*Subject Section*

# CNV-ClinViewer: Enhancing the clinical interpretation of large copy-number variants online

Marie Macnee[1], Eduardo Pérez-Palma[2], Tobias Brünger[1], Chiara Klöckner[3], Konrad Platzer[3], Arthur Stefanski[4-5], Ludovica Montanucci[4], Allan Bayat[6-7], Maximilian Radtke[3], Ryan L Collins[8-9], Michael Talkowski[8-9], Daniel Blankenberg[4], Rikke S Møller[6-7], Johannes R Lemke[3], Michael Nothnagel[1,10], Patrick May[11] and Dennis Lal[1,4-5,8,*]

[1]Cologne Center for Genomics (CCG), University of Cologne, Germany., [2]Universidad del Desarrollo, Centro de Genética y Genómica, Facultad de Medicina Clínica Alemana, Santiago, Chile., [3]Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany., [4]Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA., [5]Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH, USA., [6]Department of Epilepsy Genetics and Personalized Medicine, Member of ERN Epicare, Danish Epilepsy Centre, Dianalund, Denmark., [7]Department of Regional Health Research, Faculty of Health Sciences, University of Southern Denmark, Denmark., [8]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA., [9]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA., [10]University Hospital Cologne, Cologne, Germany., [11]Luxembourg Centre for Systems Biomedicine, University Luxembourg, Esch-sur-Alzette, Luxembourg.

*To whom correspondence should be addressed.

## Abstract

**Motivation:** Pathogenic copy number variants (CNVs) can cause a heterogeneous spectrum of rare and severe disorders. However, most CNVs are benign and are part of natural variation in human genomes. CNV pathogenicity classification, genotype-phenotype analyses, and therapeutic target identification are challenging and time-consuming tasks that require the integration and analysis of information from multiple scattered sources by experts.

**Results:** Here, we introduce the CNV-ClinViewer, an open-source web-application for clinical evaluation and visual exploration of CNVs. The application enables real-time interactive exploration of large CNV datasets in a user-friendly designed interface and facilitates semi-automated clinical CNV interpretation following the ACMG guidelines by integrating the ClassifCNV tool. In combination with clinical judgment the application enables clinicians and researchers to formulate novel hypotheses and guide their decision-making process. Subsequently, the CNV-ClinViewer enhances for clinical investigators patient care and for basic scientists translational genomic research.

**Availability:** The web-application is freely available at https://cnv-ClinViewer.broadinstitute.org and the open-source code can be found at https://github.com/LalResearchGroup/CNV-clinviewer.

**Contact:** lald@ccf.org

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

## 1 Introduction

One of the reasons for genetic disorders is copy-number changes of one or more genes, resulting from deletions, duplications, or other genomic rearrangements. Most disease-associated CNVs are unique, cause severe complex disorders and are often hard to distinguish from benign CNVs frequently found in the general population (Riggs et al., 2020).

Although the average size of pathogenic/likely pathogenic CNVs in ClinVar (Landrum et al., 2018) of 9.7 Mb (95% CI 9.3-10.1 Mb, median=2.6 Mb) is likely overestimated due to frequent imprecise discovery of CNVs with arrays, disease-associated CNVs typically cover tens to hundreds of genes. Thus, pinpointing the underlying disease driver and modifier genes represents a major challenge. Given the complexity of CNV interpretation, the American College of Medical Genetics and Genomics (ACMG) has developed technical standards (a quantitative, evidence-based scoring framework) to standardize the evaluation process of the genomic content of a CNV region, and to promote consistency and transparency in classification and reporting across clinical laboratories (Riggs et al., 2020). The application of those standards involves clinical and genetic CNV data that are scattered across registries, databases, and the literature. Therefore, it is difficult and highly time-consuming to annotate, analyze and interpret CNVs manually in a clinical setting.

Several bioinformatics tools, often fully-automated, have been developed for clinicians and researchers to facilitate large-scale CNV classification according to the ACMG guidelines (for example see (Gurbich and Ilinsky, 2020; Geoffroy et al., 2018; Requena et al., 2021). However, no algorithm or even classification framework can perfectly capture data that is usually only available to experts, such as deep clinical, CNV, or gene-level information. This specifically affects data that cannot be extracted from public online resources, such as detailed information about the phenotype, including family history, CNV inheritance pattern, and many other criteria currently not integrated in the clinical significance classification. For the semi-automated tools (Fan et al., 2021), the extracted and manually entered information is not explorable within the same web-application interface and comparative visual inspection of CNVs with various genomic and clinical data sources is not possible because current tools do not provide such functionality (Gurbich and Ilinsky, 2020; Geoffroy et al., 2018; Fan et al., 2021; Requena et al., 2021). While the current focus of those existing tools is on clinical significance classification, they are not designed to perform ad-hoc evaluations of new data such as genotype-phenotype analyses to expose undiscovered patterns of CNV localization in patient cohorts. To show a correlation of patient CNVs with known disease genes at a specific locus or to narrow down dosage sensitive genes that are likely driver genes in particular CNVs, the interactive visual inspection of affected genomic content combined with crosslinked data from various sources can enhance the interpretational analysis of CNVs. Paradoxically, the individuals who collect the most relevant clinical information for CNV clinical significance classification and interpretation, such as genetic counselors, treating physicians, and patients' families, often lack the required bioinformatical expertise to examine several patient CNVs simultaneously and integrate information from multiple scattered sources time-efficiently. To overcome current limitations for biomedical CNV interpretation of those individuals and researchers, we developed the CNV-ClinViewer, a fully open-source exploration and interpretation platform that integrates analysis, annotation, guideline-based, classification and clinical evaluation of CNVs and provides users, even families with CNVs of uncertain significance, with an interface to curate, visualize, interact with and continuously re-evaluate the CNV data.

## 2 Methods

Details about the data sources used in the CNV-ClinViewer can be found in the Supplementary Table S1.

### 2.1 Genetic CNV variants and regions datasets

CNVs with annotated clinical significance were obtained from ClinVar (Landrum et al., 2018) (updated quarterly), and CNVs identified in the general population were obtained from the UK biobank (Aguirre et al., 2019) as well as from gnomAD (Collins et al., 2020) (controls-only dataset, version 2.1). GnomAD CNVs were reduced to those with a filter equal to PASS, and SVTYPE equivalent to 'DEL' or 'DUP', and all CNVs from the general population were filtered for large CNVs >50kb. Specifically, we collected 52,344 CNVs from ClinVar (10654 pathogenic/ likely pathogenic CNVs), 195,916 CNVs from the UK-biobank (>50kb), and 10,370 CNVs from gnomAD (>50kb). The genome coordinates of CNVs from the UK biobank and gnomAD were converted to GRCh38 using the UCSC LiftOver tool (Kent et al., 2002). (Riggs et al., 2020).

To annotate neutral as well as clinically relevant regions by use of a visual summary of the variants, we further processed the variants by counting the number of deletions or duplications in the different data sets using a sliding window approach with a window size of 200kb and step size of 100kb. For the cohort-level data from the UK-biobank and gnomAD, we divided the counts by the number of samples in the cohort to retrieve region-specific allele-frequencies.

Known CNV syndromes and genomic regions with evaluated dosage sensitivity were obtained from DECIPHER (Firth et al., 2009) and ClinGen (Rehm et al., 2015) (updated quarterly) databases.

### 2.2 Gene-level annotations

We retrieved a list of 18,792 protein-coding gene symbols from the HUGO Gene Nomenclature Committee (HGNC) (Tweedie et al., 2021). These genes were annotated by the genomic boundaries of one transcript from RefSeq (O'Leary et al., 2016) (GRCh37/hg19 and GRCh38/hg38), either the MANE select transcript or alternatively the longest transcript. In addition, we annotated gene-level scores for sequence constraint and dosage sensitivity in humans (Table 1) and gene-disease associations from ClinGen (Rehm et al., 2015) (updated quarterly).

**Table 1.** Description of gene-level scores for sequence constraint and dosage sensitivity. The given thresholds are used in the CNV-ClinViewer to annotate dosage sensitivity of genes.

### 2.3 CNV-ClinViewer web server development

The CNV-ClinViewer was developed with the Shiny framework of R studio software (v.1.7.1, https://shiny.rstudio.com/) which transforms

*Article short title*

regular R code into an interactive environment that can follow and 'react' to remote-user instructions. The pre-processed data alongside the R/Shiny code was uploaded as a stand-alone Ubuntu image with Google Cloud services. The image was deployed into a Google Virtual Machine (VM) using the googleComputeEngineR package (v.0.3.0, https://github.com/cloudyr/googleComputeEngineR). The CNV-ClinViewer web server (https://cnv-ClinViewer.broadinstitute.org/) is compatible with all commonly-used internet browsers.

For the classification of uploaded CNVs, the clinical interpretation tool ClassifyCNV (Gurbich and Ilinsky, 2020) including its own data sources was integrated into the CNV-ClinViewer framework. Upon selection of one of the uploaded and classified CNVs, all pre-processed data are intersected by genomic coordinates of the selected CNV. Consequently, the HTML report describing the selected CNV is generated using the rmarkdown R package (v2.11, https://rmarkdown.rstudio.com), and the interactive visualizations and tables are rendered using the ggplot2 (v.3.3.5, https://ggplot2.tidyverse.org), plotly (v.4.9.4.1 https://plotly-r.com) and DT (v.0.19, https://CRAN.R-project.org/package=DT) R packages. For the gene set enrichment analysis, the enrichr R package (v3.0, https://CRAN.R-project.org/package=enrichR) that provides an interface to the Enrichr database is used.

The CNV-ClinViewer is an open-source project, and its code will continue to grow and improve through version control in the GitHub repository (https://github.com/LalResearchGroup/CNV-clinviewer).

## 3    Results

We present the CNV-ClinViewer, a user-friendly web-application that semi-automatedly enhances the ACMG guideline-based CNV classification (Kuleshov *et al.*, 2016) for caregivers and facilitates the biomedical interpretation of gene content at a locus of interest for scientists. Development, workflow, and main features are presented in **Fig. 1**.

**Figure 1: Development and use of CNV-ClinViewer.** CNV data as well as clinical and genomic annotations are collected, processed and annotated, and quarterly updated. Users can copy-paste CNV(s) or upload CNVs in a tab-separated or Excel file for real-time exploration and interpretation.

### 3.1    Upload of one up to thousands of CNVs for clinical interpretation and genetic reports

The CNV-ClinViewer allows simultaneous analysis of single or multiple CNVs, irrespective of the technology used to identify them. It requires as input the genomic coordinates of CNVs based on the human reference genome GRCh37/hg19 (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.13/) or GRCh38/hg38 (https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.40/), either by copy-pasting or by uploading a tab-separated text or Excel file (max. number of CNVs = 10,000). Minimal required information for each CNV, including whole chromosome trisomies and monosomies, is the chromosome, start, end and CNV type (deletion or duplication). Optionally, the user can provide sample IDs, phenotype information for filtering and a score of manually assessed evidence categories of the 2019

ACMG standard guidelines (see below). Details and example files can be found in the help section and on the about page of the CNV-ClinViewer.

After CNV submission, the user is directed to the analysis interface. Here, five different analysis panels are available. In the *first analysis panel*, the uploaded CNVs are displayed in a downloadable table comprising their annotated clinical significance and details about the scoring (**Fig. 2A**). The CNVs are classified by ClassifyCNV (Gurbich and Ilinsky, 2020), a command-line tool integrated into the CNV-ClinViewer infrastructure that automatically evaluates the evidence categories 1A/B, 2A-H, 3A-C, 4O for copy-number losses and 1A/B, 2A-H, 2J-L, 3A-C, 4O for copy-number gains from the 2019 ACMG/ClinGen Technical Standards for CNVs (Riggs *et al.*, 2020). In case the user uploaded additional scores of relevant ACMG clinical significance criteria for which additional information, such as family history or the *de novo* status, is required, the score is added to the score from ClassifyCNV to refine the final score and classification. In the *second analysis panel*, a comprehensive report of the individual CNVs can be downloaded (**Fig. 2B**). The report includes details about the clinical significance classification and the overlap with established/predicted haploinsufficient/triplosensitive and clinically relevant genes and genomic regions.

**Figure 2: Semi-automated classification and download of CNV reports.** A) The integrated semi-automated classification of uploaded CNVs is based on 2019 ACMG/ClinGen Technical Standards for CNVs by ClassifyCNV[3] and is one of the key features of the CNV-ClinViewer. The resulting scores and applied evidence categories can be inspected in a table overview and also downloaded. B) A comprehensive report on individual CNVs, including details about the clinical significance classification and the overlap with established/predicted haploinsufficient/triplosensitive and clinically relevant genes and genomic regions, can be downloaded.

### 3.2    Uploaded CNV data can be visually inspected alongside publicly available data

The *third panel* of the analysis interface is a genomic viewer (**Fig. 3, Supplementary Fig. S1A-F**). Here, the uploaded CNVs and their genomic region can be inspected alongside biomedical annotations and other pathogenic and general population CNV datasets. Included are five data tracks that the user can interactively navigate by zooming in/out, moving horizontally, and selecting genomic regions of interest. In the first track of the genomic viewer (**Supplementary Fig. S1B**), the CNV-ClinViewer integrates seamless evaluation of the gene content by visualization of all protein-coding genes. Here, the genes are highlighted by a selection of gene dosage sensitivity scores (**Table 1**) to enable fast and visual gene prioritization in context to the uploaded CNVs. Below this gene track, all uploaded CNVs that intersect the selected region are visualized and can be interactively filtered based on uploaded phenotypic annotations, their assigned clinical significance, CNV type, and sample IDs (**Supplementary Fig. S1C**). The further tracks enable the visual identification of disease-prone regions with coupled visualizations of summarized data (allele frequencies and allele counts) of >250,000 CNVs from the general population (UK Biobank (Aguirre *et al.*, 2019), gnomAD (Collins *et al.*, 2020)) and pathogenic/likely pathogenic CNVs from ClinVar (Landrum *et al.*, 2018) (**Supplementary Fig. S1D-F**). In addition, users can identify

*K.Takahashi et al.*

the overlap of uploaded CNVs with individual ClinVar CNVs (**Supplementary Fig. S1D**). The latter are displayed in an interactive plot grouped by their assigned clinical significance (pathogenic/likely pathogenic vs. uncertain significance vs. benign/likely benign) and can be filtered based on their type and clinical significance. Details about the ClinVar CNVs, such as their reported phenotypes or allele origin and the link to the ClinVar variant website, are shown in a table and can be downloaded for further analyses.

**Figure 3: Genomic viewer.** A genomic viewer allows inspecting the uploaded CNVs and their genomic region alongside biomedical annotations and other pathogenic and general population CNV datasets. Here, the user can visually compare and dynamically filter uploaded CNVs, perform a seamless evaluation of the gene content, identify disease-prone regions and find unknown patterns of CNV localization to generate hypotheses for further research. An extended view of the genomic viewer can be found in Supplementary Fig S1.

### 3.3   Additional features and analyses provide users with advanced CNV insights

Below the genomic viewer, the user can retrieve and download information about the intersecting genes, classified gene-disease associations from ClinGen (Rehm *et al.*, 2015), known dosage sensitive regions and CNV syndromes (**Fig. 4A**). The gene table contains all protein-coding genes, their genomic coordinates and transcript IDs, links to the Online Mendelian Inheritance in Man database (OMIM, https://omim.org/), and gene dosage sensitivity scores (**Table 1**). All genes classified by ClinGen with reported associations with one or several diseases are shown in the gene-disease association table, including links to the comprehensive gene-disease association reports on the ClinGen website. In the ClinGen region table and the DECIPHER (Firth *et al.*, 2009) CNV syndrome table, all well-characterized dosage sensitive regions and developmental CNV syndromes that intersect a CNV/genomic region of interest can be retrieved.

In the last panel, the user can perform gene set enrichment analyses to infer knowledge about the genes from a selected genomic region by comparing it to >180 annotated gene sets representing prior biological knowledge such as pathways, Human Phenotype Ontology (HPO) terms (Köhler *et al.*, 2021), and known (rare) diseases (**Fig. 4B**).

**Figure 4: Additional features. A)** Users can retrieve, filter, and download information about the intersecting genes, gene-disease associations, known dosage sensitive regions, and known CNV syndromes. **B)** Users can perform gene-set enrichment analyses (GSEA) to infer information on genes within a selected genomic region or CNV by comparing it to >180 annotated gene sets representing prior biological knowledge such as pathways, Human Phenotype Ontology (HPO) terms and known (rare) diseases.

### 3.4   Example analysis: 9q33.3q34.11 microdeletions

To illustrate the utility of the CNV-ClinViewer, we performed an example analysis of 14 previously published CNVs (9q33.3q34.11 microdeletions) from patients with complex developmental disorders

from the literature (n=14) (Campbell *et al.*, 2012; Ehret *et al.*, 2015; Nambot *et al.*, 2016; Saitsu *et al.*, 2012). Here, the CNVs were first uploaded and automatically classified as (likely) pathogenic (**Supplementary Fig. S2-3**). Next, the smallest region of overlap with its dosage sensitive gene *STXBP1* could be immediately identified due to the interactive visualization of the overlapping patient CNVs (**Supplementary Fig. S4-5**). In addition, the CNV-ClinViewer could resolve the clinical heterogeneity of patients by enabling the user to filter the CNVs based on assigned phenotypes (**Supplementary Fig. S7**) and retrieve region-specific information about intersecting clinically relevant genes and their associated phenotypes (**Supplementary Fig. S6**) as well as overlapping patient CNVs from ClinVar (**Supplementary Fig. S8**). Overall, we demonstrate the CNV-ClinViewer's fast and seamless biomedical interpretation of CNVs and its potential to replicate research findings due to a large number of annotations and visual inspection capabilities. To perform a comparable analysis without the CNV-ClinViewer, the user would be required to classify the CNVs manually one at a time or with the support of an existing tool, visualize the CNVs with yet another tool or genomic viewer, visit several webpages, and process and annotate information from various databases. The example analysis, including an example of benign CNVs of chromosome 9, can be found in more detail and with step-by-step instructions illustrated with images in the **Supplementary File**.

## 4   Discussion

By aggregating >250,000 population and patient CNVs, annotating various gene-scores and clinical annotations, and integrating useful existing bioinformatics tools, we developed a novel and user-friendly interface as a decision support tool for the clinical evaluation of CNVs and comparative visual inspection.

CNV analysis and interpretation tools and, in general, clinical decision support tools can make patient care more efficient, cost-effective, and guideline-concordant. However, although the prediction results from many approaches are promising, their value is limited by their lack of interpretability and human intuition. The black-box nature of some pathogenicity predictions can further exacerbate trust issues and worsen the overall experience (Cai *et al.*, 2019). To address this issue, human-centered tools with interactive mechanisms can grant end-users more agency in guiding the interpretation and can be used for critical decision-making purposes beyond an algorithm (Babione *et al.*, 2020; Cai *et al.*, 2019). Therefore, in the development of the CNV-ClinViewer, we emphasized interactive analyses of the CNVs beyond a classification score and allow the user to dynamically inspect and compare the results in detail. In comparison, existing open-source tools (Fan *et al.*, 2021; Geoffroy *et al.*, 2018; Gurbich and Ilinsky, 2020) focus more on the classification scores than on the interactive inspection and further investigations. We believe that the comprehensive approach of the CNV-ClinViewer will empower users with more trust in the results, the agency to test hypotheses, and enable them to apply their domain knowledge while simultaneously leveraging the benefits of automation.

To ensure that a tool is easy to use and is designed for target users at each development stage, human-centered design methods and principles should be applied (Babione *et al.*, 2020; Bates *et al.*, 2003).

*Article short title*

Such methods draw from well-established design principles in many disciplines, including usability, visualizations, and interface design, and consider users' context and experience. As the effectiveness of clinical tools may be severely limited without application of those principles and could even contribute to adverse medical events (Graham *et al.*, 2008), we put great emphasis on developing a user-friendly interface with simple data upload with minimal requirements, a clear analysis workflow, and real-time results, while giving the user the ability to retrieve details and help, and modify the analysis when required. We also focused on the visual and intuitive inspection of CNV data and its interaction with the user. In particular, for genomic data, the integration of visualizations is essential for interpretation and hypothesis generation, essential for combining different data sources and discovering unexpected localization patterns of genomic annotations in a short amount of time, as well as a valuable aid in communicating discoveries (Nusrat *et al.*, 2019).

On the one hand, the CNV-ClinViewer will assist clinicians, clinical geneticists and genetic counselors analyzing and interpreting CNV data and improve objectivity and accuracy in the clinical work-up. On the other hand, the collected and integrated knowledge combined with clinical judgment will enable clinicians and researchers to formulate novel hypotheses and guide their decision-making process.

The current features of the tool are focused on large CNVs that intersect several genes when gene prioritization plays a role in the analysis. Future versions will increase the granularity of the tool, for example, by providing a more detailed gene track that includes exons and introns, and different transcripts with expression data. With that, it will bring more value for interpreting intragenic deletions and duplications. Also, to integrate with existing workflows, we plan to develop an application programming interface (API) to query the system and fetch the results without using a graphical user interface. Overall, the CNV-ClinViewer is the first framework that enables interactive exploration of CNV data and semi-automated variant classification. With its scalable architecture, the web-application is well suited for both genome-wide (re-)evaluations of large data sets and small number of CNVs. The CNV-ClinViewer is an open-source tool (https://www.cnv-ClinViewer.broadinstitute.org/), including complete documentation on the use of the website.

## Data availability

The data used in the CNV-ClinViewer are listed in Supplementary Table S1. All code and data are accessible at https://github.com/LalResearchGroup/CNV-clinviewer, and detailed documentation can be found at https://www.cnv-ClinViewer.broadinstitute.org/. The web-application is free and open to all users, and there is no login requirement.

## Author contributions

Conceptualization: D.L., M.M., E.P.; Data curation: M.M., E.P.; Methodology: D.L., M.M., E.P., P.M., M.N.,; Supervision: D.L., E.P.; Validation: T.B., C.K, K.P., A.S., L.M., A.B., M.R., R.C., D.B., J.L., M.T., R.S.M.; Visualization: M.M.; Writing-original draft: D.L., M.M., E.P.; Writing-review & editing: P.M., M.N., E.P., T.B., C.K, K.P., A.S., L.M., A.B., M.R., R.C., D.B., J.L., M.T., R.S.M.

## Ethics declaration

This study does not involve human subjects. We use as reference datasets copy number variants from public resources such as databases and published research studies, in which individual investigators or organizations have consented their individuals. For example, the CNV-ClinViewer contains individual copy-number variant (CNV) data from ClinVar, which is de-identified and publicly available (https://www.ncbi.nlm.nih.gov/clinvar/). CNV data from gnomAD (https://gnomad.broadinstitute.org/downloads) and the UK Biobank (https://biobankenginedev.stanford.edu/downloads) are aggregated and de-identified data. The submitters of all databases had obtained appropriate consent for data sharing.

## Conflict of interest

Disclosure: The authors declare no conflict of interest.

## References

Aguirre,M. *et al.* (2019) Phenome-wide Burden of Copy-Number Variation in the UK Biobank. *The American Journal of Human Genetics*, **105**, 373–383.

Babione,J.N. *et al.* (2020) Human-centred design processes for clinical decision support: A pulmonary embolism case study. *International Journal of Medical Informatics*, **142**, 104196.

Bates,D.W. *et al.* (2003) Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality. *J Am Med Inform Assoc*, **10**, 523–530.

Cai,C.J. *et al.* (2019) Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19. Association for Computing Machinery, New York, NY, USA, pp. 1–14.

Campbell,I.M. *et al.* (2012) Novel 9q34.11 gene deletions encompassing combinations of four Mendelian disease genes: STXBP1, SPTAN1, ENG, and TOR1A. *Genet Med*, **14**, 868–876.

Collins,R.L. *et al.* (2020) A structural variation reference for medical and population genetics. *Nature*, **581**, 444–451.

Ehret,J.K. *et al.* (2015) Microdeletions in 9q33.3-q34.11 in five patients with intellectual disability, microcephaly, and seizures of incomplete penetrance: is STXBP1 not the only causative gene? *Mol Cytogenet*, **8**, 72.

Fan,C. *et al.* (2021) AutoCNV: a semiautomatic CNV interpretation system based on the 2019 ACMG/ClinGen Technical Standards for CNVs. *BMC Genomics*, **22**, 721.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

*K.Takahashi et al.*

Firth,H.V. *et al.* (2009) DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *The American Journal of Human Genetics*, **84**, 524–533.

Geoffroy,V. *et al.* (2018) AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics*, **34**, 3572–3574.

Graham,T.A.D. *et al.* (2008) How Usability of a Web-Based Clinical Decision Support System Has the Potential to Contribute to Adverse Medical Events. *AMIA Annu Symp Proc*, **2008**, 257–261.

Gurbich,T.A. and Ilinsky,V.V. (2020) ClassifyCNV: a tool for clinical annotation of copy-number variants. *Sci Rep*, **10**, 20375.

Kent,W.J. *et al.* (2002) The Human Genome Browser at UCSC. *Genome Res.*, **12**, 996–1006.

Köhler,S. *et al.* (2021) The Human Phenotype Ontology in 2021. *Nucleic Acids Res*, **49**, D1207–D1217.

Kuleshov,M.V. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*, **44**, W90-97.

Landrum,M.J. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*, **46**, D1062–D1067.

Nambot,S. *et al.* (2016) 9q33.3q34.11 microdeletion: new contiguous gene syndrome encompassing STXBP1, LMX1B and ENG genes assessed using reverse phenotyping. *Eur J Hum Genet*, **24**, 830–837.

Nusrat,S. *et al.* (2019) Tasks, Techniques, and Tools for Genomic Data Visualization. *Comput Graph Forum*, **38**, 781–805.

O'Leary,N.A. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, **44**, D733-745.

Rehm,H.L. *et al.* (2015) ClinGen — The Clinical Genome Resource. *N Engl J Med*, **372**, 2235–2242.

Requena,F. *et al.* (2021) CNVxplorer: a web tool to assist clinical interpretation of CNVs in rare disease patients. *Nucleic Acids Research*, **49**, W93–W103.

Riggs,E.R. *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.

Saitsu,H. *et al.* (2012) Association of genomic deletions in the STXBP1 gene with Ohtahara syndrome. *Clin Genet*, **81**, 399–402.

Tweedie,S. *et al.* (2021) Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res*, **49**, D939–D946.