TAD-GConTool and **CNV-ConTool**
to assist prediction of phenotypic outcome of chromosomal rearrangements

Fino J[1], David D[1]

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

Try our tools at
http://dgrctools-insa.min-saude.pt/
or use our QRCode!

- Introduction -

With the advance of genome sequencing technologies, it is currently possible to identify a large number of chromosomal or genomic structural variants in a single individual [1]. Therefore, the validation and manual assessment of structural variants clinical significance becomes unpractical and time consuming when performed with previous methodologies.
In order to assist the validation process, we developed two clinically inspired bioinformatics tools – **TAD-GConTool** and **CNV-ConTool.** They were developed in *python* with a *Common Gateway Interface* that allows easy and user-friendly access through any standards compliant web browser [2].

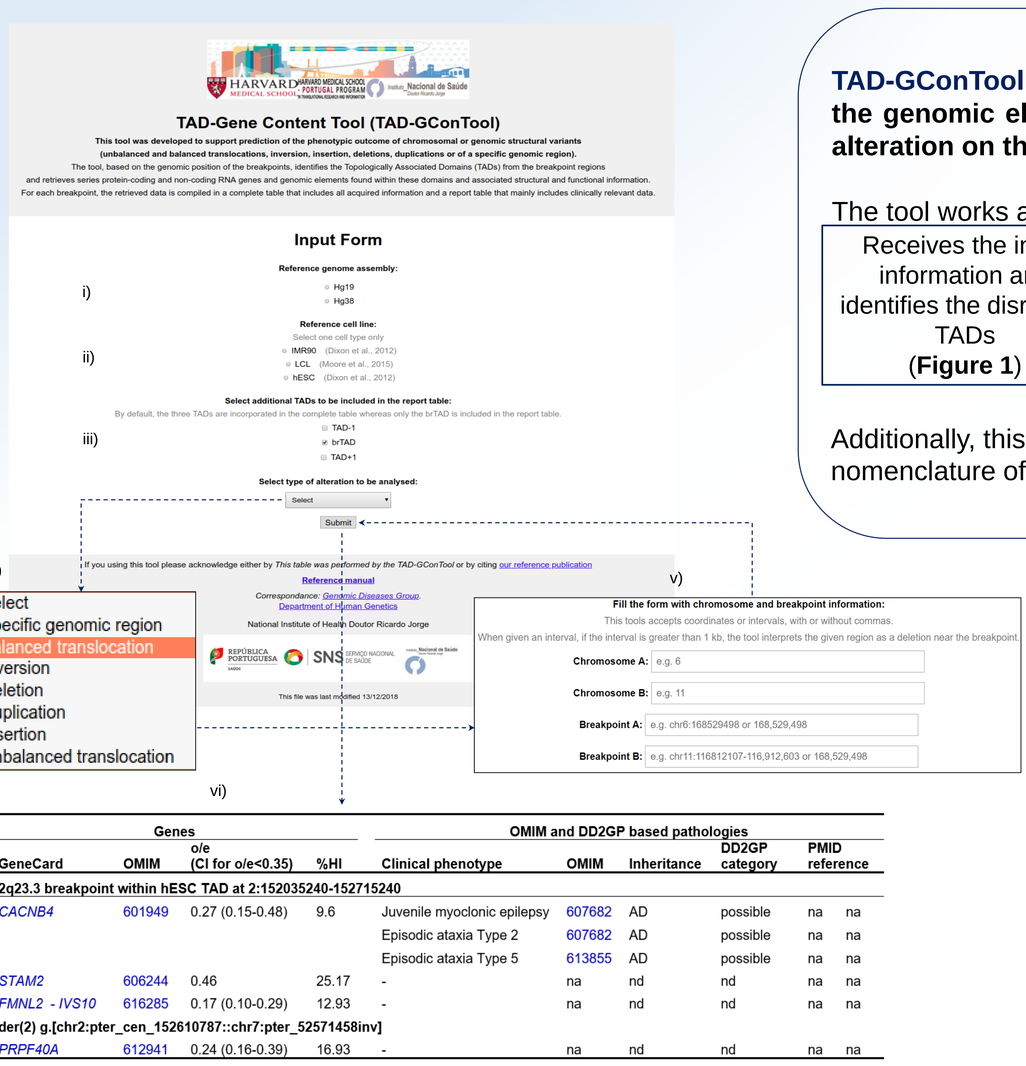
- TAD-GConTool -

TAD-GConTool allows the user to input a CNV or rearrangement and get a set of relevant and organized information about the genomic elements located on the affected Topological Associated Domains (TADs), to evaluate the importance of the alteration on the case in study.

The tool works as follows:

Receives the input information and identifies the disrupted TADs (**Figure 1**) → Searches genomic elements located inside the reference TADs using Ensembl API → Retrieves associated information from OMIM, GNOMAD, DECIPHER, MGI, DDG2P, among others → Organizes the information into a condensed table (report table) (**Figure 1**) and into an extensive table (complete table) → Retrieves the tables in a xlsx file for download

Additionally, this tool, based on the genomic position of the breakpoints and affected chromosomes, establishes the sequence based nomenclature of the rearrangement according to the International System for Human Cytogenomic Nomenclature 2016 (ISCN 2016)



Genes					OMIM and DD2GP based pathologies					
GeneCard	OMIM	o/e (CI for o/e<0.35)	%HI		Clinical phenotype	OMIM	Inheritance	DD2GP category	PMID reference	
2q23.3 breakpoint within hESC TAD at 2:152035240-152715240										
CACNB4	601949	0.27 (0.15-0.48)	9.6		Juvenile myoclonic epilepsy	607682	AD	possible	na	na
					Episodic ataxia Type 2	607682	AD	possible	na	na
					Episodic ataxia Type 5	613855	AD	possible	na	na
STAM2	606244	0.46	25.17		-	na	nd	nd	na	na
FMNL2 - IVS10	616285	0.17 (0.10-0.29)	12.93		-	na	nd	nd	na	na
der(2) g.[chr2:pter_cen_152610787::chr7:pter_52571458inv]										
PRPF40A	612941	0.24 (0.16-0.39)	16.93		-	na	nd	nd	na	na

Figure 1 Landing page of the TAD-GConTool and output example
The necessary input parameters must be selected as follows:
i) genome version; ii) reference cell line for TAD selection; iii) additional adjacent TADs (TAD-1 and TAD+1) to be included in the report table; iv) within the drop-down list, the alteration type to be analyzed; and v) data defining the SV (chromosomes and breakpoints information).
Finally, pressing the "Submit" button will launch the tool.
After running, this tool creates two output files, the vi) report table with more important condensed information and a complete table with all information found (not showed).

- CNV-ConTool -

CNV-ConTool was essentially developed to allow a faster and more informed evaluation of the CNVs or breakpoints, not by genomic content like TAD-Gene Content tool, but by comparison and overlap with benign and pathogenic CNV/Inversion public databases.

The tool works as follows:

Receives input information and identifies the genes affected by the given CNV/breakpoint, using Ensembl API(**Figure 2**) → Performs the overlap of the breakpoints/CNVs/ Inversions and affected genes vs databases (**Figure 3**) → Organizes the information into a summary table of the overlap search, with all inputted CNVs/breakpoints, affected genes and databases in the same table and extensive tables with all database and overlap information, organized by database. → Retrieves the tables in a xlsx file for download

The availability of different overlap strategies and parameters allows the user to tweak the tool as desirable.
The overlap search of the affected genes provides extra information for the CNV and breakpoint evaluation.

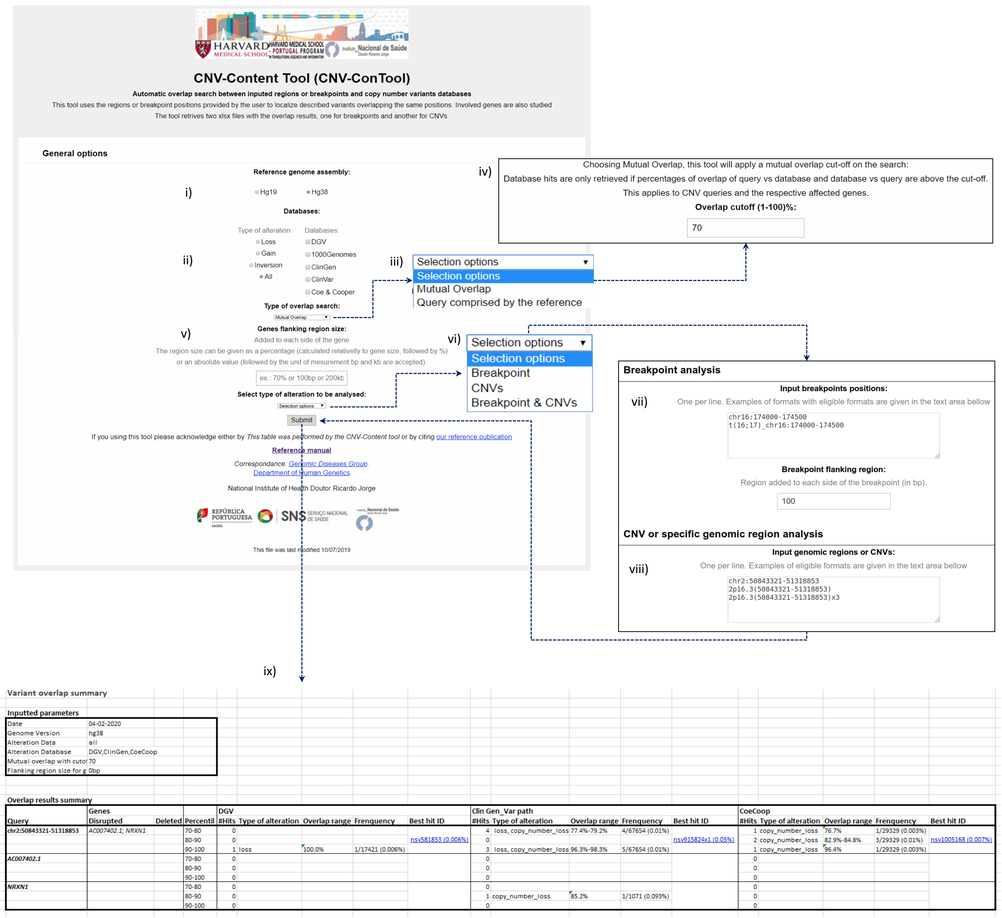


Figure 2 Landing page of CNV-ConTool and output example
Input parameters and query regions must be specified as follows:
i) genome version; ii) type of alteration to search against and reference databases; iii) type of overlap search: mutual overlap or query comprised by the reference (**Figure 3**). iv) overlap cutoff, available when mutual overlap is chosen; v) flanking gene regions to be included in the overlap search (the size of these can be given as an absolute value in bp or kb, or as a relative value, a percentage calculated relative to the size of each gene) – optional; vi) select the type of analysis (breakpoints, CNVs or both); vii) breakpoint analysis - input the breakpoint positions and the size of the breakpoint flanking region to be analyzed; definition of a region size to be added to each side of the breakpoint – optional; and viii) CNV or specific genomic region analysis - input the genomic positions of CNVs or of specific genomic regions to be analyzed.
Pressing "Submit" button will launch the tool.
The output file is composed by a ix) summary table of the overlap search and individual tables with the extensive overlap results, by database (now showed).

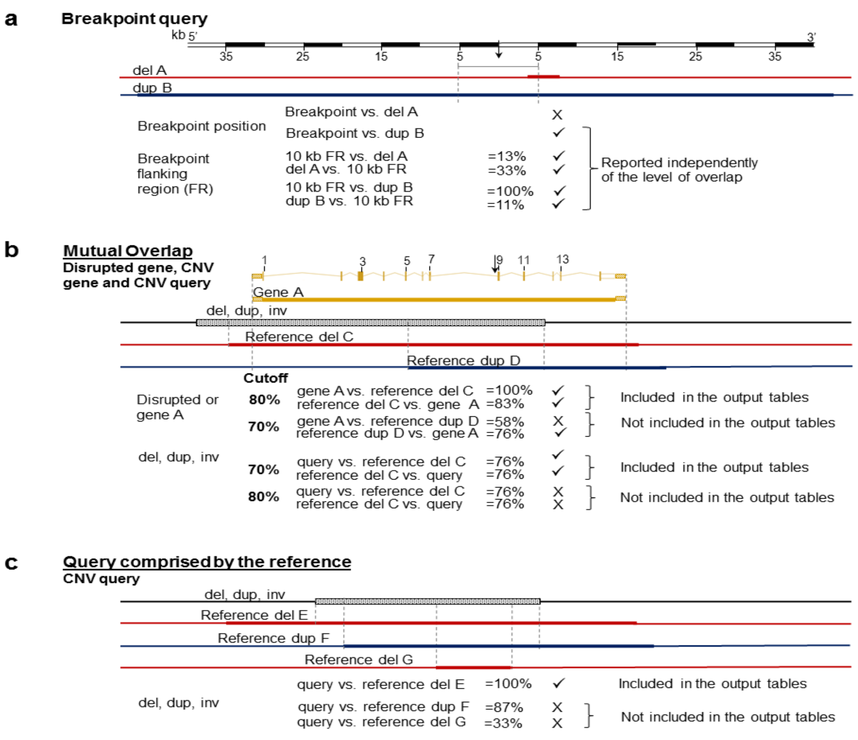


Figure 3 Schematization of overlap strategies available on CNV-ConTool
Vertical arrow indicates the breakpoint position. Reference deletions are in red whereas duplications are in blue. Gray dashed vertical lines outline the overlap between the query and reference alterations.
a Breakpoint genomic position and breakpoint-flanking region. All CNVs overlapping the query genomic position or breakpoint-flanking region are retrieved.
b Mutual overlap approach applied in CNVs, inversions and affected genes. Affected genes by either CNVs or breakpoints are depicted in yellow. Additional gene flanking genomic sequences included in the overlap search are depicted by shaded regions flanking the query genes. The query CNV is illustrated by the dashed back bar. In this approach, the percentage of overlap between query vs. reference and reference vs. query is computed, which ensures that, these show approximately the same size, and can be considered as similar. Only alterations with overlaps above the cutoff value (defined by the user), are included in the output tables.
c **Query comprised by the reference search for deletions, duplications and inversions**. In this approach, only the query vs. reference overlap percentage is taken into consideration: reference CNVs that cover 100% of the query are included in the output tables.

- Application & Conclusions-

These tools were already successfully applied to about 40 cases studied under the project "Next-gen cytogenetics enters clinical care and annotates the human genome" (HMSP-ICT/0016/2013) and are now being made available to the broader scientific community.
These tools allowed a faster and more informed evaluation of the genomic structural variants, helping select potential pathogenic variants, either by identifying phenotype-associated genes, or by overlapping deletions and duplications with already described benign or pathogenic CNVs [2].
As genome sequencing is becoming more and more a routine method for identification of chromosomal and genomic structural variants, such clinically oriented bioinformatics tools are crucial and represent the first level of analysis toward personalized genomic medicine.

- References -

[1] Collins *et al*., Genome Biology, 18.1 (2017)
[2] David *et al*., Human Genetics (2020)

Study supported by Fundação para a Ciência e Tecnologia project HMSP-ICT/0016/2013.