessner JT, Mason T, Pregno G, Dorrani N, Mandrile G, Giachino D, Perrin D, Walsh C, Cipicchio M, Costello M, Stortchevoi A, An JY, Currall BB, Seabra CM, Ragavendran A, Margolin L, Martinez-Agosto JA, Lucente D, Levy B, Sanders SJ, Wapner RJ, Quintero-Rivera F, Kloosterman W, Talkowski ME. (2017) Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. Genome Biol 18:36.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niyazov D, Leppig K, Thiese H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE. (2011) A copy number variation morbidity map of developmental delay. Nat Genet 43:838-846.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376-380.

Durand NC, Robinson JT, Shamim, MS, Machol I, Mesirov JP, Lander ES, Aiden EL. (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Systems 3:99-101.

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Vooren SV, Moreau Y, Pettett RM, Carter NP. (2009) DECIPHER: database of chromosomal

imbalance and phenotype in humans using Ensembl resources. Am J Hum Genet 84:524-533.
Hodgetts Morton V, Quinlan-Jones E, Butts N, Williams D, Hamilton S, Marton T, Morris K. (2017) The first antenatal diagnosis of KBG syndrome: a microdeletion at chromosome 16q24.2q24.3 containing multiple genes including ANKRD11 associated with the disorder. Clin Case Rep 6:189-191.
Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J, Lizio M, Kawaji H, Kasukawa T, Itoh M, Burroughs AM, Noma S, Djebali S, Alam T, Medvedeva YA, Testa AC, Lipovich L, Yip CW, Abugessaisa I, Mendez M, Hasegawa A, Tang D, Lassmann T, Heutink P, Babina M, Wells CA, Kojima S, Nakamura Y, Suzuki H, Daub CO, de Hoon MJ, Arner E, Hayashizaki Y, Carninci P, Forrest AR. (2017) An atlas of human long non-coding RNAs with accurate 5' ends. Nature 543:199-204.
Huang N, Lee I, Marcotte EM, Hurles ME. (2010) Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 6:e1001154.
Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mulle JG, Warren ST, Richard G, Compton JG, Fuller AE, Gliem TJ, Huang S, Collinson MN, Beal SJ, Ackley T, Pickering DL, Golden DM, Aston E, Whitby H, Shetty S, Rossi MR, Rudd MK, South ST, Brothman AR, Sanger WG, Iyer RK, Crolla JA, Thorland EC, Aradhya S, Ledbetter DH, Martin CL. (2011) An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med 13:777-784.
Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. (2016) ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 44:D862-868.
Layer RM, Chiang C, Quinlan AR, Hall IM. (2014) LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15:R84.
Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285-291.
Li H, Durbin R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.
MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 42:D986-992.
McGowan-Jordan J, Simons A, Schmid M. (2016) ISCN 2016: An International system for human cytogenomic nomenclature. Karger, Basel, New York.

McKusick VA. (1998) Mendelian inheritance in Man. A catalog of human genes and genetic disorders. 12th edn. Johns Hopkins University Press, Baltimore, MD.

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla JA, Eichler EE, Epstein CJ, Faucett WA, Feuk L, Friedman JM, Hamosh A, Jackson L, Kaminsky EB, Kok K, Krantz ID, Kuhn RM, Lee C, Ostell JM, Rosenberg C, Scherer SW, Spinner NB, Stavropoulos DJ, Tepperberg JH, Thorland EC, Vermeesch JR, Waggoner DJ, Watson MS, Martin CL, Ledbetter DH. (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.

Moore BL, Aitken S, Semple CA. (2015) Integrative modeling reveals the principles of multiscale chromatin boundary formation in human nuclear organization. Genome Biol 16,110.

Novara F, Rinaldi B, Sisodiya SM, Coppola A, Giglio S, Stanzial F, Benedicenti F, Donaldson A, Andrieux J, Stapleton R, Weber A, Reho P, van Ravenswaaij-Arts C, Kerstjens-Frederikse WS, Vermeesch JR, Devriendt K, Bacino CA, Delahaye A, Maas SM, Iolascon A, Zuffardi O. (2017) Haploinsufficiency for ANKRD11-flanking genes makes the difference between KBG and 16q24.3 microdeletion syndromes: 12 new cases. Eur J Hum Genet 25:694-701.

Ordulu Z, Wong KE, Currall BB, Ivanov AR, Pereira S, Althari S, Gusella JF, Talkowski ME, Morton CC. (2014) Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. *Am* J Hum Genet 94:695–709.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159:1665-1680.

Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, Raca G, Ritter DI, South ST, Thorland EC, Pineda-Alvarez D, Aradhya S, Martin CL. (2019) Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med 1-13.

Quinlan AR, Hall IM. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

Sánchez-Ferrero E, Coto E, Beetz C, Gámez J, Corao AI, Díaz M, Esteban J, del Castillo E, Moris G, Infante J, Menéndez M, Pascual-Pascual SI, López de Munaín A, Garcia-Barcina MJ, Alvarez V. (2013) SPG7 mutational screening in spastic paraplegia patients supports a dominant effect for some mutations and a pathogenic role for p.A510V. Clin Genet 83:257-262.

Skraban CM, Wells CF, Markose P, Cho MT, Nesbitt AI, Au PYB, Begtrup A, Bernat JA, Bird LM, Cao K, de Brouwer APM, Denenberg EH, Douglas G, Gibson KM, Grand K, Goldenberg A, Innes AM, Juusola J, Kempers M, Kinning E, Markie DM, Owens MM, Payne K, Person R, Pfundt R, Stocco A, Turner CLS, Verbeek NE, Walsh LE, Warner TC, Wheeler PG, Wieczorek D, Wilkens AB, Zonneveld-Huijssoon E; Deciphering Developmental Disorders Study, Kleefstra T, Robertson SP, Santani A, van Gassen KLI, Deardorff MA. (2017) WDR26 haploinsufficiency causes a recognizable syndrome of intellectual disability, seizures, abnormal gait, and distinctive facial features. Am J Hum Genet 101:139-148.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. (2016) The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 54:1.30.1-1.30.33.

Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK, Shen Y, Borowsky M, Daly MJ, Morton CC, Gusella JF. (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.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. (2015) Sambamba: fast processing of NGS alignment formats. Bioinformatics 31:2032-2034.

Wright CF, Fitzgerald TW, Jones WD, Clayton S, McRae JF, Kogelenberg M, King DA, Ambridge K, Barret DM, Bayzetinova T, Bevan AP, Bragin E, Chatzimichali EA, Gribble S, Jones P, Krishnappa N, Mason LE, Miller R, Morley KI, Parthiban V, Prigmore E, Rajan D, Sifrim A, Swaminathan J, Tivey AR, Middleton A, Parker M, Carter NP, Barrett JC, Hurles ME, FitzPatrickDR, Firth HV. (2015) Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. Lancet 385:1305-1314.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. (2018) Ensembl 2018. Nucleic Acids Res 46:D754-D761.