A crowdsourcing database for the copy-number variation of the Spanish population

Daniel López-López[1,2,3], Gema Roldán[1], Rosario Carmona[1,3], Gerrit Bostelmann[1], Jose Luis Fernandez-Rueda[1], Virginia Aquino[1], María Peña-Chilet[1,2,3], Rubén García[1], Rocío Nuñez[4], Guillermo Pita[4], Anna Gonzalez[4] and Joaquín Dopazo[1,2,3,5]

[1]Computational Medicine Platform, Andalusian Public Foundation Progress and Health (FPS), [2]Computational Systems Medicine, Institute of Biomedicine of Seville (IBiS), [3]Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), [4]Human Genotyping Unit, Spanish National Cancer Research Centre (CNIO), [5]FPS-ELIXIR-ES, Andalusian Public Foundation Progress and Health (FPS)

Motivation

Despite being a very common type of genetic variation, the distribution of copy-number variations in the population is still poorly understood. The knowledge of the genetic variability, especially at the level of the **local population**, is a critical factor for distinguishing pathogenic from non-pathogenic variation in the discovery of new disease variants.

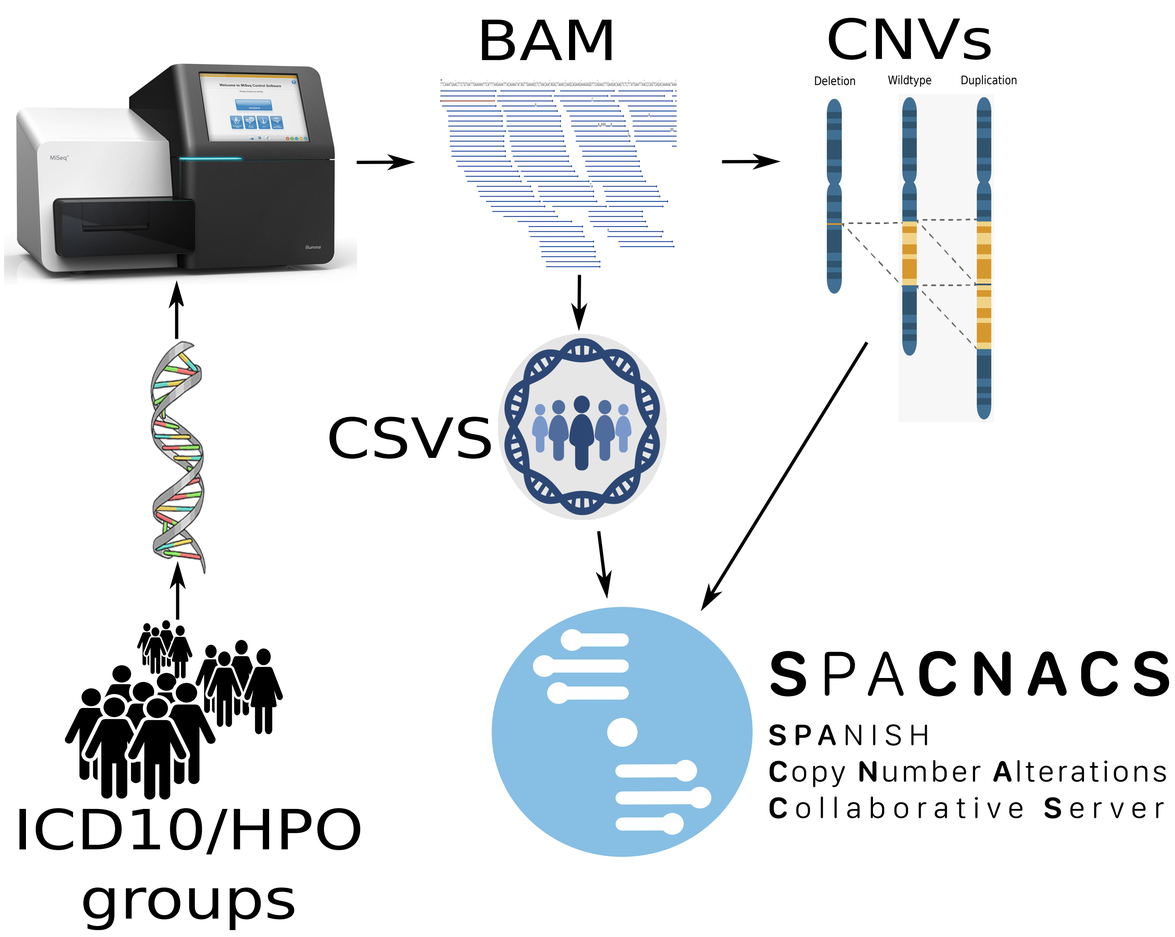
Results

Here, we present the SPAnish Copy Number Alterations Collaborative Server (**SPACNACS**), which currently contains:

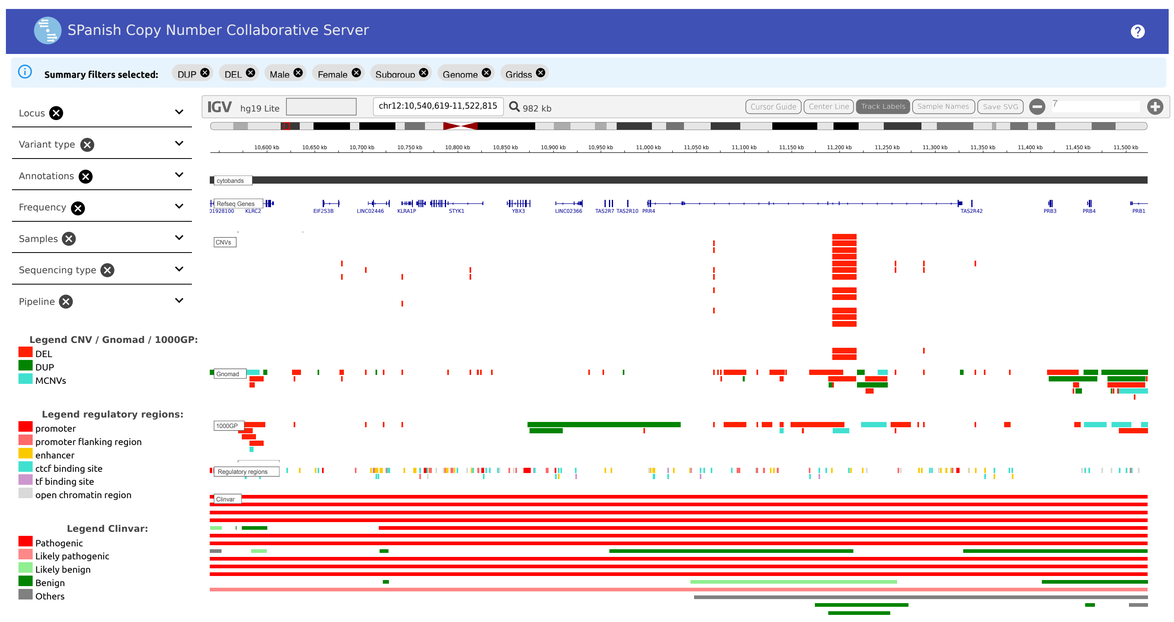
- Copy number variation (**CNVs**) profiles obtained from more than 400 genomes and exomes
- Sample **locality** and potential **kinship** testing
- Extensive **annotation** of CNVs with clinically relevant databases
- Samples binned by the top levels of **ICD10** and Human Phenotype Ontology (**HPO**) codes
- A **web** interface to query and filter results

http://csvs.clinbioinfosspa.es/spacnacs

Method



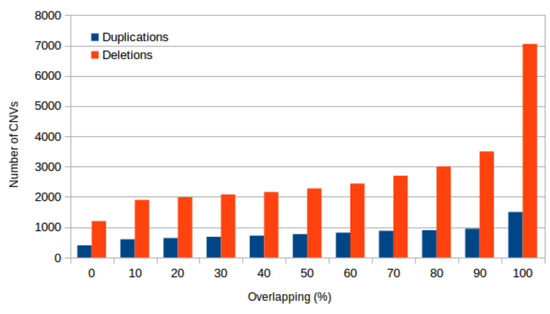
Web interface



(**A**) Genome browser panel consisting of an embedded Integrative Genomics Viewer preloaded with the Spanish CNV database and other useful tracks. (**B**) Search and selection panel, which provides several filters for specifying the genomic region and the data to be shown. (**C**) Filtering status panel, which shows information about the active filters.

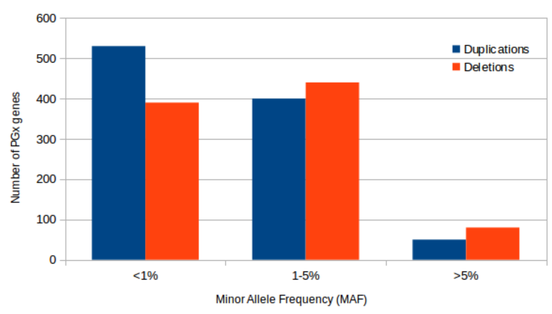
Comparison with other databases

Most SPACNACS CNVs overlap to a greater or lesser extent with some CNVs from both, 1000 genomes or Gnomad projects. Interestingly, a remarkable amount of SPACNACS CNVs do not overlap with any other CNV, which would play a crucial role in CNVs **prioritization** processes.



Case study

A 5,58% of the **pharmacogenomic** relevant genes reported in the clinical annotations from PharmGKB database, showed a minor allele frequency (MAF) higher than 5% in SPACNACS, suggesting a relevant role of this kind of variation in the pharmacogenomics field.



Beacon

SPACNACS implements a Beacon (version 1.0), a standard protocol used to query the database to check whether a specific region is involved in a CNV. The Beacon is an initiative of the Global Alliance for Genomics and Health (GA4GH) that allows genomic **data sharing** across federated networks.

Junta de Andalucía | Consejería de Salud y Consumo | Fundación Progreso y Salud
IBIS INSTITUTO DE BIOMEDICINA DE SEVILLA
ciberer Centro de Investigación Biomédica en Red Enfermedades Raras
Gobierno de España | Ministerio de Ciencia e Innovación
Instituto de Salud Carlos III
European Union
European Regional Development Fund