ARTICLE

ClinTAD: a tool for copy number variant interpretation in the context of topologically associated domains

Jacob D. Spector[1] · Arun P. Wiita [1]

Abstract

Standard clinical interpretation of DNA copy number variants (CNVs) identified by cytogenomic microarray involves examining protein-coding genes within the region and comparison to other CNVs. Emerging basic research suggests that CNVs can also exert a pathogenic effect through disruption of DNA structural elements such as topologically associated domains (TADs). To begin to integrate these discoveries with current practice, we developed ClinTAD, a free browser-based tool to assist with interpretation of CNVs in the context of TADs (www.clintad.com). We used ClinTAD to examine 209 randomly selected single-nucleotide polymorphism microarray cases with a total of 236 CNVs. We compared 118 CNVs classified as variants of uncertain clinical significance (VUS), where additional insight into pathogenicity of these CNVs would be of greatest utility, to 118 CNVs classified as benign. We found that a higher proportion of VUS had at least two genes in a nearby TAD related to a phenotype seen in the patient based on Human Phenotype Ontology (HPO) annotation. We present example cases demonstrating scenarios where ClinTAD may either increase or decrease clinical suspicion of pathogenicity for VUS, depending on disruption of TAD boundaries and HPO phenotype match. ClinTAD is an easy-to-use tool, based on emerging research in chromatin architecture, that can help inform CNV interpretation.

Introduction

DNA copy number variants (CNVs) have been associated with a wide variety of diseases in humans, including autism, intellectual disability, and congenital malformations. For this reason, American College of Medical Genetics and Genomics guidelines recommend cytogenomic microarrays for CNVs as "first-tier" tests for the postnatal evaluation of individuals with these conditions [1]. Currently, clinical interpretation of CNV pathogenicity takes into consideration properties such as the protein-coding genes included within the region and CNV size. Another factor in clinical interpretation of CNVs is comparison to similar CNVs from other individuals. Attempts can be made to establish a genotype–phenotype correlation with affected individuals with similar CNV breakpoints, such as those in the DECIPHER [2] and ClinGen [3] databases. Analogously, CNVs are more likely to be classified as benign if they appear frequently in the Database of Genomic Variants [4] (DGV) describing the normal population.

However, there are pitfalls in this current view of CNV interpretation, especially for CNVs with no similar cases in the above databases. While CNVs have long been known to cause disease by directly affecting the gene coding sequence, recent studies have demonstrated that CNVs can also lead to disease when they disrupt normal chromatin architecture [5, 6]. One of the most important mediators of this phenomenon appears to be topologically associated domains (TADs), which are neighborhoods of DNA interaction that demarcate and limit physical interactions between genomic regions. These recently discovered chromatin features play both structural and functional roles, regulating gene-enhancer interactions and influencing epigenetic modifications such as histone methylation [6, 7]. CNVs disrupting TAD boundaries may be most likely to cause clinical effect, though CNVs internal to TADs may also lead to alterations of transcriptional regulation [5, 6, 8]. Ibn-Salem et al. [9] suggested that as many as 11.8% of

✉ Arun P. Wiita
 Arun.Wiita@ucsf.edu

[1] Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA, USA

pathogenic CNVs may exert their effect through disruption of TADs. This important finding suggests that in current routine practice we may be missing clinically significant CNVs by ignoring their effect on TADs.

However, a limitation of Ibn-Salem et al. [9] above was that they primarily examined very large (typically >3 Mb) genomic deletions, which are seen infrequently in healthy individuals and already typically interpreted as pathogenic [1, 10]. In clinical practice, a major issue is deciding on the potential pathogenicity of much smaller CNVs that would otherwise be classified as variants of uncertain clinical significance (VUS). Therefore, rooted in this emerging area of genome research, we developed ClinTAD, a browser-based tool that can assist in determining the clinical significance of a CNV in the context of TADs. ClinTAD allows a user to input a chromosome number, genomic coordinates, and phenotypic information as a Human Phenotype Ontology (HPO) [11] ID number for either a single case or multiple cases. The main functionality of the tool is to relate this entered data to nearby TAD boundaries and genes.

Here, we first use ClinTAD to evaluate 236 CNVs extracted from clinical cases at our institution with an equal distribution between those classified as VUS and those classified as benign. We find that CNVs classified as VUS are more likely to demonstrate HPO matches to genes in adjacent TADs than those classified as benign. We then describe four specific use cases of the software, focusing on examples classified as VUS and without similar cases in typical comparison databases. We show how ClinTAD analysis to identify TAD boundary disruption and HPO matches to genes outside the CNV boundaries can inform suspicion of clinical pathogenicity for VUS. We anticipate that this decision support tool, freely available at www.clintad.com, will provide a useful resource to begin to integrate these basic discoveries with clinical practice.

Materials and methods

Clinical case review

This study protocol was approved by the University of California, San Francisco (UCSF) Committee on Human Research. For each case, we manually reviewed the electronic medical record to obtain the clinical indication for microarray testing and all associated phenotypes. Based on medical record review, we searched for HPOs using ClinTAD's "Lookup HPO" function, and attempted to assign all HPOs that described the patients' phenotypes to each of their variants. More generic HPOs that described the patients' phenotypes were also assigned, but more specific HPOs were not assigned if we were not certain they

matched a patient's phenotype. For example, if a clinical note described a patient as having a "median cleft palate" the patient would be assigned the HPO "median cleft palate" and more generic HPOs such as "cleft palate." However, the case would not be assigned more specific HPOs such as "cleft hard palate" or "submucous cleft of soft and hard palate" because it is not clear these describe the patient's exact phenotype. Clinical classification of variants as VUS and benign were as per the ACMG (American College of Medical Genetics and Genomics) guidelines [1].

Code and data sources

The tool was created using Django as the web framework, with Javascript and D3.js for visualization of results. The code for the tool is available at github.com/Jacob-SpectorMD/ClinTAD, and can be downloaded, modified, and run locally if desired. An online version of the tool is currently available at www.clintad.com.

The TAD boundaries file used [12] was generated using H1 human embryonic stem cells (hESCs), chromosome build GRCh37, a bin size of 40 kb, and a window size of 2 Mb [7]. Total chromosome lengths are taken from the Genome Reference Consortium [13] using build GRCh37 (release date 2009–02–27). The gene–phenotype associations are based on the genes to phenotype file for all sources/all frequencies from the HPO website [14, 15]. The list of HPO terms and their descriptions is based on the HPO. obo file, also downloaded from the HPO website. Genes and their coordinates used in the tool are based on the human gene set for GRCh37 (Homo_sapiens.GRCh37.87. gtf) downloaded from the Ensembl archive website [16, 17]. Only elements with a feature type of "gene" and gene biotype of "protein_coding" are used. Enhancer data were taken from the VISTA Enhancer Browser [18], and include only human enhancers that were positive for in vivo enhancer activity as defined by VISTA. The data in the DGV track were generated using the DGV Gold Standard Variants file with the release date of 2016–05–15, and were downloaded from DGV [4].

ClinTAD features and functionality

When using the single case view, ClinTAD returns a visualization showing the CNV, TAD boundaries, VISTA enhancers, and genes, with genes that have phenotypic matches highlighted in orange. If a gene is associated with one or more of the patient's HPOs, it is considered a "gene match" and each individual matching HPO is considered an "HPO match." The number of "unique HPO matches" is defined as the number of HPO matches when only counting each individual HPO ID a maximum of once. The default

functionality of the single case view is to search from the left (lower number) CNV coordinate for the nearest left TAD boundary with a lesser coordinate, and from the right CNV coordinate for the nearest right TAD boundary with a greater coordinate. We also include a multiple case view that may be useful for research or cohort studies, which returns a text document showing the HPO matches for each case. Additionally, the tool has a statistics function that randomly places the user's CNV in 500 locations across the genome. A visualization is then generated showing the number of gene matches, the number of HPO matches, and a weighted match score for the actual CNV and the 500 random CNVs. The weighted match score takes into account the frequency that each HPO phenotype occurs, and can be used to compare the original CNV to the randomly generated CNVs. To generate an easy-to-read score for each CNV, the value of a single HPO match was arbitrarily defined as 20 divided by the number of genes associated with that HPO phenotype, with the total weighted HPO match score being equal the sum of all the individual HPO match values. For example, case A described below had HPO matches for patent ductus arteriosus, bicuspid aortic valve, coarctation of the aorta (two nearby genes had this HPO), and abnormal aortic arch morphology. The number of genes associated with these HPO terms is 193, 54, 59, and 12, respectively. The total weighted match score for this case would therefore be $20/193 + 20/54 + 2 \times (20/59) + 20/12 = 2.82$.

Results

Clinical case review

We reviewed 236 CNVs from 209 single-nucleotide polymorphism microarrays (Illumina CytoSNP 850 k platform), with 118 of the variants previously interpreted as benign and 118 as VUS. These cases represent a randomly selected subset of those performed for clinical diagnosis at UCSF between 2014 and 2018 with at least one CNV classified as benign or VUS. Of these variants, 132 were duplications (minimum 35 kb, median 328 kb, maximum 2665 kb) and 104 were deletions (minimum 23 kb, median 278 kb, maximum 9728 kb). Twenty-nine of the VUS overlapped at least one of the other VUS and 94 of the benign variants overlapped at least one of the other benign variants. All coordinates for these cases are in genome build GRCh37. The number of total HPOs assigned to the patient presentation in each clinical case, based on a manual review of medical records, was similar between those with benign variants and VUS (Fig. 1). Additional clinical information is provided in Supplementary Table 1 (VUS cases) and Supplementary Table 2 (benign cases).

After analyzing these cases using ClinTAD, a chi-square test was performed to compare gene matches and HPO matches (see Materials and methods for definition of terms) for benign variants and VUS. A higher proportion of VUS have two or more gene matches (Fig. 2a, $p = 0.006$), two or more HPO matches (Fig. 2b, $p = 0.006$), and two or more unique HPO matches (Fig. 2c, $p = 0.002$). We further compared benign variants and VUS that specifically cross TAD boundaries. For these variants, although not statistically significant due to smaller number of cases, VUS showed a trend toward a higher proportion of variants with two or more gene matches (Fig. 2d, $p = 0.131$), two or more HPO matches (Fig. 2e, $p = 0.297$), and two or more unique HPO matches (Fig. 2f, $p = 0.173$). These findings are perhaps unsurprising given standard reporting guidelines, whereby many VUS are CNVs that include genes within the region but lack further data to stratify as benign or pathogenic, or genes that may carry some suspicion of clinical relevance. However, these findings offer a baseline from which to assess the ClinTAD output of number of HPO-matching genes. In addition, these findings demonstrate the ability of ClinTAD HPO-matching function to suggest specific genes outside the immediate CNV region for consideration. Notably, these genes would be ignored by standard microarray reporting guidelines [1].

Example use cases in clinical CNV interpretation

With these baseline data in hand, in Fig. 3, we show example cases where ClinTAD analysis can potentially inform the probability of clinical pathogenicity for VUS. These cases illustrate examples of some of the most challenging scenarios in microarray interpretation, where any additional data may be beneficial to aid in interpretation. Case A is a patient with several cardiac abnormalities including aortic arch interruption or coarctation of the aorta, bicuspid aortic valve, patent ductus arteriosus, and patent foramen ovale. Figure 3a demonstrates the patient's ~1.3 Mb deletion in Chr15 with approximate coordinates of 95,407,056–96,696,462. Notably, this region includes no protein-coding genes and deletions with similar breakpoints were not identified in the ClinGen, DECIPHER, or DGV databases, leaving little space for further interpretation using standard reporting criteria. Using ClinTAD, however, we found that this deletion disrupted a TAD boundary, which is less common in benign CNVs based on our analysis above (Fig. 2), as well as the analysis of Ibn-Salem et al. [9], and therefore increases suspicion for clinical effect. Using ClinTAD HPO-matching function, in the TADs adjacent to the CNV, *MCTP2* had matches for abnormal aortic arch morphology (HP:0012303), bicuspid aortic valve (HP:0001647), patent ductus arteriosus (HP:0001643), and coarctation of the aorta (HP:0001680),

Fig. 1 Distribution of Human Phenotype Ontology (HPO)-annotated phenotypes assigned to the presentation of clinical cases (**a**) and summary statistics for these assignments (**b**). The mean and median number of HPOs assigned is similar between the two categories. The 118 benign variants have a total of 1066 HPOs assigned and the 118 variants of uncertain clinical significance (VUS) have a total of 1062 HPOs assigned
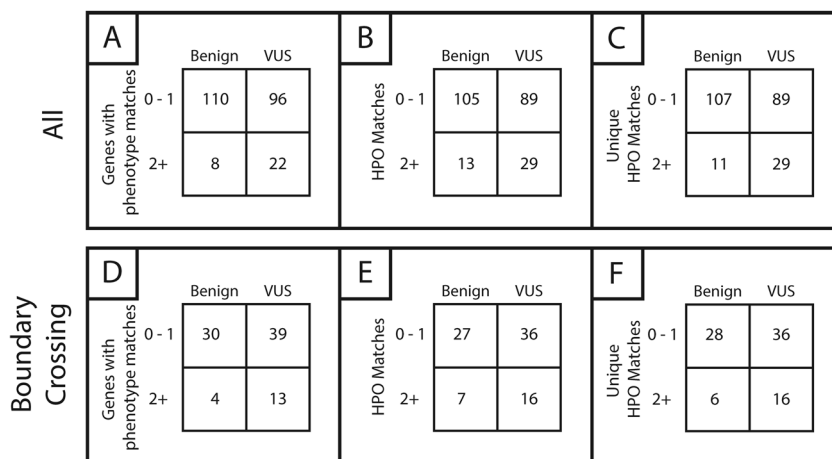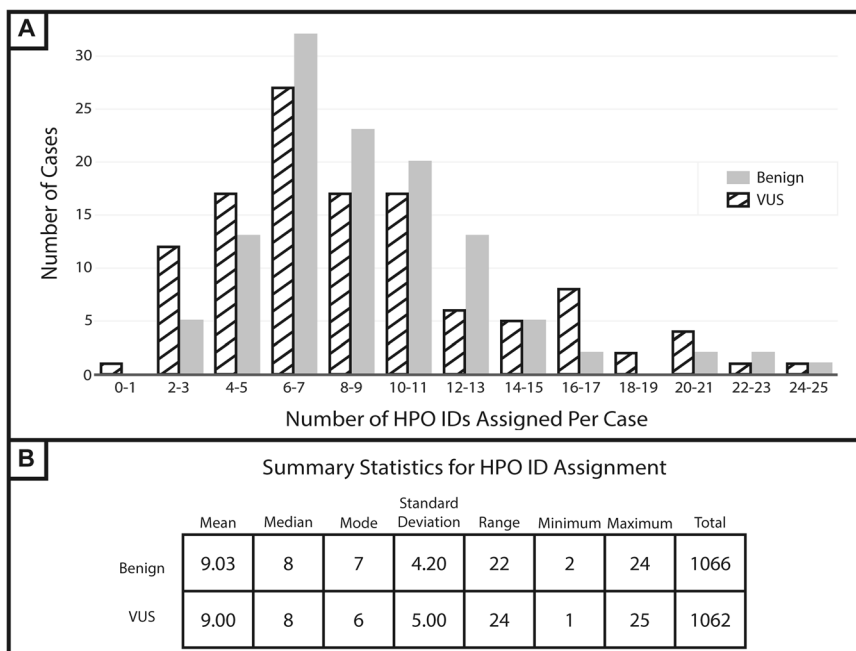

Fig. 2 Comparison of benign vs. variants of uncertain clinical significance (VUS) copy number variants (CNVs) by ClinTAD. The 118 benign variants and 118 VUS examined characterized by number of gene matches (**a**), Human Phenotype Ontology (HPO matches (**b**), and unique HPO matches (**c**). Of the variants we reviewed using ClinTAD, 34 benign variants and 52 VUS were predicted to cross topologically associated domain (TAD) boundaries. The number of gene matches (**d**), HPO matches (**e**), and unique HPO matches (**f**) for only boundary crossing variants are also shown

and *NR2F2* had a match for coarctation of the aorta. In particular, *NR2F2* (also known as COUP-TFII; OMIM*107773) is a critical transcription factor in cardiac development and single-copy loss of function of *NR2F2* is linked to cardiac abnormalities. While no VISTA-validated enhancer was present within the region, review of ENCODE data demonstrated the presence of a strong H3K27Ac peak within this region, suggestive of a possible enhancer element. ClinTAD therefore leads to the hypothesis that the enhancer present within the TAD may be disrupted by the CNV leading to misexpression of *NR2F2* and/

or *MCTP2*, increasing suspicion for pathogenicity beyond standard reporting criteria. Notably, the ClinTAD statistics function for case A demonstrated that this CNV had a higher weighted score than at 493 out of 500 random locations, further increasing suspicion for causality.

Case B is a patient with a clinical history including developmental delay, autism, aggressive behavior, headache/migraines, and alopecia. Figure 3b demonstrates this patient's ~227 kb duplication in Chr6, with approximate coordinates of 33,202,640–33,429,672, which disrupts a TAD boundary. Duplications with similar breakpoints were
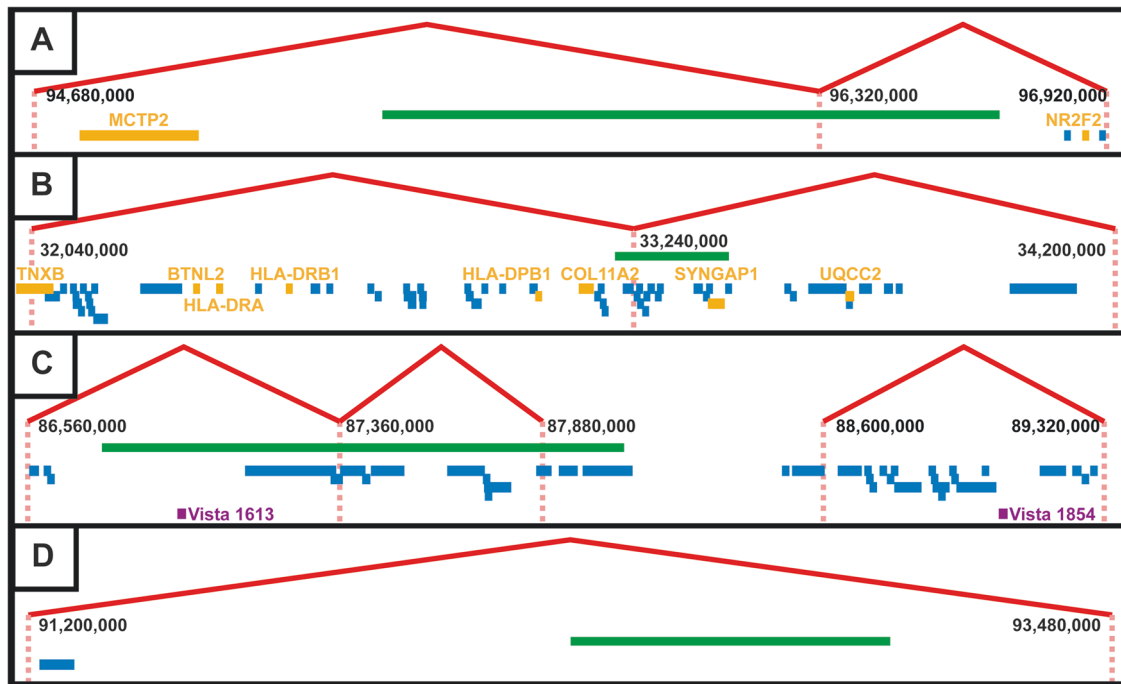
Fig. 3 Output from ClinTAD for cases **a–d**. Red lines and dashed red lines represent topologically associated domains (TADs) and their boundaries, green lines represent the patients' copy number variants (CNVs), purple lines represent VISTA enhancers, blue lines represent genes, and orange lines represent genes with an Human Phenotype Ontology (HPO) phenotype match. Cases A and B had CNVs overlapping a TAD boundary and HPO phenotype matches in adjacent TADs, making it plausible that TAD interruption contributed to these patients' phenotypes. Case C overlapped two TAD boundaries but had no HPO phenotype matches, whereas case D did not overlap a TAD boundary and also had no phenotype match, decreasing suspicion for clinical relevance

not identified in the ClinGen, DECIPHER, or DGV databases. The TADs adjacent to this CNV had eight gene matches with 11 HPO matches. The *COL11A2*, *SYNGAP1*, and *UQCC2* genes had matches for global developmental delay (HP:0001263), with *UQCC2* having an additional match for aggressive behavior (HP:0000718). Multiple genes had HPO matches for alopecia (HP:0001596) or sparse scalp hair (HP:0002209). Finally, the *TNXB* gene had a match for migraines (HP:0002076) and the *HLA-DPB1* gene had a match for headaches (HP:0002315). The statistics function showed this CNV had a higher weighted score at this location than at 497 out of 500 random locations. ClinTAD analysis therefore suggests that transcriptional dysregulation of neighboring genes based on TAD boundary disruption may plausibly relate to the patient phenotype, somewhat increasing suspicion for clinical significance.

Case C is a patient with a clinical history including velopharyngeal insufficiency and language processing disorder and a family history of relatives with a similar presentation. This patient had a deletion on Chr16 of ~1.4 Mb with approximate coordinates of 86,750,307–88,089,160 which included six protein-coding genes (*HTR1E, ZNF292, CGA, GJB7, SMIM8,* and *C6orf163*), none of which had any reported disease association per the OMIM (Online Mendelian Inheritance in Man) database. No deletions with

similar breakpoints were reported in any of the databases referenced above. In this case, ClinTAD analysis demonstrated disruption of two TAD boundaries but no genes with HPO phenotype matches in adjacent TADs. Therefore, this deletion retains suspicion for possible clinical significance based on TAD boundary disruption, size, and protein-coding genes in the region, but does not have supporting evidence of potential transcriptional disruption of more distal genes with known relevant phenotype.

Case D is a patient with a history of pulmonary artery hypertension. This patient's deletion on Chr6 of ~0.7 Mb, with approximate coordinates of 92,340,802–93,013,237, included no protein-coding genes and no cases with similar breakpoints in any of the databases referenced above. ClinTAD analysis demonstrated no disruption of a TAD boundary and no genes within the TAD with HPO matches. This analysis therefore helped decrease suspicion of pathogenicity for this VUS.

Discussion

Here we describe the use of ClinTAD, a clinical decision-support tool to aid in interpretation of clinical microarray data. Specifically, ClinTAD assesses TAD boundary disruption and reports matches to HPO phenotypes associated

with genes in CNV-adjacent TADs. In addition, ClinTAD statistics function uses simulations in comparison to randomly located CNVs of the same size to assess a weighted HPO phenotype score. Through ClinTAD analysis of 236 clinically annotated CNVs from our institution, we found that VUS shows significantly more gene matches and HPO matches than benign variants. These HPO-matching genes may potentially have altered transcription due to disruption of TADs or other regulatory elements. Matching genes outside the CNV region but within adjacent TADs may particularly aid in the clinical interpretation of a given CNV, as they would otherwise be ignored based on current reporting criteria. We further present four example cases to illustrate how ClinTAD can inform suspicion of pathogenicity for VUS found on cytogenomic microarray.

There are several limitations to the current version of ClinTAD. While the HPO system is extensive, correlating a patient's phenotype as described in clinical notes to an exact HPO ID is a recognized challenge [19]. Additionally, some HPO phenotypes, such as "global developmental delay" (HP:0001263), are so frequent that they do not provide meaningful information when a match is found. Furthermore, we use a single set of TAD boundary calls from H1 hESCs. While current data suggest that TAD boundaries are largely invariant across ~60–70% of human tissues [7], it is possible that tissue-specific TAD boundaries may be desired for interpretation. Therefore, users can easily implement custom tracks for TAD boundaries, enhancers, and benign or pathogenic CNVs by using the "Tracks" page. A very clear limitation of ClinTAD is that it cannot prove functionality of any phenotype-matching genes. Our analysis suggests that even well-known benign variants can demonstrate many HPO matches to nearby genes. Therefore, while ClinTAD may assist in broadening the scope of genes for consideration in any given clinical case, it must be noted that the mere presence of an HPO match by no means assures clinical relevance. Further literature exploration of gene function is absolutely required.

In general, much remains unknown about the relationship between TADs and clinical effects. For instance, in clinical testing, unlike in the research setting [20], there is no way to validate the differential effects of duplications and deletions on TAD boundaries or neighboring gene expression. A recent elegant study used a physics-based model to predict changes in TAD architecture in the context of a CNV [21], potentially coming closer to this goal, but this approach is not yet available for clinical use. Based on current reporting guidelines [1], we anticipate that ClinTAD may be most valuable in modulating suspicion for VUS regions with no or few protein-coding genes and with no similar CNVs in relevant patient databases. Currently, we do not anticipate ClinTAD to change a final clinical diagnosis, and results here may remain too speculative to

incorporate in clinical reporting given currently available data. However, as recently proposed by Mundlos and colleagues [8], interpretation algorithms that incorporate both epigenetic features and TAD effects could one day establish a new paradigm for clinical CNV interpretation. We propose ClinTAD as an easy-to-use tool to help move toward this goal.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med. 2011;13:680–85.
2. Flirth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84:524–33.
3. Kirkpatrick BE, Riggs ER, Azzariti DR, Miller VR, Ledbetter DH, Miller DT, et al. GenomeConnect: matchmaking between patients, clinical laboratories, and researchers to improve genomic knowledge. Hum Mutat. 2015;36:974–78.
4. MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 2014;42:986–92.
5. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161:1012–25.
6. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 2016;538:265–69.
7. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485:376–80.
8. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet. 2018;19:453–67.
9. Ibn-Salem J, Köhler S, Love MI, Chung HR, Huang N, Hurles ME. Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome Biol. 2014;15:423.
10. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011;43:838–46.

11. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017;45:865–76.

12. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015;518:331–36.

13. Assembling the Genome [Internet]. [Place unknown]. Genome Reference Consortium; Human Genome Assembly GRCh37. 2013. https://www.ncbi.nlm.nih.gov/grc/human/data?asm=GRCh37. Accessed 13 Sep 2017.

14. Köhler S, Carmody L, Vasilevsky N, Jacobsen JO, Danis D, Gourdine JP. et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019;47:D1018–27.

15. Human Phenotype Ontology. Human Phenotype Ontology [Internet]. Berlin: Human Phenotype Ontology; 2008. https://hpo.jax.org/app/. Accessed 13 Sep 2017.

16. Ensembl. Ensembl Genome Browser [Internet]. Hinxton: Ensembl; 2000. http://grch37.ensembl.org. Accessed 25 Sep 2018.

17. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46:754–61.

18. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 2007;35:88–92.

19. Son JH, Xie G, Yuan C, Ena L, Li Z, Goldstein A, et al. Deep phenotyping on electronic health records facilitates genetic diagnosis by clinical exomes. Am J Hum Genet. 2018;103:58–73.

20. Flöttmann R, Kragesteen BK, Geuer S, Socha M, Allou L, Sowińska-Seidler A, et al. Noncoding copy-number variations are associated with congenital limb malformation. Genet Med. 2018;20:599–607.

21. Bianco S, Lupiáñez DG, Chiariello AM, Annunziatella C, Kraft K, Schöpflin R. Polymer physics predicts the effects of structural variants on chromatin architecture. Nat Genet. 2018;50:662–67.