



**HAL**  
open science

## The AnnotSV webserver in 2023: updated visualization and ranking

Véronique Geoffroy, Jean-Baptiste Lamouche, Thomas Guignard, Samuel Nicaise, Arnaud Kress, Sophie Scheidecker, Antony Le Béhec, Jean Muller

### ► To cite this version:

Véronique Geoffroy, Jean-Baptiste Lamouche, Thomas Guignard, Samuel Nicaise, Arnaud Kress, et al.. The AnnotSV webserver in 2023: updated visualization and ranking. *Nucleic Acids Research*, 2023, 10.1093/nar/gkad426 . hal-04151442

**HAL Id: hal-04151442**

**<https://hal.science/hal-04151442>**

Submitted on 4 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The AnnotSV webserver in 2023: updated visualization and ranking

Véronique Geoffroy<sup>1,2,\*</sup>, Jean-Baptiste Lamouche<sup>2,3,†</sup>, Thomas Guignard<sup>4</sup>, Samuel Nicaise<sup>3</sup>, Arnaud Kress<sup>5</sup>, Sophie Scheidecker<sup>2,6</sup>, Antony Le Béchech<sup>3,‡</sup> and Jean Muller<sup>2,3,6,\*</sup>

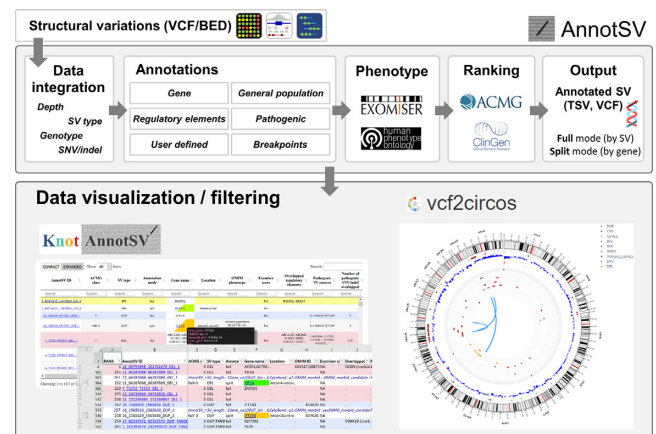
<sup>1</sup>Université de Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France, <sup>2</sup>Laboratoire de Génétique Médicale, UMR 1112, INSERM, IGMA, Université de Strasbourg, Strasbourg, France., <sup>3</sup>Unité Fonctionnelle de Bioinformatique Médicale appliquée au diagnostic (UF7363), Hôpitaux Universitaires de Strasbourg, Strasbourg, France, <sup>4</sup>Unité de Génétique Chromosomique, CHU Montpellier, France, <sup>5</sup>Complex Systems and Translational Bioinformatics, ICube, UMR 7357, University of Strasbourg, CNRS, FMTS, Strasbourg, France. and <sup>6</sup>Laboratoires de Diagnostic Génétique, IGMA, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.

Received March 26, 2023; Revised April 20, 2023; Editorial Decision May 08, 2023; Accepted May 09, 2023

## ABSTRACT

Much of the human genetics variant repertoire is composed of single nucleotide variants (SNV) and small insertion/deletions (indel) but structural variants (SV) remain a major part of our modified DNA. SV detection has often been a complex question to answer either because of the necessity to use different technologies (array CGH, SNP array, Karyotype, Optical Genome Mapping...) to detect each category of SV or to get an appropriate resolution (Whole Genome Sequencing). Thanks to the deluge of pangenomic analysis, Human geneticists are accumulating SV and their interpretation remains time consuming and challenging. The AnnotSV webserver (<https://www.lbgi.fr/AnnotSV/>) aims at being an efficient tool to (i) annotate and interpret SV potential pathogenicity in the context of human diseases, (ii) recognize potential false positive variants from all the SV identified and (iii) visualize the patient variants repertoire. The most recent developments in the AnnotSV webserver are: (i) updated annotations sources and ranking, (ii) three novel output formats to allow diverse utilization (analysis, pipelines), as well as (iii) two novel user interfaces including an interactive circos view.

## GRAPHICAL ABSTRACT



## INTRODUCTION

In the era of high throughput sequencing, structural variants (SV) are commonly described as genetics variants of a minimal size of 50 bp (base pairs) (1). They include a variety of genetic events grouped in two categories: balanced (inversion, intra or interchromosomal translocation) and unbalanced (deletion, insertion) that can have multiple origins (tandem duplication, mobile element insertion, segmental duplication...) and degrees of complexity (i.e. chromoanagenesis) (2). Accessing the full SV repertoire and its dynamic is crucial for understanding their role in human genome evolution (3). From cancer to inherited disorders, SV have been implicated in a wide range of diseases, making it an area of intense interest for medical scientists. Thanks to the

\*To whom correspondence should be addressed. Tel: +33 3 69 55 07 77; Fax: +33 3 69 55 18 94; Email: jeanmuller@unistra.fr

Correspondence may also be addressed to Véronique Geoffroy. Email: veronique.geoffroy@inserm.fr

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors

‡The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint last authors

progress of molecular tools more and more reliable SV calls are achieved leading to a deluge of genetic variants to interpret in diagnostic and research laboratories. Considering all these constraints, AnnotSV was developed and frequently updated since 2017 (4). A prioritization module was introduced in the second version. The third version integrated a phenotype matching module and a web based visualization (5). AnnotSV and its webserver have a wide range of applications, including single case reports and broader analyses in rare diseases (6,7) and cancer (8,9) demonstrating clinical application (10,11). Here, we provide an updated version of AnnotSV integrating new developments to improve ranking and visualization.

## PROGRAM DESCRIPTION AND METHODS

### General description of the web server

Keeping with the general functioning of the AnnotSV webserver, our workflow is separated into the annotation part (AnnotSV engine) and the analysis and visualization part including the knotAnnotSV and vcf2circos modules (Figure 1).

### Genomic annotations: categories and data sources

The annotation engine has been updated to version 3.3 and includes updated datasets covering the whole range of annotations types: Gene, Regulatory elements, Pathogenic or Benign genomic regions and Breakpoints (Table 1). Some of the annotations are linked to the gene name and thus provided independently of the genome build. In order to make sure AnnotSV remains a prime resource for researchers and geneticists, data annotations are regularly updated (once or twice a year) and novel sources are integrated. As an example, curated gene-disease relationships are now reported using the Gene Curation Coalition (GenCC) (12). Around 328000 benign regions (Gain/Loss/Insertion/Inversion) were retrieved from different sources (gnomAD (13), ClinVar (14), ClinGen (15)...). As an example, the Children's Mercy Research Institute database has been added (~100 000 genomes from long read Whole Genome Sequencing or WGS). Around 27 300 pathogenic regions (Gain, Insertion, Loss) were retrieved from ClinVar, ClinGen, dbVar (16) and OMIM (17). The regulatory elements annotation category has been complemented with the miR-TargetLink database (18). The exhaustive list of annotations is given in Supplementary Table S1. As previously described, the AnnotSV server is providing annotations using two complementary modes. The 'Full mode' considers the SV as a whole and integrates annotations and ranking of the highest scoring elements overlapped. The 'Split mode' considers each gene within the SV and provides detailed annotations for each overlapped gene (transcripts, LOEUF score ...).

### Updated ranking of SV

AnnotSV provides for each SV the five standard classes based on the ACMG/ClinGen recommendations (19). The detailed scoring schemes are available in Supplementary Table S2. The automatic ranking has been updated with 8

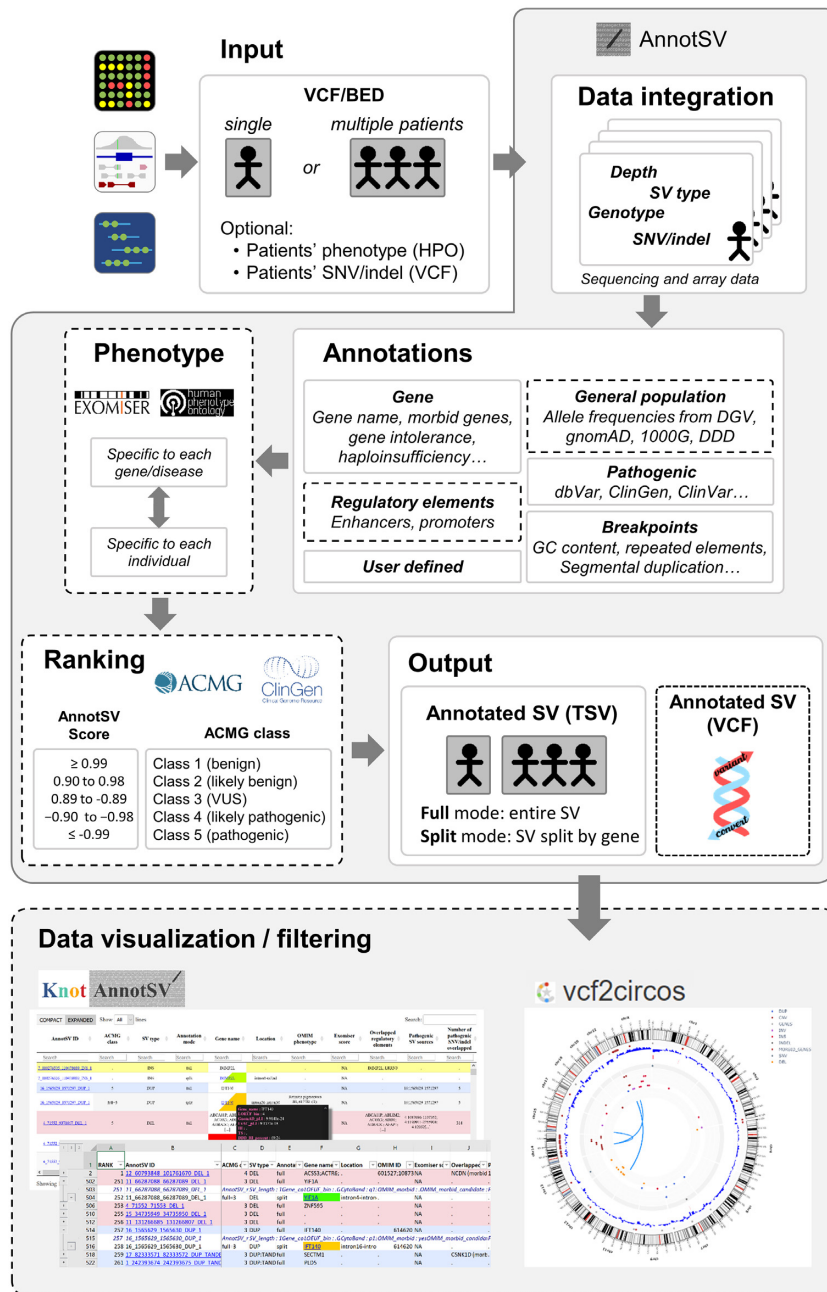
**Table 1.** Summary of annotation sources and their versions available in the current version for the human GRCh38 version. Updated and novel annotation sources are highlighted in bold

Annotations source	Version
<b>... Gene annotations</b>	
Gene annotations (RefSeq)	17-08-2020
Gene annotations (ENSEMBL)	<b>04-06-2022</b>
<b>... Regulatory Elements annotations</b>	
Promoter data (RefSeq)	17-08-2020
Promoter data (ENSEMBL)	24-10-2020
EnhancerAtlas 2.0	11-06-2019
GeneHancer	Licence required
miRTargetLink	<b>11/12/2020 (v2.0, data provided by the authors)</b>
<b>... Gene-based annotations</b>	
GenCC	<b>02-09-2022</b>
OMIM	<b>03-09-2022</b>
ACMG	ACMG SF v3.1
Gene intolerance (gnomAD)	V2.1.1
Gene intolerance (ExAC)	23-08-2016
Haploinsufficiency (DDD)	13-07-2020
Haploinsufficiency and triplosensitivity (ClinGen)	<b>05-09-2022</b>
Exomiser	<b>06/09/2022 (v2202)</b>
NCBI gene ID	<b>06-09-2022</b>
<b>... Annotations with known pathogenic genes or genomic regions</b>	
ClinVar	<b>03-09-2022</b>
ClinGen	<b>06-09-2022</b>
dbVar	<b>03-08-2022</b>
OMIM	<b>03-09-2022</b>
<b>... Annotations with known pathogenic SNV/indel</b>	
ClinVar	<b>03-09-2022</b>
<b>... Annotations with known benign genes or genomic regions</b>	
gnomAD (GRCh37)	06/03/2019 (v2.1)
ClinVar	<b>03-09-2022</b>
ClinGen	<b>08-09-2022</b>
DGV annotations	25-02-2020
DDD annotations (GRCh37)	09/2015 (v9.2)
1000 genomes annotations	21-05-2017
Children's Mercy Research Institute annotations	<b>27-10-2021</b>
Ira M. Hall's lab annotations	31-12-2018
<b>... Annotations with features overlapped with the SV</b>	
COSMIC annotations	Licence required
TAD boundaries annotations	User downloaded
<b>... Breakpoints annotations</b>	
GRCh38 FASTA genome	23-01-2014
Cytoband	<b>11-03-2019</b>
Repeated sequences annotations	<b>08-09-2022</b>
Segmental Duplication annotations	<b>15-10-2021</b>
ENCODE blacklist annotations	2018 (v2)
GAP regions annotations	<b>15-10-2021</b>

subsections (2B, 2G and 4O criteria for the 'Loss' category and 2B, 2C, 2F, 2G and 4O for the 'Gain' category). To be more comprehensive, we have now included an improved list of criteria and scores for each SV (e.g. AnnotSV\_ranking\_criteria column in the annotated file).

### Phenotype-driven prioritization of SV

To link the patient's phenotypic data to the already available knowledge for each gene, we provide a phenotype driven prioritization module based on the HPO (Human Phenotype Ontology) dataset (20) and Exomiser (21) to score each SV (from 0 to 1.0). We generally recommend a score above 0.7 for a known disease gene and 0.5 for a gene not yet re-



**Figure 1.** Schematic overview of the AnnotSV webservice workflow. The webservice architecture comprises a two-tiered framework: first an annotation engine with AnnotSV (submitting a query, processing the annotations) and second a visualization and filtering interface with knotAnnotSV and vcf2circo (showing results, generating the visualization and filtering system). The variantconvert module is used for VCF conversion. Part of the workflow with an update compared to the previous version is highlighted using a black dotted rectangle.

lated to a disease. The phenotype driven module has been updated with one of the latest versions of Exomiser (v2202).

**Inputs/outputs**

Multiple input and output formats are supported by the AnnotSV webservice corresponding to different usages. One can query directly the server using SV coordinates (format is Chromosome:Begin-End SVType) or using direct URL pre-filling (Supplementary data and supplementary Table S3) as an easy and fast way to gather annotations and classi-

fication for a single SV. Alternatively and if one requires a larger analysis (multiple SV), a BED (browser extensible data) file (22) or a VCF (Variant Call Format) file (23) can be submitted. It is to notice that the square-bracketed notation (ALT field in the VCF format) is now fully supported. Each square-bracketed event is interpreted as either a deletion, duplication, insertion, inversion or translocation and not only as a breakend. The user can add optional sample related information such as the HPO terms to help prioritization or SNV/indels (additional VCF) to help identify false positive deletion. Besides the classical TSV (Tab Sep-

arated Values) and HTML files, we have three new output files including a XLSM, a circos and a VCF file. The latter one is supported by our new module variantconvert that allows a reliable conversion from TSV to VCF files. Outputs can be downloaded for later use or visualized directly in a web browser.

## INTERFACE: VISUALIZATION AND FILTERING SYSTEM

Several user-friendly interfaces are provided with the AnnotSV webserver to annotate, rank and display the SV of interest to the user (Figure 1).

### Spreadsheet XLSM

Besides the HTML Table browser, knotAnnotSV provides a new downloadable spreadsheet XLSM file. It shares the same features as the HTML output, displaying Full and Split lines but is dedicated to larger input files (e.g. >1000 SV). For example, Case study 5 (WGS, supplementary data) that contains 13468 SV is converted into a 7.3 Mo XLSM file. It provides annotations enriched formula bar to facilitate handling and interpretation (Supplementary Figure S1). The user can benefit from the native filtering options of the spreadsheets programs.

### Vcf2circos

Circos (24) was developed to provide a visual framework capable of displaying large amount of genomic data. Our implementation in vcf2circos allows a circular display based on a VCF file organized following multiple layer (Figure 2). The outer band carries each chromosome and their cytoband. The user can display either all chromosomes or only chromosomes overlapping a SV (default). From the outer rings to the inner rings, RefSeq genes (gray) and OMIM morbid genes (red) (17) are shown following additional genomic features such as the GC rate thanks to the UCSC (25). Next layers are used to display SV divided in seven 'rings'. The upper ring is used to display additional SNV/Indels (false positive checkup) following by SV from 0 to 5 and more copies. The Plotly library combined with the javascript backend offers a completely interactive range of graphics charts for a fast web-based visualization (standalone HTML files) or directly from python-build applications. Hover is available for every item (SV and genes) to present specific annotations and zooming as well as automatic rescaling is possible. The legend is also interactive and allows users to select one or more items to display. Moreover, an easy cross-platform export utility is provided (e.g. kaleido) for generating static high quality images (.png).

## RESULTS AND DISCUSSION

Performance and usefulness of the webserver has been assessed on multiple datasets (5 case studies) from different sources including CGH/SNP arrays, WGS and OGM (Optical Genome Mapping) (26) (Table 2, Supplementary Data and Supplementary Figures S4–S6). In total, from 1 to 13 468 SV ranging in size from 50 bp to 6 Mb were annotated using the GRCh37 or GRCh38 build of the human

genome with or without HPO terms. As a result, the webserver completed the annotation within a reasonable time frame (~1–5 min) and the vcf2circos too (few seconds to 7 min) (Table 2). Adding HPO terms increases the running time.

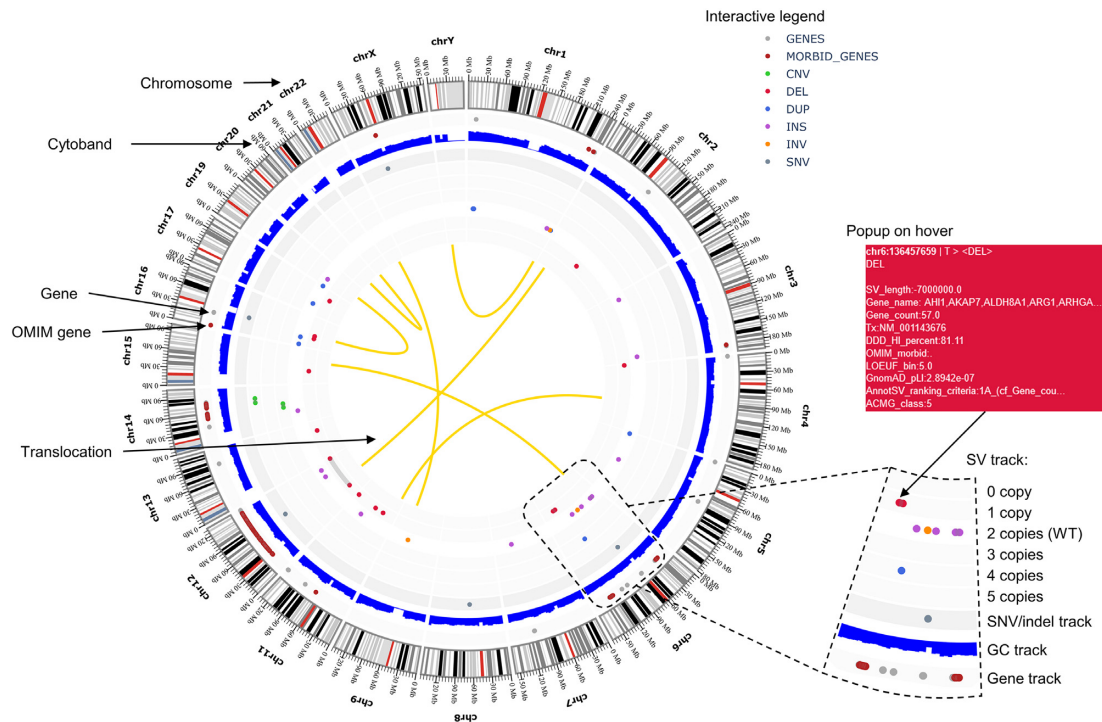
### Case study 1: SNP array (causal regulatory element overlapped)

SNP array (Illumina Infinium HumanCytoSNP-12) in a patient with severe developmental delay, hypotonia, seizure and strabismus detected one loss and two gains of one copy. Among those SV, we identified a 3 Mb heterozygous deletion on 5q14.3 encompassing 5 OMIM genes and the *MEF2C* regulatory elements. Deletions of coding regions or regulatory elements of this gene are known to cause the 5q14.3 deletion syndrome (27). This illustrates how annotation of regulatory elements can identify causative SV (Supplementary Figure S2).

### Case study 2: SNP array and OGM (rare complex chromosomal mechanism)

SNP array (Illumina Infinium HumanCytoSNP-12) in a patient with global developmental delay and recurrent infections identified 4 losses and one gain of a single copy. Among those a *de novo* 5.6 Mb heterozygous 12q deletion including 17 OMIM genes. Complementary analysis using OGM revealed an additional pericentromeric inversion of chromosome 12 likely sharing a common breakpoint with the deletion (Supplementary Figure S3).

These examples illustrate how SV detection is increasingly supported in research and diagnostic laboratories (28), leading to a growing demand for SV annotations. Our webserver processes on average 1800 submissions per month, corresponding to 8700000 annotated SV. Several tools have been developed for SV annotation such as VEP (29), CNVExplorer (30) or ClassifyCNV (31) (for review, see (32)). Thanks to the ACMG/ClinGen guidelines, SV ranking is based on a quantitative scoring framework. However, several methods have been proposed to assess SV pathogenicity such as CADD-SV (33), DeepSVP (34), StrVCTVRE (35) or SvAnna (36). AnnotSV is one of the most comprehensive tool available for annotation and prioritization of human SV (Supplementary Table S4). Visualization of SV is another complex topic (for review (37)) with multiple ways to represent each variant either linear like IGV (38) or circular like circos (24) that should ultimately help human geneticist to interpret clinically relevant SV within a large number of false positives. The current version of the server is supporting the last 2 versions of the human genome (e.g. GRCh37/38) and the next version shall integrate the T2T-CHM13 build (39). Regarding the interface, the vcf2circos is a new display that allows an integrated view of the genetic repertoire of one sample. Several tools exists to display genomic data in a circular way (for review (40)) but none fulfilled our requirements. However, further developments will include more communication between the various interfaces of the webserver to allow a combined analysis. With the recent updates, AnnotSV handles almost all input/output formats describing SV (BED, VCF,



**Figure 2.** vcf2circos interface overview. The circos represents the chromosome in a circular view allowing a quick overview along the whole genome or selected chromosomes. The interface includes multiple annotations tracks: chromosomes, cytobands, genes and morbid OMIM genes (red), GC content, CNV dedicated regions (from 0 to 5 and more copies) including inversions. The different SV are represented using the following scheme (gnomAD inspired (42)): deletion (DEL in dark red), duplication (DUP in blue), insertion (INS in purple), inversion (INV in orange) and CNV with no specific type (CNV in green). The popup includes various information such as the SV coordinates and length, overlapping transcripts and genes, haploinsufficiency cytoband...

**Table 2.** Comparison of computational speeds for the annotation and rendering of the AnnotSV webserver using different datasets

SV count	SV type	Min SV size (bp) DEL/DUP	Max SV size (bp) DEL/DUP	Mean SV size (bp) DEL/DUP	Median SV size (bp) DEL/DUP	Running time without HPO	Running time with HPO	Running for vcf2circos
86	40 DEL; 46 DUP	427/3 066	492 402/1 210 830	98 628/124 539	61 891/64 472	~1 min	~1 min	~12 s
4 190	1356 DEL 2691 INS 34 DUP 109 INV	338/11 473	6 155 870/2 190 530	88 798/276 151	2 292/57 364	~2 min		~3 min
13 468	DEL 5823; DUP 283; INS 1204; INV 203; BND 5955	50/108	660 427/442 801	811/4 468	87/652	~4 min	~5 min	~7 min
Smallest SV	1 DEL	50				~10 s		~3 s
Largest SV	1 DEL		660 427			~1 min		~3 s

SV datasets from different sources (SNP arrays, WGS and OGM) were annotated to evaluate the running time depending on the number of SV and the integration of human phenotype. The SNP arrays data is based on 20 samples and the average running time is given. DUP: duplication, DEL: deletion, INS: insertion, INV: inversion, TRA: translocation, BND: breakend.

TSV). However, this is not always an easy step for biologists. The recently proposed BEDPE format (browser extensible data paired-end format), shall overcome several limitations. The forthcoming release of our web server will be able to handle this new format. The annotation engine undergoes regular updates, incorporating the latest versions of the largest available datasets. Additionally, novel datasets, like the 100KGP (41), emerge on a regular basis, necessitating careful handling for a seamless integration.

Our webserver addresses the need for an integrated annotation and user-friendly visualization SV tool. It should help medical geneticists and scientists to analyze and understand the SV repertoire derived from any molecular biology assay that can produce SV calls including array based techniques (SNP/CGH), sequencing techniques (NGS datasets such as WGS) or other technologies such as OGM, and to guide precision medicine.

## DATA AVAILABILITY

The web server is available for free public usage at <https://lbgf.fr/AnnotSV> together with examples. A detailed documentation and a youtube channel have been extensively prepared (<https://www.youtube.com/channel/UCfIYQytaZJmVnpVxk480JKw>). The different source codes are available from our github repositories: <https://github.com/lmggeo/AnnotSV>, <https://github.com/mobidic/knotAnnotSV>, <https://github.com/bioinfo-chru-strasbourg/vcf2circos>, and <https://github.com/SamuelNicaise/variantconvert/>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We wish to thank AnnotSV users for their constant support suggesting new annotations and features. We would like to thank especially Rodrigo Martín for his assistance to interpret the square-bracketed SV break-ends notations (<https://github.com/EUCANCan/variant-extractor#homogenization-rules>). We would like also to acknowledge Dr Jin Cui for providing the PCircos repository and the members of the diagnostic laboratories at the University Hospital in Strasbourg for technical assistance. We wish to thank Pierre-Alain Rollat and Caroline Schluth-Bolard for providing us with the chromoanagenesis examples. WGS was performed at Integragen. OGM was performed thanks to Bionano Genomics and the GENTYANE platform. The authors would like to thank the BiGest-ICube Platform for assistance.

## FUNDING

University of Strasbourg, Strasbourg University Hospital [EJP RD Solve-RET EJPRD19-234, CREGEMES]. Funding for open access charge: Strasbourg University Hospital. *Conflict of interest statement.* None declared.

## REFERENCES

- Ho, S.S., Urban, A.E. and Mills, R.E. (2020) Structural variation in the sequencing era. *Nat. Rev. Genet.*, **21**, 171–189.
- Holland, A.J. and Cleveland, D.W. (2012) Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. *Nat. Med.*, **18**, 1630–1638.
- Soto, D.C., Uribe-Salazar, J.M., Shew, C.J., Sekar, A., McGinty, S.P. and Dennis, M.Y. (2023) Genomic structural variation: a complex but important driver of human evolution. *Am. J. Biol. Anthropol.*, <https://doi.org/10.1002/ajpa.24713>.
- Geoffroy, V., Herenger, Y., Kress, A., Stoetzel, C., Piton, A., Dollfus, H. and Muller, J. (2018) AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics*, **34**, 3572–3574.
- Geoffroy, V., Guignard, T., Kress, A., Gaillard, J.-B., Solli-Nowlan, T., Schalk, A., Gatinois, V., Dollfus, H., Scheidecker, S. and Muller, J. (2021) AnnotSV and knotAnnotSV: a web server for human structural variations annotations, ranking and analysis. *Nucleic Acids Res.*, **49**, W21–W28.
- Garret, P., Chevarin, M., Vitobello, A., Verdez, S., Fournier, C., Verloes, A., Tisserant, E., Vabres, P., Prevel, O., Philippe, C. *et al.* (2022) A second look at exome sequencing data: detecting mobile elements insertion in a rare disease cohort. *Eur. J. Hum. Genet.*, <https://doi.org/10.1038/s41431-022-01250-3>.
- Oud, M.S., Smits, R.M., Smith, H.E., Mastroianni, F.K., Holt, G.S., Houston, B.J., de Vries, P.F., Alobaidi, B.K.S., Batty, L.E., Ismail, H. *et al.* (2022) A de novo paradigm for male infertility. *Nat. Commun.*, **13**, 154.
- Talsania, K., Shen, T.-W., Chen, X., Jaeger, E., Li, Z., Chen, Z., Chen, W., Tran, B., Kusko, R., Wang, L. *et al.* (2022) Structural variant analysis of a cancer reference cell line sample using multiple sequencing technologies. *Genome Biol.*, **23**, 255.
- Shah, J.B., Pueschl, D., Wubbenhorst, B., Fan, M., Pluta, J., D'Andrea, K., Hubert, A.P., Shilan, J.S., Zhou, W., Kraya, A.A. *et al.* (2022) Analysis of matched primary and recurrent BRCA1/2 mutation-associated tumors identifies recurrence-specific drivers. *Nat. Commun.*, **13**, 6728.
- Liu, Z., Roberts, R., Mercer, T.R., Xu, J., Sedlazeck, F.J. and Tong, W. (2022) Towards accurate and reliable resolution of structural variants for clinical diagnosis. *Genome Biol.*, **23**, 68.
- Spillmann, R.C., Tan, Q.K.-G., Reuter, C., Schoch, K., Undiagnosed Diseases Network, Kohler, J., Bonner, D., Zastrow, D., Alkelai, A., Baugh, E. *et al.* (2022) A concurrent dual analysis of genomic data augments diagnoses: experiences of 2 clinical sites in the Undiagnosed Diseases Network. *Genet. Med.*, **25**, 100353.
- DiStefano, M.T., Goehringer, S., Babb, L., Alkuraya, F.S., Amberger, J., Amin, M., Austin-Tse, C., Balzotti, M., Berg, J.S., Birney, E. *et al.* (2022) The Gene Curation Coalition: a global effort to harmonize gene-disease evidence resources. *Genet. Med.*, **24**, 1732–1742.
- Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H. *et al.* (2020) A structural variation reference for medical and population genetics. *Nature*, **581**, 444–451.
- 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.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
- 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.
- Lappalainen, I., Lopez, J., Skipper, L., Hefferon, T., Spalding, J.D., Garner, J., Chen, C., Maguire, M., Corbett, M., Zhou, G. *et al.* (2013) DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res.*, **41**, D936–D941.
- Hamosh, A., Scott, A.F., Amberger, J., Valle, D. and McKusick, V.A. (2000) Online Mendelian Inheritance in Man (OMIM). *Hum. Mutat.*, **15**, 57–61.
- Kern, F., Aparicio-Puerta, E., Li, Y., Fehlmann, T., Kehl, T., Wagner, V., Ray, K., Ludwig, N., Lenhof, H.-P., Meese, E. *et al.* (2021) miRTargetLink 2.0—interactive miRNA target gene and target pathway networks. *Nucleic Acids Research*, **49**, W409–W416.
- Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C. *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.
- Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J.O.B., Danis, D., Gourdi, J.-P., Gargano, M., Harris, N.L., Matentzoglou, N., McMurry, J.A. *et al.* (2019) Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.*, **47**, D1018–D1027.
- Smedley, D., Jacobsen, J.O.B., Jager, M., Köhler, S., Holtgrewe, M., Schubach, M., Siragusa, E., Zemojtel, T., Busch, O.J., Washington, N.L. *et al.* (2015) Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat. Protoc.*, **10**, 2004–2015.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.
- Lee, C.M., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Gonzalez, J.N., Hinrichs, A.S., Lee, B.T., Nassar, L.R., Powell, C.C. *et al.* (2019) UCSC Genome Browser enters 20th year. *Nucleic Acids Res.*, **48**, D756–D761.
- Mantere, T., Neveling, K., Pebrel-Richard, C., Benoist, M., van der Zande, G., Kater-Baats, E., Baatout, I., van Beek, R., Yammine, T., Oorsprong, M. *et al.* (2021) Optical genome mapping enables constitutional chromosomal aberration detection. *Am. J. Hum. Genet.*, **108**, 1409–1422.
- Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.-A. *et al.* (2017) The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.*, **49**, 36–45.
- Raca, G., Astbury, C., Behlmann, A., De Castro, M.J., Hickey, S.E., Karaca, E., Lowther, C., Riggs, E.R., Seifert, B.A., Thorland, E.C. *et al.* (2023) Points to consider in the detection of germline structural variants using next-generation sequencing: a statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.*, **25**, 100316.

29. Hunt,S.E., Moore,B., Amode,R.M., Armean,I.M., Lemos,D., Mushtaq,A., Parton,A., Schuilenburg,H., Szpak,M., Thormann,A. *et al.* (2022) Annotating and prioritizing genomic variants using the Ensembl Variant Effect Predictor-A tutorial. *Hum. Mutat.*, **43**, 986–997.
30. 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.
31. Gurbich,T.A. and Ilinsky,V.V. (2020) ClassifyCNV: a tool for clinical annotation of copy-number variants. *Sci. Rep.*, **10**, 20375.
32. Pös,O., Radvanszky,J., Styk,J., Pös,Z., Buglyó,G., Kajsik,M., Budis,J., Nagy,B. and Szemes,T. (2021) Copy number variation: methods and clinical applications. *Appl. Sci.*, **11**.
33. Kleinert,P. and Kircher,M. (2022) A framework to score the effects of structural variants in health and disease. *Genome Res.*, **32**, 766–777.
34. Althagafi,A., Alsubaie,L., Kathiresan,N., Mineta,K., Aloraini,T., Al Mutairi,F., Alfadhel,M., Gojobori,T., Alfares,A. and Hoehndorf,R. (2022) DeepSVP: integration of genotype and phenotype for structural variant prioritization using deep learning. *Bioinformatics*, **38**, 1677–1684.
35. Sharo,A.G., Hu,Z., Sunyaev,S.R. and Brenner,S.E. (2022) StrVCTVRE: a supervised learning method to predict the pathogenicity of human genome structural variants. *Am. J. Hum. Genet.*, **109**, 195–209.
36. Danis,D., Jacobsen,J.O.B., Balachandran,P., Zhu,Q., Yilmaz,F., Reese,J., Haimel,M., Lyon,G.J., Helbig,I., Mungall,C.J. *et al.* (2022) SvAnna: efficient and accurate pathogenicity prediction of coding and regulatory structural variants in long-read genome sequencing. *Genome Med*, **14**, 44.
37. Yokoyama,T.T. and Kasahara,M. (2020) Visualization tools for human structural variations identified by whole-genome sequencing. *J. Hum. Genet.*, **65**, 49–60.
38. Thorvaldsdóttir,H., Robinson,J.T. and Mesirov,J.P. (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192.
39. Nurk,S., Koren,S., Rhie,A., Rautiainen,M., Bzikadze,A.V., Mikheenko,A., Vollger,M.R., Altemose,N., Uralsky,L., Gershman,A. *et al.* (2022) The complete sequence of a human genome. *Science*, **376**, 44–53.
40. Parveen,A., Khurana,S. and Kumar,A. (2019) Overview of genomic tools for circular visualization in the next-generation genomic sequencing era. *Curr. Genomics*, **20**, 90–99.
41. Greene,D., Consortium,G.E.R., Pirri,D., Frudd,K., Sackey,E., Al-Owain,M., Giese,A.P.J., Ramzan,K., Riaz,S., Yamanaka,I. *et al.* (2023) Genetic association analysis of 77,539 genomes reveals rare disease etiologies. *Nat. Med.*, **29**, 679–688.
42. Lek,M., Karczewski,K.J., Minikel,E.V., Samocha,K.E., Banks,E., Fennell,T., O'Donnell-Luria,A.H., Ware,J.S., Hill,A.J., Cummings,B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.