W744–W752 *Nucleic Acids Research, 2022, Vol. 50, Web Server issue* *Published online 7 May 2022*
https://doi.org/10.1093/nar/gkac318

TADeus2: a web server facilitating the clinical diagnosis by pathogenicity assessment of structural variations disarranging 3D chromatin structure

Barbara Poszewiecka [1,*], Victor Murcia Pienkowski [2,3], Karol Nowosad [4,5,6], Jérôme D. Robin [2], Krzysztof Gogolewski [1] and Anna Gambin [1,*]

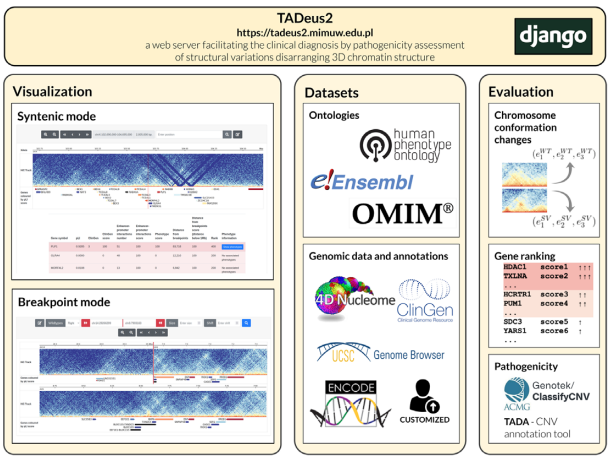
[1]Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, 2 Banacha street, 02-097 Warsaw, Poland, [2]Aix Marseille Univ, INSERM, Marseille Medical Genetics, MMG, Marseille, France, [3]Department of Medical Genetics, Medical University of Warsaw, Adolfa Pawińskiego 3c, 02-106 Warsaw, Poland, [4]Department of Cell Biology, Erasmus Medical Center, Doctor Molewaterplein 40, 3015 GD Rotterdam, Netherlands, [5]Department of Biomedical Sciences, Laboratory of Molecular Genetics, Medical University of Lublin, Doktora Witolda Chodźki 1, 20-400 Lublin, Poland and [6]The Postgraduate School of Molecular Medicine, Medical University of Warsaw, Żwirki i Wigury 61, 02-091 Warsaw, Poland

Received March 12, 2022; Revised April 12, 2022; Editorial Decision April 19, 2022; Accepted April 21, 2022

ABSTRACT

In recent years great progress has been made in identification of structural variants (SV) in the human genome. However, the interpretation of SVs, especially located in non-coding DNA, remains challenging. One of the reasons stems in the lack of tools exclusively designed for clinical SVs evaluation acknowledging the 3D chromatin architecture. Therefore, we present TADeus2 a web server dedicated for a quick investigation of chromatin conformation changes, providing a visual framework for the interpretation of SVs affecting topologically associating domains (TADs). This tool provides a convenient visual inspection of SVs, both in a continuous genome view as well as from a rearrangement's breakpoint perspective. Additionally, TADeus2 allows the user to assess the influence of analyzed SVs within flaking coding/non-coding regions based on the Hi-C matrix. Importantly, the SVs pathogenicity is quantified and ranked using TADA, ClassifyCNV tools and sampling-based *P*-value. TADeus2 is publicly available at https://tadeus2.mimuw.edu.pl.

GRAPHICAL ABSTRACT



INTRODUCTION

Motivation

Next generation sequencing (NGS) revolutionized the field of clinical genetics (1). One of the most popular approaches in the identification of molecular etiology of human diseases is sequencing of coding regions, which compose <3% of the genome (2). However, the application of whole exome sequencing (WES) characterized the genotype-phenotype correlation in only 25–40% of patients (3). This diagnostic rate could be explained by the limitation in WES targeting only coding regions, while other causes of pathology such as structural variants (SVs) and copy number variants (CNVs) are often located in non-coding DNA (4,5).

*To whom correspondence should be addressed. Tel: +48 225544521; Fax: +48 225544200; Email: aniag@mimuw.edu.pl
Correspondence may also be addressed to Barbara Poszewiecka. Email: b.poszewiecka@mimuw.edu.pl

SV and CNV might lead to disease through 4 main mechanisms: (i) direct disruption or deletion of a gene leading to haploinsufficiency (6); (ii) generation of SV/CNV derived fusion genes – a mechanism often present, but not limited to cancer cells (7–9); (iii) changes in the gene dosage due to duplication (10) and (iv) disruption of the epigenetic equilibrium caused by displacement of regulatory elements (11,12). The latter mechanism is often caused by changes in the three-dimensional (3D) chromatin structure. The 3D genome architecture controls spatiotemporal gene expression and plays a key role in the development and disease (13). At the (sub)megabase scale, the 3D chromatin structure is organized into topologically associated domains (TADs), delimited by boundaries enriched in CTCF binding sites (14,15). TADs facilitate enhancer-promoter contacts within their bodies and insulate inter-TADs chromatin interactions (16,17).

Interestingly, the disruption of TADs or the formation of a novel TADs can lead to gene deregulation associated with congenital disorders e.g. brachydactyly, a limb malformation affecting finger development (17), and Cooks syndrome, a congenital disorder affecting digits and nails (18).

Despite the characterization of the genotype-phenotype correlation in a number of diseases linked to SVs and CNVs in non-coding DNA, the evaluation of their pathogenicity caused by the disruption of long-range regulatory interactions remains challenging (12,19). More evidence points towards the fact that computational investigation of non-coding variants severity is helpful in the diagnosis of genetic disorders (20–22).

For this reason, there is a clear need for an easy to use web server that enables a quick evaluation of chromatin conformation changes and provides a visual framework for the interpretation of SVs/CNVs affecting TADs structures.

Related research

The growing number of NGS data from symptomatic SVs/CNVs carriers led to the emergence of multiple databases inspiring further studies regarding the role of gene position effect (GPE) in the patient phenotype. For example, Zepada et. al. applied the haploinsufficiency and triplosensivity scores to characterize GPE derived from balanced translocations in 17 subjects from the Developmental Genome Anatomy Project (GDAP) (23). Ruifeng *et al.* using data from ClinVar database (24) introduced the Structure Influence score to prioritize and designate SVs that are likely to disturb gene regulation through TADs disorganization. Salem et al. (25) based on the DECIPHER database (containing nearly one thousand deletions) (26) showed that only 4.5% of analyzed deletions can disrupt TAD boundaries leading to the gene misexpression due to enhancer adoption.

Investigation of NGS data from symptomatic patients carrying SVs led also to the development of multiple tools designed for clinical evaluation of SVs, including: 3Disease Browser (27), PhenogramViz (28), GeCCO (29) or rankings based on HI scores (30). However, PhenogramViz, GeCCO and HI scores do not use the chromatin conformation and regulatory data for the SV-induced pathogenecity estimation, while 3Disease Browser does not evaluate any novel SVs.

Among the newest web-services providing an extended palette of features for SVs/CNVs evaluation in the clinical context are 3D Genome Browser (31) and CNVxplorer (32). The main functionalities provided by 3D Genome Browser consist of analysis of disease-associated SVs and 3D chromatin structure by visualization and integration of Hi-C and ChIP-seq data. This tool also enables inspection of inter-chromosomal interactions. However, it is limited only to the uni-directed strand view (hg19), and does not provide any views for comparisons of the rearranged genome structure versus wild-type. The CNVxplorer mines a comprehensive set of clinical, genomic, and epigenomic features associated with CNVs making it one of the most versatile diagnosis tools available online. Nonetheless, it focuses only on CNVs and does not consider balanced rearrangements events. Summarized evaluation of tools with purpose comparable to `TADeus2` is presented in the Table 1.

Our results

To address existing challenges and provide a competitive tool, we have developed `TADeus2`—a web server for a quick evaluation of SVs/CNVs that provides a visual framework to aid the medical expert in the interpretation of variant pathogenicity in the context of changes in the TADs organization (see Graphical Abstract). Based on the type of a variant `TADeus2` allows to visualize the affected region in two modes: (i) *syntenic* mode dedicated for the analysis of deletions and duplications; (ii) *breakpoint* mode designed for translocations and inversions. The former mode allows visualization of one genomic region, while the latter enables analysis of two different genome loci joined together during the rearrangement event. Both modes integrate multiple datasets (e.g. Hi-C or ChIPseq) either available on the server or provided by the user. The tool is user-friendly and allows to perform a customized analysis by providing the genomic coordinates of the analyzed variants. It should be emphasized that `TADeus2` and its previous version `TADeus` (39) was successfully used in the recently published studies (40,41).

MATERIALS AND METHODS

Publicly available experimental data

Currently, `TADeus2` allows the user to upload genomic data in BED and BEDGraph format. The user is also provided with over 30 preloaded, publicly available datasets described in detail at https://tadeus2.mimuw.edu.pl/datasources/.

Evaluation of the gene pathogenicity

To assess and rank the pathogenicity of a gene `TADeus2` *score* is introduced. For a given gene g its value is calculated based on the following indicators: (i) CG(g): the ClinGen haploinsufficiency/triplosensitivity score (42), (ii) EPdis(g): the number of distant candidate enhancer–promoter predicted interactions disrupted by

Table 1. Listing of all available tools that provide similar functionalities to TADeus2 regarding typical clinical use cases, TAD analysis and breakpoint viewing

	Options					Breakpoint browser options			
Name of service	HiC Heat map	Tad annotation	Virtual 4C view	Data upload or local installation	Breakpoint browser mode	Wildtype View	Customised Tracks	Unlimited Number HiC tracks	Head/Tail Orientation of coordinates
TADeus[b]	✓	✓	✓	✓	✓	✓	✓	✓	✓
3D Genome Browser[a] (31)	✓	✓	✓	✓	✓		✓[b]		
3DIV[c] (33)	✓	✓	✓	✓	✓				
HiGlass (34)	✓	✓	✓	✓					
Juicebox (35)	✓		✓	✓					
WashU Epigenome Browser (36)	✓			✓					
HUGIN (37)	✓	✓	✓						
3Disease Browser (38)	✓	✓							

[a]Inter-chromosomal interaction mode is regarded as breakpoint mode for comparisons.
[b]As UCSC or WashU session link.
[c]Complex SV and 3D genome view is regarded as breakpoint mode for comparisons.

the breakpoints based on (43), (iii) HPO(g): the number of names and links to the associated phenotype described in Human Phenotype Ontology (HPO)—the ontology of phenotypic abnormalities with associated diseases and genes (44), (iv) dist(g): the distance from the rearrangement breakpoints. The final formula is accordingly a scaled sum of the four compounds:

$$score(g) = \mathbb{I}(\mathrm{CG(g)} = 1) + \mathbb{I}(\mathrm{EPdis(g)} > 0) + \mathbb{I}(\mathrm{HPO(g)} > 0) + \mathbb{I}(\mathrm{dist(g)} < 1\mathrm{Mb}),$$

where CG(g) is equal to 1 when g falls into one of the ClinGen categories: Sufficient Evidence, Emerging Evidence, Autosomal Recessive; and $\mathbb{I}(c) = 1$, when a condition c is met and 0 otherwise.

As an example consider gene g = TBR1 ranked as first in evaluation of structural variant involving chromosome 2 (exemplary case study from the main page of the web service). This gene is annotated in ClinGen, so CG(g) = 1, there are 38 disrupted distant enhancer-promoter interactions, i.e. EPdis(g) > 0, also the HPO(g) > 0 as several abnormal phenotypes have been found in HPO, and finally the distance of the TBR1 gene from the breakpoint is small enough, i.e. dist(g) < 1Mb. In total we get a Total Pathogenecity Score equal to 4.

TADeus2 rank gene table

The table is ordered by the aforementioned ranking score in the descending order. Genes with the equal scores are secondary sorted by the number of the disrupted enhancer–promoter interactions and the distance from the rearrangement breakpoints. For the convenience of the user the final importance ranking is color-coded with a dark-pink to white scale. Additionally, for a broader perspective, the disease and inheritance type from Online Mendelian Inheritance in Man (OMIM) (45) and pLI score from Genome Aggregation Database (gnomAD) (46), were added.

SVs/CNVs pathogenicity assessment

To evaluate the pathogenicity of a rearrangement TADeus2 uses the state-of-the-art, third-party software: TADA (available only for autosomes) (47) and ClassifyCNV (48) as well as an original statistical significance p-value.

TADA automatically ranks Copy Number Variants (CNVs) based on a extensive catalogue of functional annotations supported by enrichment analysis. The software is based on a machine-learning classifier to accurately predict and prioritize pathogenic deletion or duplication to produce a well-calibrated pathogenicity score.

ClassifyCNV, uses pre-parsed publicly available databases to calculate a pathogenicity score for each duplication and deletion. Importantly, the tool is an implementation of the 2019 ACMG guidelines for variant interpretation that provide a set of criteria to score variants and place them into one of the five classification tiers (49).

Finally, we propose calculation of an empirical p-value to assess the statistical significance of the predicted number of interrupted interactions between the *cis*-regulatory elements, which is a good predictor of SVs pathogenicity (50).

To this end we compute the null hypothesis probability distribution—the probability distribution for all possible numbers of disrupted interactions induced by a random breakpoint. We sample 10^4 loci and compute the number of potential disruptions that would occur because of a breakpoint in that loci. The resulting distribution is approximated by a mixture of the 0-concentrated Dirac distribution and the geometric distribution:

$$\mathbb{P}(X = k|P, \delta) = \begin{cases} \delta, & \text{for } k = 0 \\ (1-\delta)P(1-P)^{k-1}, & \text{for } k > 0 \\ 0, & \text{for } k < 0 \end{cases}$$

for some $P \in (0, 1)$ and $\delta \in (0, 1)$ parameters, which we estimated from the data. The proportion of breakpoints in all genome loci that do not disrupt any predicted promoter–enhancer interactions is an estimator of

δ (0.0924), while P (0.01174) was estimated as the maximum likelihood estimation (MLE). As a result, the smallest number of enhancer–promoter interactions broken by a rearrangement breakpoint that is statistically significant (i.e. P-value ≤ 0.05) is equal to 246.

Implementation and architecture of TADeus2 web server

TADeus2 is implemented as a Django application using MySQL database as data storage. Fragments of code from HiCExplorer (51) are reused in the track plot module. All the presented command-line tools (e.g. TADA, classifyCNV) incorporated in TADeus2 are accompanied with a responsive and user-friendly graphical interface developed with Bootstrap v4. Additionally, java-script snippets enable customization of forms and export of any user-defined plots that are produced using the python matplotlib package. Finally, TADeus2 provides the rest-API that can be utilized by external applications.

RESULTS AND DISCUSSION

Diagnosis workflow of TADeus2 web server

The input of TADeus2 consists of breakpoint coordinates (hg38) previously identified experimentally. The analysis starts with the selection of an adequate Hi-C matrix (currently, data from 8 different cell-types are available). The Hi-C matrix is displayed as *a triangle plot* (upper triangle rotated by 45°) and TADs structure is visualized (red lines), in order to investigate the SV role in reorganization of the domains (often linked to symptomatic gene misexpression). Based on the type of a variant the user should choose either *syntenic* mode (deletion or duplication), or *breakpoint* mode (translocation, inversion). The former, presents a continuous sequence of a genomic region and is designed to generate a ranked table of putative disease-causing genes. The latter integrates two genomic fragments from different loci (e.g. located at distinct chromosomes) as if it was a continuous genome chunk, allowing visualization of the new rearrangement. To use this mode the user should specify the coordinates of the breakpoints and the direction of the fused strands (e.g. in the case of translocation between P and q arms, a forward reverse strand fusion occurs). The breakpoint view also includes wild-type regions that compose the rearrangement. Further analysis (in both modes) includes an integration of additional tracks containing publicly available or self-uploaded datasets. After the selection of disease causing putative genes (based on the TADeus2rank gene table or the interpretation of particular tracks e.g. gnomAD pLI) the user should focus on the characterization of the cis-regulatory landscape within the region of interest. This can be achieved by: (i) identification of candidate enhancers (CEs) based on the histone marks (H3K27ac, H3K4me1), chromatin accessibility (DNAseI), and conservation; (ii) investigation of the track with experimentally validated enhancers; (iii) analysis of putative enhancer-promoter interactions based on Virtual 4C. Finally, verification of the regulatory elements, enhancer-promoter interactions, disrupted genes should be further confirmed experimentally. For the convenience of the reader, the overview of the clinical diagnosis workflow is also presented in Figure 1.

Validation of TADeus2 gene ranking scheme

In order to test TADeus2 accuracy a detailed search of literature was performed to find well-described cases of position effects generated by SVs and CNVs. For each case, breakpoints, clinical presentation and genes that according to authors contribute to the disease were collected. A table ranking disease-causing putative genes close to the SV/CNV was generated only for the chromosome where the disease causing gene was present. In total, 21 distinct cases were found and used to validate the TADeus2 ranking method (Supplementary Table S1). For 18 (85.7%) cases, the genes contributing to the disease were predicted as the strongest candidates (top scores if compared to the rest of the genes localized on the same chromosome). In the remaining 3 (14.3%) cases, genes were predicted as the most probable candidates with lower scores (see Supplementary Table S1).

Below we present in detail two selected cases from the set described above.

Examples of analysis workflows provided by TADeus2 in a clinical diagnosis setting

Case 1—Position effect in a balanced translocation neighbouring FOXG1 *gene.* Murcia Pienkowski *et al.* identified the exact structure of a balanced chromosomal translocation (46,XX,t(6;14)(p25.1;q12)) in a patient with epileptic seizures and severe developmental delay (40). The breakpoint on chromosome 6 disrupted the*TXNDC5* gene in the second intron (chr6:7903201 hg38), while chromosome 14 was damaged in non-coding DNA (chr14:29266318). TADeus2 ranking was used to establish genes that have a higher probability of being responsible for the disease. On chromosome 6 the highest score was assigned to gene DSP. At the same time on chromosome 14 *FOXG1* was chosen as the gene with the highest probability to impact the phenotype (Supplementary Table S1).

Based on the clinical picture that the patient displayed we decided to conduct the next steps of the analysis using data Hi-C from neural progenitor cells (NPC). The following tracks have been added to the breakpoint view: Genes coloured by pLI score, Virtual 4C for promoter regions (±2.5 kb from transcription start site (TSS)) for *DSP* and *FOXG1* and experimentally validated functional elements from NCBI database (Figure 2). The result indicate that *FOXG1* lost contact with 4 experimentally verified enhancers, while *DSP* did not lose any, making *FOXG1* the most probable cause of the disease (see the red box in Figure 2). Indeed, at least 11 translocations located in proximity of *FOXG1* in patients with Rett-like syndrome have been described. Furthermore, it has been proposed that loss of only one active *FOXG1* enhancer is not clinically relevant (52). This is consistent with our results as the analyzed translocation leads to displacement of two apparently important enhancers: hs433 and hs342; and two enhancers with unknown impact on textitFOXG1: hs1168 and hs598. hs433 is thought to bring other enhancers into physical contact with *FOXG1* (53), while hs342 is the main candidate for the regulation of *FOXG1* expression (52). Overall,

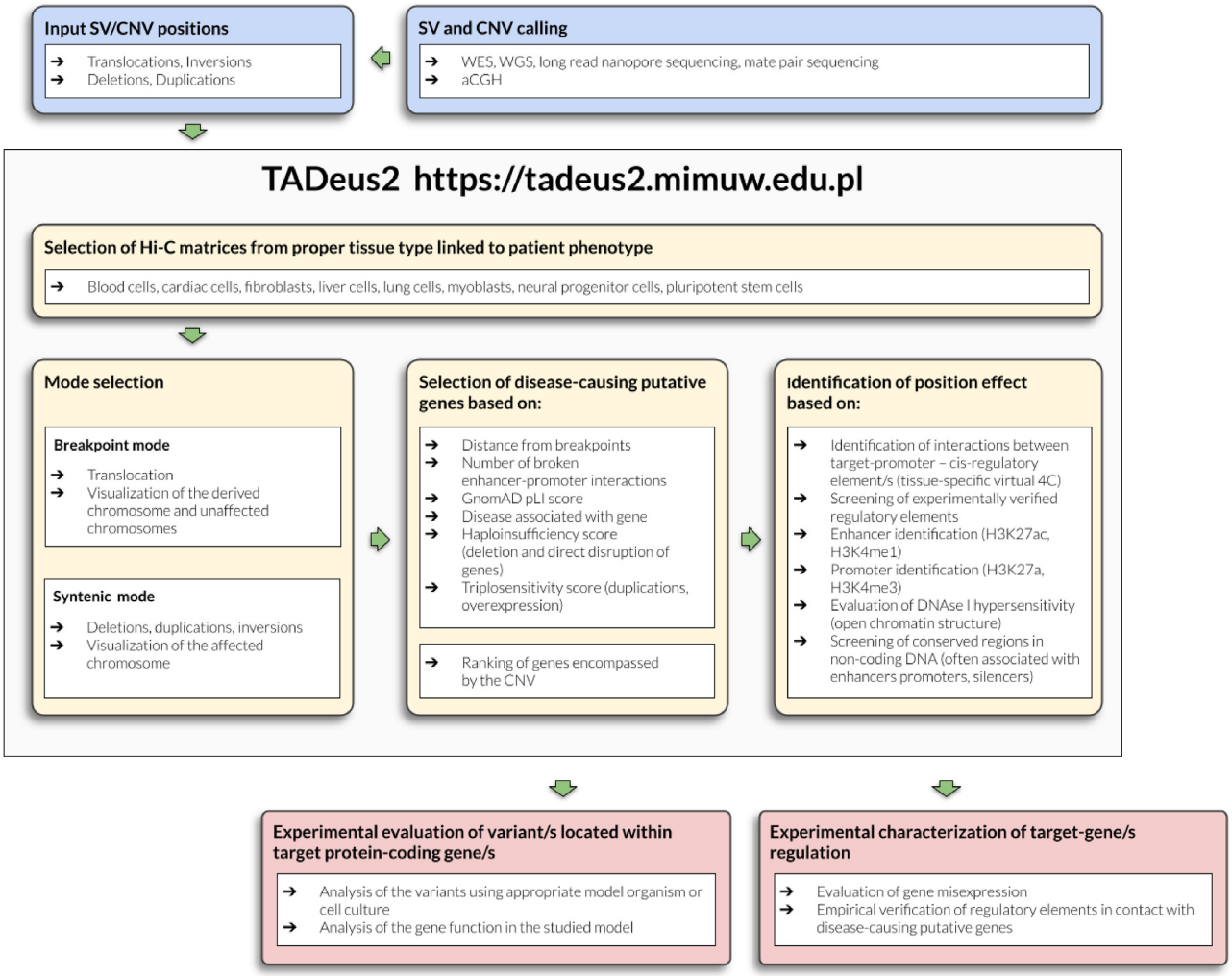


Figure 1. TADeus2 overview of the clinical diagnosis workflow. The above diagram outlines the subsequent steps recommended for the SVs/CNVs evaluation. The pipeline starts with the specification of a pair of coordinates accompanied by a Hi-C matrix of interest. Using two different views the structural rearrangements can be visualized and inspected. Next, based on numerous indicators of disease-causing putative genes candidates for further analysis can be selected. Finally, position effects can be evaluated thanks to provided databases from UCSC e.g. H3K27ac or experimentally verified regulatory elements. All the obtained outcomes are candidates for any further confirmation with wet experiments.

TADeus2 indicated correctly the gene responsible for the disease as well as important regulatory elements that might have been at the root of the patient phenotype. Importantly, further experimental investigation of *cis*-regulatory landscape encompassing FOXG1 might shed more light into the etiology of the patients phenotype.

Case 2—Position effect in an inversion accompanied with a deletion located nearby DLX5 and DLX6 genes. Kerry Brown et al. identified a paracentric inversion of the long arm of chromosome 7, inv(7)(q21.3q35) in five patients with hearing loss and craniofacial defects (54). The breakpoint in q21.3 (accompanied by a 5.1 kb deletion (chr7:96,935,329-96,940,443; hg38)) is located within non-coding region, 65–80 kb away from DLX5 and DLX6 genes, while the 7q35 breakpoint disrupts the *CNTNAP2* gene, between exon two and ten.

First, we usedTADeus2 to analyze the potential role of 7q35 breakpoint in the molecular etiology of hearing loss and craniofacial defects. Characterization of the chromatin architecture surrounding the breakpoint region revealed no genes within the TAD except *CNTNAP2* (Figure 3A). However, the pLI score for *CNTNAP2* was equal to zero, suggesting that disruption of one copy of *CNTNAP2* has a low probability of being responsible for the abovementioned defects. Indeed, the spatiotemporal expression analysis of *CNTNAP2* during mouse embryonic development revealed no expression in the majority of tissues linked to the patient's phenotype (54). Moreover, the *CNTNAP2*-/- mice present no abnormalities in craniofacial and inner ear development (55). Therefore, for further analysis we focused on the 7q21.3 breakpoint region.

The TADeus2 ranking list pointed DLX5 and DLX6 as genes with a high probability of being responsible for

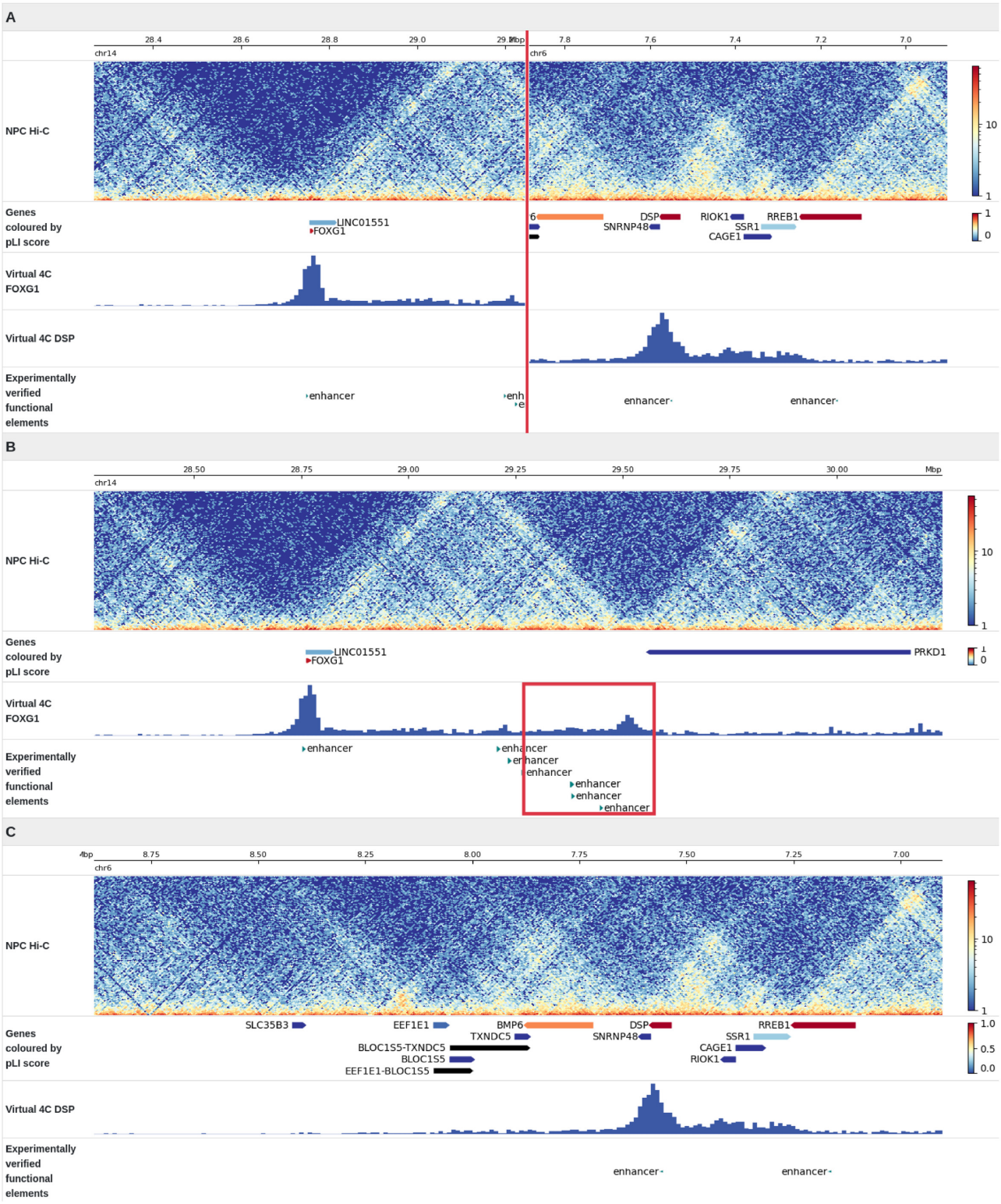


Figure 2. Analysis of a balanced translocation 46,XX,t(6;14)(p25.1;q12) in a patient with epileptic seizures and severe developmental delay. Visualization of the fusion chromosome der(14) (**A**) and wild type chromosomes: chromosome 14 (**B**) chromosome 6 (**C**). NPC Hi-C data was used to generate the chromatin architecture (top track). The figure contains the following tracks: gene gnomAD pLI score (bottom track) in color-scale (0 = blue and 1 = red), virtual 4C for promoter regions (±2.5 kb from TSS) for FOXG1 and DSP and experimentally validated functional elements. The enhancers displaced by the translocation are marked with the red rectangle.

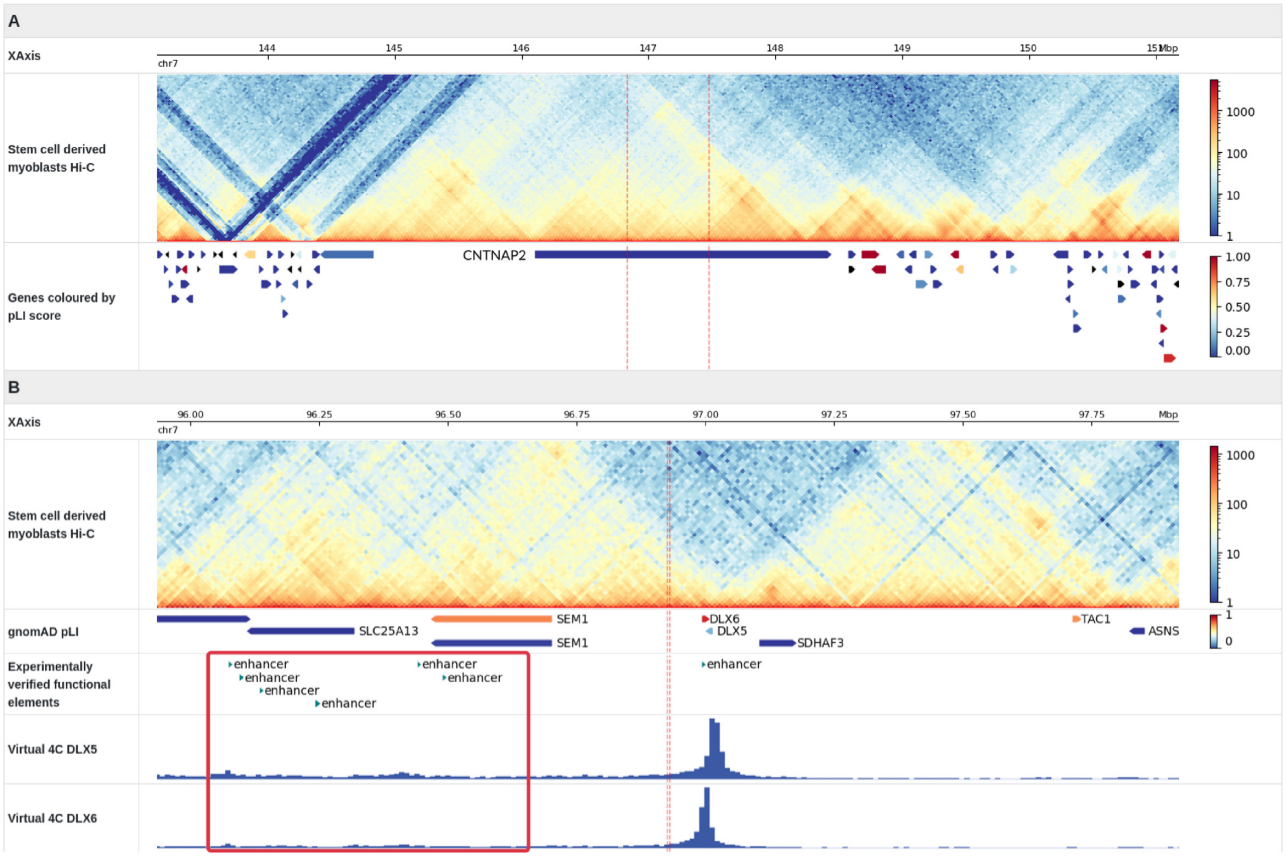


Figure 3. Analysis of the inversion inv(7)(q21.3q35) linked to the craniofacial defects and hearing loss. (**A**) Visualisation of the 7q35 breakpoint located between *CNTNAP2* exon two and ten. Stem cell derived myoblasts were used to visualise the chromatin architecture (top track) combined with gene gnomAD pLI score (bottom track) in color-scale (0 = blue and 1 = red). (**B**) Visualisation of the breakpoint in q21.3 located centromeric to DLX5 and DLX6. The Hi-C data was used from stem cell derived myoblasts (top track), followed by analysis of gene gnomAD pLI scores, and also cis-regulatory elements (middle tracks). The putative interactions between DLX5, DLX6 promoters (±2.5 kb from TSS) and non-coding DNA were investigated using Virtual 4C (bottom track). The enhancers affected by inversion are marked with the red rectangle.

these developmental abnormalities (top two genes with rank score: 300), with the pLI score equal to 0.2184 and 0.9213, respectively (Supplementary Table S1). Based on these results we extended our analysis by implementing the TADeus2 browser tracks with: experimentally validated functional elements from NCBI database and virtual 4C for *DLX5* and *DLX6* promoter regions (±2.5 kb from TSS). Multiple enhancers are located within the genomic region affected by the inversion, and in close 3D proximity with *DLX5* and *DLX6* promoters (based on virtual 4C).

This suggests that the inversion might disrupt long-range contacts between six experimentally validated enhancers (in the red rectangle) and promoters of *DLX5* and *DLX6*, most likely leading to a misregulation of these two genes (Figure 3B).

Importantly, *DLX5* is expressed in the otic placode and vesicle (involved in formation of inner ear vestibular structures), and in the semicircular canals of the inner ear (56). Deletion of *DLX5* and *DLX6* leads to dysplastic ears and congenital deafness (Chromosome 7 Annotation Project, http://www.chr7.org), which is consistent with the mouse phenotype. The *DLX5* knockout as well as *DLX5* and *DLX6* double knockout mice are characterized by the craniofacial malformations and abnormalities in the ear development (56–58). Collectively, TADeus2 results align with literature results suggesting that changes in the *DLX5* and *DLX6* gene dosage caused by disruption of enhancer-promoter connections may lead to the patient phenotype.

CONCLUSIONS AND FUTURE PLANS

In the following article we presented TADeus2 web server with all its features. The tool is an online solution for visualization and analysis of SVs and CNVs in the context of chromatin conformation. Thanks to an easy to use and user-friendly interface, TADeus2 is suitable as a handful framework for preliminary clinical diagnosis of patients for non-bioinformatician experts.

The future directions for the improvement of the web server include additional expansions of the breakpoint view that would allow the view of more breakpoint coordinates at once (particularly useful for visualisation of inversions, duplications, and complex rearrangements, i.e. chromothripsis). In addition, we plan to implement a functionality that provides a visualization of chromatin conformation in 3D.

Moreover, to predict the TAD structure affected by the SVs, Akita software (59) will be used to generate a new Hi-

C matrix for the rearranged TAD. This functionality is currently under development, with a beta version available for the user.

Furthermore, the statistical modeling of chromatin organization disorders caused by structural variants will be further developed to achieve higher accuracy in its assessment. Lastly, we state that TADeus2 will benefit from regular updates using the wealth of data publicly available.

DATA AVAILABILITY

TADeus2 is publicly available at https://tadeus2.mimuw.edu.pl. Detailed tutorial, help pages, short videos presenting functionality of application, and a FAQ section are provided. The tool is free and open to all users and there is no login requirement. TADeus2 source code and installation instructions are available at https://github.com/bposzewiecka/tadeus2. The software is distributed under a GNU General Public License v3.0.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Authors want to thank Paweł Stankiewicz (Baylor College of Medicine) and Tomasz Gambin (Warsaw University of Technology) for their invaluable contribution to the development of the first version of the TADeus tool.

FUNDING

B.P. was supported by the grant of National Science Centre [NCN 2019/33/N/ST6/03110]; K.G. and A.G. were supported by the grant of National Science Centre [NCN 2018/30/M/NZ2/00054]; V.M.P. is supported by the Foundation for Polish Science, START Programme and MarMaRa grant. Funding for open access charge: Polish National Science Centre [NCN 2018/30/M/NZ2/00054].
Conflict of interest statement. None declared.

REFERENCES

1. Zhang,J., Chiodini,R., Badr,A. and Zhang,G. (2011) The impact of next-generation sequencing on genomics. *J. Genet. Genomics.*, **38**, 95–109.
2. Hangauer,M.J., Vaughn,I.W. and McManus,M.T. (2013) Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.*, **9**, e1003569.
3. Sawyer,S., Hartley,T., Dyment,D., Beaulieu,C., Schwartzentruber,J., Smith,A., Bedford,H., Bernard,G., Bernier,F., Brais,B. *et al.* (2016) Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin. Genet.*, **89**, 275–284.
4. Kumar,A., Adhikari,S., Kankainen,M. and Heckman,C.A. (2021) Comparison of structural and short variants detected by linked-read and whole-exome sequencing in multiple Myeloma. *Cancers*, **13**, 1212.
5. Zare,F., Dow,M., Monteleone,N., Hosny,A. and Nabavi,S. (2017) An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC bioinformatics*, **18**, 286.
6. Harewood,L., Chaignat,E. and Reymond,A. (2012) Structural variation and its effect on expression. *Methods Mol. Biol.*, **838**, 173–186.
7. Mitelman,F., Johansson,B. and Mertens,F. (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer*, **7**, 233–245.
8. Mertens,F., Johansson,B., Fioretos,T. and Mitelman,F. (2015) The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer*, **15**, 371–381.
9. Eykelenboom,J.E., Briggs,G.J., Bradshaw,N.J., Soares,D.C., Ogawa,F., Christie,S., Malavasi,E.L., Makedonopoulou,P., Mackie,S., Malloy,M.P. *et al.* (2012) A t (1; 11) translocation linked to schizophrenia and affective disorders gives rise to aberrant chimeric DISC1 transcripts that encode structurally altered, deleterious mitochondrial proteins. *Hum. Mol. Genet.*, **21**, 3374–3386.
10. Lupski,J.R. (1999) Charcot-Marie-Tooth polyneuropathy: duplication, gene dosage, and genetic heterogeneity. *Pediatr. Res.*, **45**, 159–165.
11. Harewood,L., Schütz,F., Boyle,S., Perry,P., Delorenzi,M., Bickmore,W.A. and Reymond,A. (2010) The effect of translocation-induced nuclear reorganization on gene expression. *Genome Res.*, **20**, 554–564.
12. Harewood,L. and Fraser,P. (2014) The impact of chromosomal rearrangements on regulation of gene expression. *Hum. Mol. Genet.*, **23**, R76–R82.
13. Aigner,A., Müller,G., Knapp,E. and Raas,E. (1974) Systolic time intervals, phonocardiograms and sound spectrograms in patients with Starr-Edwards aortic valve prostheses. *Cardiology*, **59**, 30–40.
14. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
15. Chen,X., Ke,Y., Wu,K., Zhao,H., Sun,Y., Gao,L., Liu,Z., Zhang,J., Tao,W., Hou,Z. *et al.* (2019) Key role for CTCF in establishing chromatin structure in human embryos. *Nature*, **576**, 306–310.
16. Flavahan,W.A., Drier,Y., Johnstone,S.E., Hemming,M.L., Tarjan,D.R., Hegazi,E., Shareef,S.J., Javed,N.M., Raut,C.P., Eschle,B.K. *et al.* (2019) Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature*, **575**, 229–233.
17. Lupiáñez,D.G., Kraft,K., Heinrich,V., Krawitz,P., Brancati,F., Klopocki,E., Horn,D., Kayserili,H., Opitz,J.M., Laxova,R. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, **161**, 1012–1025.
18. Franke,M., Ibrahim,D.M., Andrey,G., Schwarzer,W., Heinrich,V., Schöpflin,R., Kraft,K., Kempfer,R., Jerković,I., Chan,W.-L. *et al.* (2016) Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, **538**, 265–269.
19. Lettice,L.A., Daniels,S., Sweeney,E., Venkataraman,S., Devenney,P.S., Gautier,P., Morrison,H., Fantes,J., Hill,R.E. and FitzPatrick,D.R. (2011) Enhancer-adoption as a mechanism of human developmental disease. *Hum. Mutat.*, **32**, 1492–1499.
20. Wells,A., Heckerman,D., Torkamani,A., Yin,L., Sebat,J., Ren,B., Telenti,A. and di Iulio,J. (2019) Ranking of non-coding pathogenic variants and putative essential regions of the human genome. *Nat. Commun.*, **10**, 5241.
21. Zhang,F. and Lupski,J.R. (2015) Non-coding genetic variants in human disease. *Hum. Mol. Genet.*, **24**, R102–R110.
22. Momozawa,Y. and Mizukami,K. (2020) Unique roles of rare variants in the genetics of complex diseases in humans. *J. Hum. Genet.*, **66**, 11–23.
23. Zepeda-Mendoza,C.J., Ibn-Salem,J., Kammin,T., Harris,D.J., Rita,D., Gripp,K.W., MacKenzie,J.J., Gropman,A., Graham,B., Shaheen,R. *et al.* (2017) Computational prediction of position effects of apparently balanced human chromosomal rearrangements. *Am. J. Hum. Genet.*, **101**, 206–217.
24. Landrum,M.J., Lee,J.M., Benson,M., Brown,G.R., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Jang,W. *et al.* (2017) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
25. Ibn-Salem,J., Köhler,S., Love,M.I., Chung,H.-R., Huang,N., Hurles,M.E., Haendel,M., Washington,N.L., Smedley,D., Mungall,C.J. *et al.* (2014) Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.*, **15**, 423.
26. Firth,H.V., Richards,S.M., Bevan,A.P., Clayton,S., Corpas,M., Rajan,D., Van Vooren,S., Moreau,Y., Pettett,R.M. and Carter,N.P.

(2009) DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.*, **84**, 524–533.
27. Li,R., Liu,Y., Li,T. and Li,C. (2016) 3Disease Browser: a web server for integrating 3D genome and disease-associated chromosome rearrangement data. *Sci. Rep.*, **6**, 34651.
28. Köhler,S., Schoeneberg,U., Czeschik,J.C., Doelken,S.C., Hehir-Kwa,J.Y., Ibn-Salem,J., Mungall,C.J., Smedley,D., Haendel,M.A. and Robinson,P.N. (2014) Clinical interpretation of CNVs with cross-species phenotype data. *J. Med. Genet.*, **51**, 766–772.
29. Hehir-Kwa,J.Y., Wieskamp,N., Webber,C., Pfundt,R., Brunner,H.G., Gilissen,C., de Vries,B.B., Ponting,C.P. and Veltman,J.A. (2010) Accurate distinction of pathogenic from benign CNVs in mental retardation. *PLoS Comput. Biol.*, **6**, e1000752.
30. Huang,N., Lee,I., Marcotte,E.M. and Hurles,M.E. (2010) Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.*, **6**, e1001154.
31. Wang,Y., Song,F., Zhang,B., Zhang,L., Xu,J., Kuang,D., Li,D., Choudhary,M.N.K., Li,Y., Hu,M. *et al.* (2018) The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.*, **19**, 151.
32. Requena,F., Abdallah,H.H., García,A., Nitschké,P., Romana,S., Malan,V. and Rausell,A. (2021) CNVxplorer: a web tool to assist clinical interpretation of CNVs in rare disease patients. *Nucleic Acids Res.*, **49**, W93–W103.
33. Yang,D., Jang,I., Choi,J., Kim,M.-S., Lee,A.J., Kim,H., Eom,J., Kim,D., Jung,I. and Lee,B. (2018) 3DIV: A 3D-genome Interaction Viewer and database. *Nucleic Acids Res.*, **46**, D52–D57.
34. Kerpedjiev,P., Abdennur,N., Lekschas,F., McCallum,C., Dinkla,K., Strobelt,H., Luber,J.M., Ouellette,S.B., Azhir,A., Kumar,N. *et al.* (2018) HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.*, **19**, 125.
35. Durand,N.C., Robinson,J.T., Shamim,M.S., Machol,I., Mesirov,J.P., Lander,E.S. and Aiden,E.L. (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.*, **3**, 99–101.
36. Li,D., Hsu,S., Purushotham,D., Sears,R.L. and Wang,T. (2019) WashU epigenome browser update 2019. *Nucleic Acids Res.*, **47**, W158–W165.
37. Martin,J.S., Xu,Z., Reiner,A.P., Mohlke,K.L., Sullivan,P., Ren,B., Hu,M. and Li,Y. (2017) HUGIn: Hi-C unifying genomic interrogator. *Bioinformatics*, **33**, 3793–3795.
38. Li,R., Liu,Y., Li,T. and Li,C. (2016) 3Disease Browser: a web server for integrating 3D genome and disease-associated chromosome rearrangement data. *Sci. Rep.*, **6**, 34651.
39. Poszewiecka,B., Stankiewicz,P., Gambin,T. and Gambin,A. (2018) TADeus-a tool for clinical interpretation of structural variants modifying chromatin organization. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, pp. 84–87.
40. Pienkowski,V.M., Kucharczyk,M., Młynek,M., Szczałuba,K., Rydzanicz,M., Poszewiecka,B., Skórka,A., Sykulski,M., Biernacka,A., Koppolu,A.A. *et al.* (2019) Mapping of breakpoints in balanced chromosomal translocations by shallow whole-genome sequencing points to EFNA5, BAHD1 and PPP2R5E as novel candidates for genes causing human Mendelian disorders. *J. Med. Genet.*, **56**, 104–112.
41. Pienkowski,V.M., Kucharczyk,M., Rydzanicz,M., Poszewiecka,B., Pachota,K., Młynek,M., Stawiński,P., Pollak,A., Kosińska,J., Wojciechowska *et al.* (2020) Breakpoint mapping of symptomatic balanced translocations links the EPHA6, KLF13 and UBR3 genes to novel disease phenotype. *J. Clin. Med.*, **9**, 1245.
42. Rehm,H.L., Berg,J.S., Brooks,L.D., Bustamante,C.D., Evans,J.P., Landrum,M.J., Ledbetter,D.H., Maglott,D.R., Martin,C.L., Nussbaum,R.L. *et al.* (2015) ClinGen—the clinical genome resource. *N. Engl. J. Med.*, **372**, 2235–2242.
43. Thurman,R.E., Rynes,E., Humbert,R., Vierstra,J., Maurano,M.T., Haugen,E., Sheffield,N.C., Stergachis,A.B., Wang,H., Vernot,B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75.
44. Köhler,S., Vasilevsky,N.A., Engelstad,M., Foster,E., McMurry,J., Aymé,S., Baynam,G., Bello,S.M., Boerkoel,C.F., Boycott,K.M. *et al.* (2016) The human phenotype ontology in 2017. *Nucleic Acids Res.*, **45**, D865–D876.
45. Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
46. Karczewski,K.J., Francioli,L.C., Tiao,G., Cummings,B.B., Alföldi,J., Wang,Q., Collins,R.L., Laricchia,K.M., Ganna,A., Birnbaum,D.P., Gauthier,L.D. *et al.* (2020) The mutational constraint spectrum quantified from variation in 141, 456 humans. *Nature*, **581**, 434–443.
47. Hertzberg,J., Mundlos,S., Vingron,M. and Gallone,G. (2022) TADA—a machine learning tool for functional annotation-based prioritisation of pathogenic CNVs. *Genome Biol.*, **23**, 67.
48. Gurbich,T.A. and Ilinsky,V.V. (2020) ClassifyCNV: a tool for clinical annotation of copy-number variants. *Sci. Rep.*, **10**, 203757.
49. Riggs,E.R., Andersen,E.F., Cherry,A.M., Kantarci,S., Kearney,H., Patel,A., Raca,G., Ritter,D.I., South,S.T., Thorland *et al.* (2020) 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.*, **22**, 245–257.
50. D'haene,E. and Vergult,S. (2021) Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet. Med.*, **23**, 34–46.
51. Ramírez,F., Bhardwaj,V., Arrigoni,L., Lam,K.C., Grüning,B.A., Villaveces,J., Habermann,B., Akhtar,A. and Manke,T. (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.*, **9**, 189.
52. Mehrjouy,M.M., Fonseca,A.C.S., Ehmke,N., Paskulin,G., Novelli,A., Benedicenti,F., Mencarelli,M.A., Renieri,A., Busa,T., Missirian,C. *et al.* (2018) Regulatory variants of FOXG1 in the context of its topological domain organisation. *Eur. J. Hum. Genet.*, **26**, 186–196.
53. Ibn-Salem,J., Köhler,S., Love,M.I., Chung,H.-R., Huang,N., Hurles,M.E., Haendel,M., Washington,N.L., Smedley,D., Mungall,C.J. *et al.* (2014) Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.*, **15**, 423.
54. Brown,K.K., Reiss,J.A., Crow,K., Ferguson,H.L., Kelly,C., Fritzsch,B. and Morton,C.C. (2010) Deletion of an enhancer near DLX5 and DLX6 in a family with hearing loss, craniofacial defects, and an inv (7)(q21. 3q35). *Hum. Genet.*, **127**, 19.
55. Poliak,S., Salomon,D., Elhanany,H., Sabanay,H., Kiernan,B., Pevny,L., Stewart,C.L., Xu,X., Chiu,S.-Y., Shrager,P. *et al.* (2003) Juxtaparanodal clustering of Shaker-like K+ channels in myelinated axons depends on Caspr2 and TAG-1. *J. Cell Biol.*, **162**, 1149–1160.
56. Merlo,G.R., Paleari,L., Mantero,S., Zerega,B., Adamska,M., Rinkwitz,S., Bober,E. and Levi,G. (2002) The Dlx5 homeobox gene is essential for vestibular morphogenesis in the mouse embryo through a BMP4-mediated pathway. *Dev. Biol.*, **248**, 157–169.
57. Acampora,D., Merlo,G.R., Paleari,L., Zerega,B., Postiglione,M.P., Mantero,S., Bober,E., Barbieri,O., Simeone,A. and Levi,G. (1999) Craniofacial, vestibular and bone defects in mice lacking the distal-less-related gene Dlx5. *Development*, **126**, 3795–3809.
58. Merlo,G.R., Paleari,L., Mantero,S., Genova,F., Beverdam,A., Palmisano,G.L., Barbieri,O. and Levi,G. (2002) Mouse model of split hand/foot malformation type I. *Genesis*, **33**, 97–101.
59. Fudenberg,G., Kelley,D.R. and Pollard,K.S. (2020) Predicting 3D genome folding from DNA sequence with Akita. *Nat. Methods*, **17**, 1111–1117.