ELIXIR-hCNV

Implementation Study kick-off workshop

ELIXIR All-Hands 2019 Lisbon Portugal

www.elixir-europe.org

h-CNV workshop Agenda

- **The ELIXIR h-CNV community** – Christophe Béroud
- **WP1: Optimal CNV detection pipelines for research and diagnostics** – David Salgado
- **WP2: Definition of reference datasets** - Leslie Matalonga
- **WP3: Improvement of community formats for CNV exchange** – Cristina Yenyxe Gonzalez
- **WP4: Enabling CNV data discovery in diagnostic and phenotypic context** – Marc Hanauer and Michael Baudis
- **WP5: Creation of innovative tools** – Christophe Béroud
- **WP6: FAIRification of h-CNV databases and datasets** – Christophe Béroud
- **WP7: Dissemination -** Victoria Dominguez del Angel
- **Collaborative Discussion: human Copy Number Variation priorities, bottlenecks, …** - David Salgado & Gary Saunders
- **Workshop round up: summary, actions, and next steps** – David Salgado & Gary Saunders

H-CNV introduction

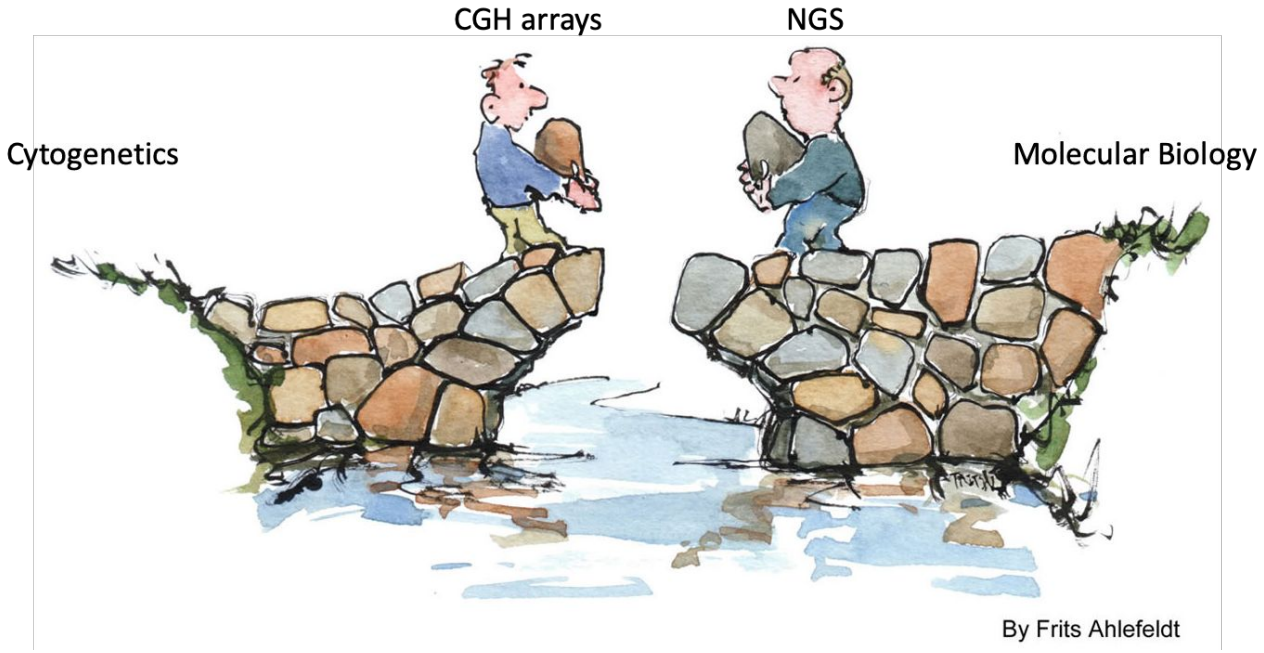
- **Structural variants have been the first ones to be detected in humans** *(late 1950s)*
- **Genes' mutations shortly followed** *(Ingram et al. 1957)*

Cytogenetics

Molecular Biology

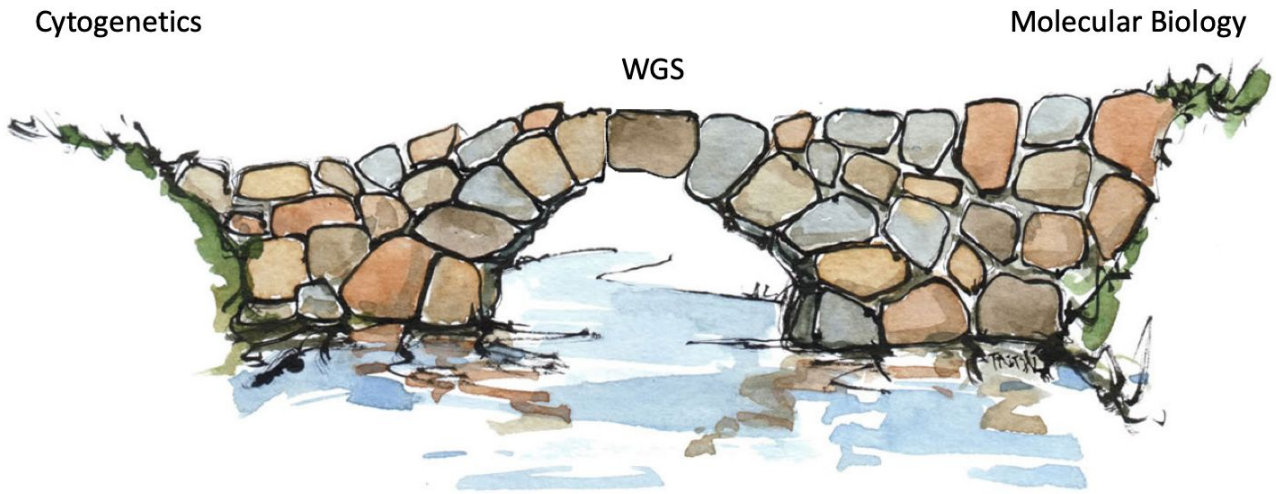
H-CNV introduction

- **Structural variants have been the first ones to be detected in humans** *(late 1950s)*
- **Genes' mutations shortly followed** *(Ingram et al. 1957)*



H-CNV introduction

- **Structural variants have been the first ones to be detected in humans** *(late 1950s)*
- **Genes' mutations shortly followed** *(Ingram et al. 1957)*



H-CNV introduction

- **Structural variants have been the first ones to be detected in humans**
- **Genes' mutations shortly followed**
- **CNV (Copy Number Variation) are the most frequent genetic cause of diseases**
  - Submicroscopic aberration in 1:270 fetuses *(general population - Srebniaket al. 2017)*
  - Account for 5 to 60% of mutations in RD genes
  - LOH (Loss Of Heterozygosity) somatic mutations in Oncogenetics
  - Involvement in common diseases (obesity, cardiovascular diseases …)
- **Detection, annotation and interpretation is not as efficient as for SNV**
  - High needs from the research and clinics for improvements

Current achievements

- 1st community meeting held in September 2018 - Hinxton
- F1000 community paper – needs to be adapted and resubmitted as a full Article based on HoNs comments
- h-CNV community approved in February 2019
- Implementation study accepted for a start 1st of June 2019
- Grant proposal submitted (*AMU, EJP-RD*, INSERM transversal project)
- Organized the HGVS meeting – Human CNV – June 14th – Goteborg Sweden
  - A great success!! Thanks again to all the members who contributed !!!
- 1st community workshop ELIXIR All-Hands Lisbon
- Regular participation to the HDC conference calls
- ELIXIR h-CNV community webpage - https://elixir-europe.org/communities/hcnv
- h-CNV GitHub repository – for visibility, communication and sharing https://github.com/hcnv + dedicated community website https://hcnv.github.io

WP1: Optimal CNV detection pipelines for research and diagnostics

Lead	Alfonso Valencia, Salvador Capella (**ELIXIR-ES**)
Members	**CH** (Michael Baudis), **DE** (Jan Korbel), **ES** (Joaquin Dopazo, Steven Laurie, Gemma Bullich), **FR** (Christophe Béroud and David Salgado), **HU** (Attila Gyenesei), **NL** (Bauke Ylstra and Daoud Sie, Leon Mei, Lennart Johansson), **SI** (Brane Leskošek, Polonca Ferk, Marko Vidak), **UK** (Krzysztof Poterlowicz),
Delivery	**M1-M18**

WP1: Main objectives

- CNVs can be detected through multiple technologies (aCGH, MLPA, WGS, WES, panel,…)
- CNVs are of different nature (DEL / DUP / INS / Tandem Repeats)
- A lot of software and pipelines have been released to facilitate detection
- Problem:
  - Sequencing methods can in theory detect any type of CNVs, no single computational algorithm can accurately and sensitively detect all types and all size CNVs
  - Low consensus among tools calling CNVs specifically for WES/WGS (Zare et al BMC Bioinformatics, 2016, Yao et al, Molecular Cytogenetics, 2017, Zook et al, BiorXiv 2019) and panels
- Solution:
  - Provide sensitive, reliable, optimized and validated pipelines for different type of CNVs and from different experimental data/technologies
  - A benchmark infrastructure to evaluate efficiency of new developed algorithms

WP1: Milestones & Deliverables

Milestones:

- **M1.1** Evaluation of available systems to detect CNV and documentation in ELIXIR Bio.tools
- **M1.2** Installation of systems to be benchmarked within ELIXIR compute platform and OpenEbench, work on metrics to be collected and compared
- **M1.3** Proceed to benchmark and provide results
- **M1.4** Benchmark systems to detect CNV for diagnostic (ISO) requirements.

Deliverables:

- **D1.1** Deliver the list of available pipelines/software as well as partners' local solutions to detect CNV from gene panels, WES, WGS, array CGH and SNP arrays.
- **D1.2** Develop a generic benchmarking platform to evaluate new tools and new datasets
- **D1.3** Benchmark the various systems using datasets from WP2 to select the most sensitive, specific, reliable and rapid systems for each dataset for germline and somatic CNVs.
- **D1.4** Deliver optimized pipelines from D1.3 to increase performance on ELIXIR compute nodes and define optimal parameters and guidelines to help end-users to efficiently and reliably detect CNV in various situations through the ELIXIR training platform.
- **D1.5** Genomic external quality assessments (EQAs) for CNV patient sample data, germline and somatic.

WP1: Requirements

- Identify CNV detection tools (cf CNV software catalogue)
    - Gather some specific details (made for deletion/duplication, algorithm type (read depth, split reads, assembly, …), technology (short/long reads, single cell, array, panel, exome, genome))
- Prepare these tools to be easily installed within compute platforms (biocontainers, bioconda, docker,…)
- Prepare the benchmarking infrastructure (OpenEbench, gold standard dataset (WP2), metrics to be collected)

Remarks:

- Need to interact with **Compute** / **Tool** (bio.tools) / **Data** (curation of published CNV tools) and **Galaxy** community (e.g. collaborative benchmarking)
- Strong links with h-CNV WP2: reference datasets
- Various recent initiatives are currently benchmarking SV (**only WGS**)
    - Zook et al, bioRxiv June, 2019 (provide datasets and comparisons metrics)
    - Kosugi et al, Genome Biology, 2019 (provide real and simulated datasets and comparisons metrics)

WP2: Definition of reference datasets

Lead	Steven Laurie (ELIXIR-ES)
Members	DE (Jan Korbel), EMBL-EBI (Thomas Keane and Cristina Y. Gonzalez), FR (Christophe Béroud and David Salgado), SI (Marko Vidak)
Delivery	M1-M15

WP2: Main objectives

- Problem:
  - No gold standards to evaluate CNV software efficiency
  - Some efforts are starting to emerge such as GIAB-SV but focus on one type of technology and nature (mainly WGS and not including WES technologies)

- Goals:
  - In coordination with international initiatives (NIST, GIAB, GA4GH, …), WP2 will attempt to define reference datasets for the various technologies available for CNV detection.
  - This effort will focus on WGS, WES and targeted sequencing, and on various sequencing technologies (short/long reads) and CNV types (DEL/INS… Germline/somatic/*de novo*…)

WP2: Challenges

- Challenges and points to be discussed:
  - Ambitious IS as it involves several use cases (germline and somatic analysis)and also different types of NGS technologies (WES, WGS, short and long reads)
  - WES CNV analysis requires large control datasets (most of the current tools requires a batch of control samples for data comparison on CNV calls)
  - For creating new robust datasets we need human genomic DNA (cell lines are not a realistic solution - previous experience): ethical issues (consent) and laboratory to sequence those (enough DNA to seq with different technologies) (collaboration with other initiatives?)
  - Possible practical solutions
  - Define the requirements and needs of each specific use case and focus on one specific approach (e.g. germline WES short reads)
  - For WGS data align with GIAB efforts and SV workstream (Ashkenazi Jewish trio). (https://www.biorxiv.org/content/10.1101/664623v2)
  - We could use DNA from a non-human model organism to test programs, but unlikely to be equivalent to human samples, so only as a last option

WP2: Milestones & Deliverables

Milestones:

- **M2.1** List of reference datasets for NGS (M3)
- **M2.2** Updated list of reference datasets for NGS (M15)

Deliverables:

- **D2.1** Deliver WES reference datasets (M6, M15)
- **D2.2** Deliver WGS reference datasets (M6, M15)

WP2: Requirements

- Work with the **Human data** and the **Rare Disease** communities to identify which data could be of interest (known/confirmed CNV in patients), **Coriell**, **BBMRI, …**
    - Try to identify other projects that have generated large amounts of WES data that could be used for our purposes
- Need to take into account GDPR impact (ELSI, consent,…)
    - Involvement of ELSI specialists, lawyers, true anonymisation
- Need to generate sequencing data for various sequencing kits/technologies
    - Long/short reads, single cell, distinct exome kits (up-to-date, and most popular)
    - Need to keep these "Gold standard" datasets "up to date", evolving technologies, release of new kits…
- Need to interact with ELIXIR **Data platform**

WP3: Improvement of community formats for CNV exchange

Lead	Thomas Keane (**EMBL-EBI**)
Members	**CH** (Michael Baudis), **FR** (Marc Hanauer), **EMBL-EBI** (Cristina Y. Gonzalez), **SI** (Brane Leskošek).
Delivery	M1-M12

WP3: Main objectives

- International collaborative projects require harmonization and standardization of results in order to ensure efficient data aggregation and comparison
- The GA4GH Genomic Knowledge Standards and Large-Scale Genomics groups, are currently globally addressing aspects of this issue
- Problem:
  - No robust and exhaustive standard CNV annotation format has emerged so far
  - Cytogeneticists / molecular biologists and other users communities use different formalisms
- Solution:
  - Adapt existing formats to handle proper CNVs description.
  - Improve VCF standard to properly describe CNVs or adopt future formats as replacement

WP3: Milestones & Deliverables

Milestones:

- **M3.1** Perform survey to identify community use of file formats and storage formats for representation and exchange of CNV data

Deliverables:

- **D3.1** Catalog of identified issues and limitations of file formats and data schemas for representing and exchanging CNV data
- **D3.2** Create consensus collection of perceived requirements for efficient and effective CNV file and data exchange formats
- **D3.3** Propose and communicate specification changes to pre-existing formats (e.g. VCF) to address the highest priority limitations
- Participate in discussions about future formats that could replace VCF

WP3: Requirements

- Work with the **Human Data** community to understand what limitations do they currently face when representing which types of variation.
- Datasets identified in WP2 should be converted to the format recommended in WP3.
- The recommended formats must be (based on) existing ones, to facilitate adoption by the community.
- Efforts to improve formats must be aligned with those from the Global Alliance for Genomics and Health (GA4GH).

WP4: Enabling CNV data discovery in diagnostic and phenotypic context

Lead	Michael Baudis (**ELIXIR-CH**), Marc Hanauer (**ELIXIR-FR**)
Members	**DE** (Jan Korbel), **EMBL-EBI** (Denise Carvalho-Silva), **ELIXIR-Hub** (Gary Saunders, Leyla Garcia), **ES** (Alfonso Valencia, Salvador Capella, Joaquin Dopazo, Laura I. Furlong, Janet Piñero, Steven Laurie, Gemma Bullich), **FR** (Christophe Béroud and David Salgado, Marc Hanauer), **HU** (Katalin Monostory), **NL** (Bauke Ylstra), **SI** (Marko Vidak, Polonca Ferk), **UK** (Krzysztof Poterlowicz)
Delivery	**M1-M12**

WP4 Main objectives: select ontologies and relevant common data elements

- **Select ontologies required to efficiently capture phenotypic description useful for data interpretation and matchmaking**

- **Provide lists of common data elements that should be used in various situations**

  - Enabling CNV data discovery in clinical context (diagnosis and phenotypic knowledge)

  - Finding similar cases across databases

  - Identify diseases-causing genes and improve genotype/phenotype correlations knowledge

WP4: Deliverables and Milestones

- **D4.1** Deliver a list of select ontologies required to efficiently capture phenotypic description useful for data interpretation for any genetic disease. (**M6**)

- **D4.2** Deliver lists of common data elements that should be provided in various situations such as rare diseases, oncology or common diseases. (**M6**)

- **M4.1** List of selected ontologies to capture phenotypic description of patients and samples (**M12**)

WP4: Activities

- **Benchmark of available standards & existing workflows**

Based on datasets and data exchanges, taking account of FAIR principles and approaches, we will explore the accuracy of the existing standards regarding the CNV (WP6)

- **Stakeholders involvement to determine accurate Common Data Elements**

Based on a community approach, we will determine the minimum valuable data model needed. We will work closely with WP2 & WP3

WP4: Projects crosslinking & ontologies

- **GA4GH BEACON and standardisation**
the Beacon API and planned ontology-based phenotype queries

- **HPO and Phenopackets**
Human Phenotype Ontology and new data model provided by Phenopacket to exchange phenotypic information

- **Orphanet Rare Diseases Ontology and HOOM** (HPO Orphanet Ontological Module)
Rare diseases dedicated nomenclature, with genes, enabling mappings across terminologies and 70k annotations with HPO

- **EJP-RD (European Joint Program on Rare diseases)**
Coordinated access to data and services for transformative rare diseases research.

NCIT, Uberon, ICD-O, others…
Part of the process will be to select relevant ontologies for HCNV field

WP5: Creation of innovative tools

Lead	Christophe Béroud and David Salgado (**ELIXIR-FR**), Joaquin Dopazo (**ELIXIR-ES**)
Members	**CH** (Michael Baudis), **DE** (Jan Korbel), **EMBL-EBI** (Sarah Hunt), **ES** (Laura I. Furlong, Alfonso Valencia, Salvador Capella), **HU** (Katalin Monostory), **NO** (Eivind Hovig, Pubudu Samarakoon), **UK** (Krzysztof Poterlowicz)
Delivery	**M1-M24**

WP5: Main objectives

- Problem:
  - CNVs can involve large genomic regions and affect multiple genes
  - With relevance to recessive diseases and tumor suppressor genes, CNVs can cooperate with other types of genomic or regulatory alterations affecting alleles of the same genetic target.
  - In situations where large CNV are involved, it is difficult to unanimously identify the specific gene(s) whose alterations are directly associated to the patient's phenotype.

- Solution:
  - The h-CNV community will develop innovative tools
    - Functional annotation of CNVs
    - Combinatorial approaches to CNV interpretation
    - Identification of landmark genes in regions of interest

WP5: Milestones & Deliverables

Milestones:

- **M5.1** Creation of innovative tools to facilitate CNV interpretation (M12, M24)

Deliverables:

- **D5.1** Deliver mandatory **CNV annotations including**: type; genotype; genes and transcripts; expression level; exons; regulatory elements; breakpoints/fusion fragments. (M6)
- **D5.2** Creation of a specific pipeline to **interpret duplications as tandem,** inverted or translocation duplications may result in very different phenotypes. (M12)
- **D5.3** Creation of specific bioinformatics tools to **select candidate genes** localized in the CNV region by combining genes' annotations and patients' phenotype. (M18)
- **D5.4** Deliver tools to **determine (tumor) heterogeneity and mosaicisms**. (M24)

WP5: Requirements

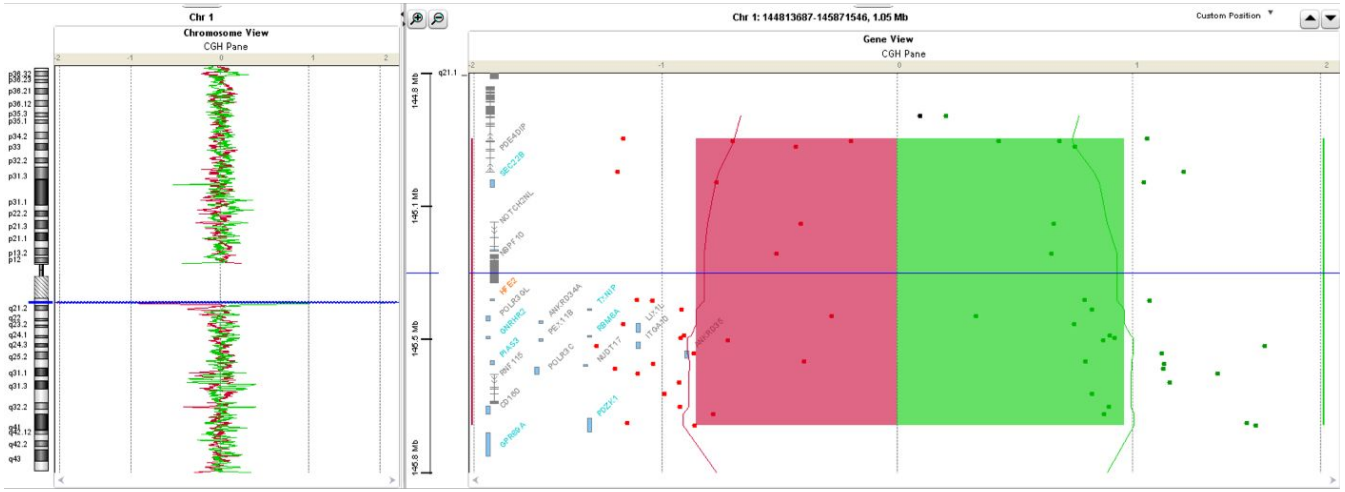
- Strong interaction with WP4 & WP6
- Strong interaction with various communities (FHC, RD, onco, …)
- Need to interact with **ELIXIR Beacon**
  - Should we develop a hCNV central beacon?

- Tools should be compatibles with data description (WP4) to ensure the easy implementation into the hCNV national databases from partners
- Tools should be made available to a wider community to ensure hCNV recognition
  - CNV translation
  - Beacon API

WP5: Requirements

- Make sure we adopt standards

1q21. deletion (TAR syndrome)



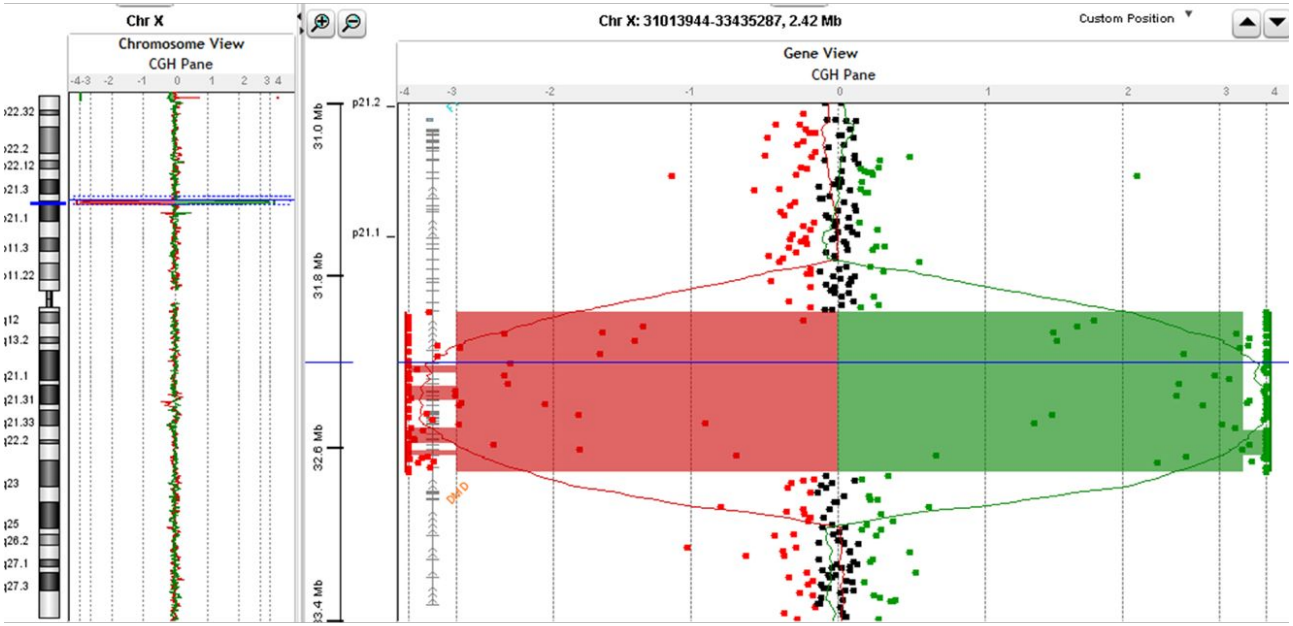
Radius malformation, thrombocytopenia: **Pathogenic**
and search for a second hit on the other allele

Isolated intellectual disability: **Benign**,
no further molecular studies at this locus

WP5: Requirements

- Make sure we adopt standards

Intragenic *DMD* deletion



Male: **pathogenic**
Female: **benign**, carrier

WP6: FAIRification of h-CNV databases and datasets

Lead	Christophe Béroud and David Salgado (**ELIXIR-FR**), Joaquin Dopazo (**ELIXIR-ES**)
Members	**CH** (Michael Baudis), **EMBL-EBI** (Cristina Y. Gonzalez), **FR** (Victoria Dominguez Del Angel), **HU** (Katalin Monostory), **NL** (Morris Swertz), **NO** (Eivind Hovig, Pubudu Samarakoon, Lennart Johansson), **UK** (Krzysztof Poterlowicz)
Delivery	**M1-M24**

WP6: Main objectives

- Problem:
  - Various national CNV databases, curated CNV data resources, and ELIXIR deposition databases are currently being developed by ELIXIR h-CNV partners
  - In order to allow interoperability (including resource and data discovery), the FAIR principles (Findable, Accessible, Interoperable, Reusable) will be applied to those systems to demonstrate the feasibility and utility of distributed CNV databases
  - This will respect databases' ownerships and national regulations' compliance while allowing searching for similar patients across the network
- Solution:
  - Implement WP4 recommendations and FAIR principles
  - Implement Beacons

WP6: Milestones & Deliverables

Milestones:

- **M6.1** Release of a FAIR CNV database. (M18)

Deliverables:

- **D6.1** FAIRification of the French **BANCCO** database (http://bancco.fr) developed at Aix Marseille University, the **CIBERER** (Spanish network for research in rare diseases) database developed at the Fundación Progreso y Salud of Sevilla and the **VKGL CNV** database (Dutch clinical genetics diagnostics) CNV database as prototypes to demonstrate the benefits of using the FAIR data principles for CNV in diagnostic and research contexts. (M18)
- **D6.2** Extension of the FAIRification to other non-specific CNV databases such as the **European Variation Archive** (EVA), **RD-CONNECT**, **arrayMap**, **Progenetix** and others. (M24)

WP6: Requirements

- Use **same CNV description**
  - Molecular level
  - Phenotypic level

ID	Chromosome	hg18	hg19	hg38	nature	Classification
104834	17	Début: 31889665 Fin: 33323678	Début: 34815552 Fin: 36249565	Début: 36459738 Fin: 37889943	Délétion	Bénin=100%
2430993	17	Début: 31889664 Fin: 33318471	Début: 34815551 Fin: 36244358	Début: 36459737 Fin: 37884738	Délétion	VUS=100%

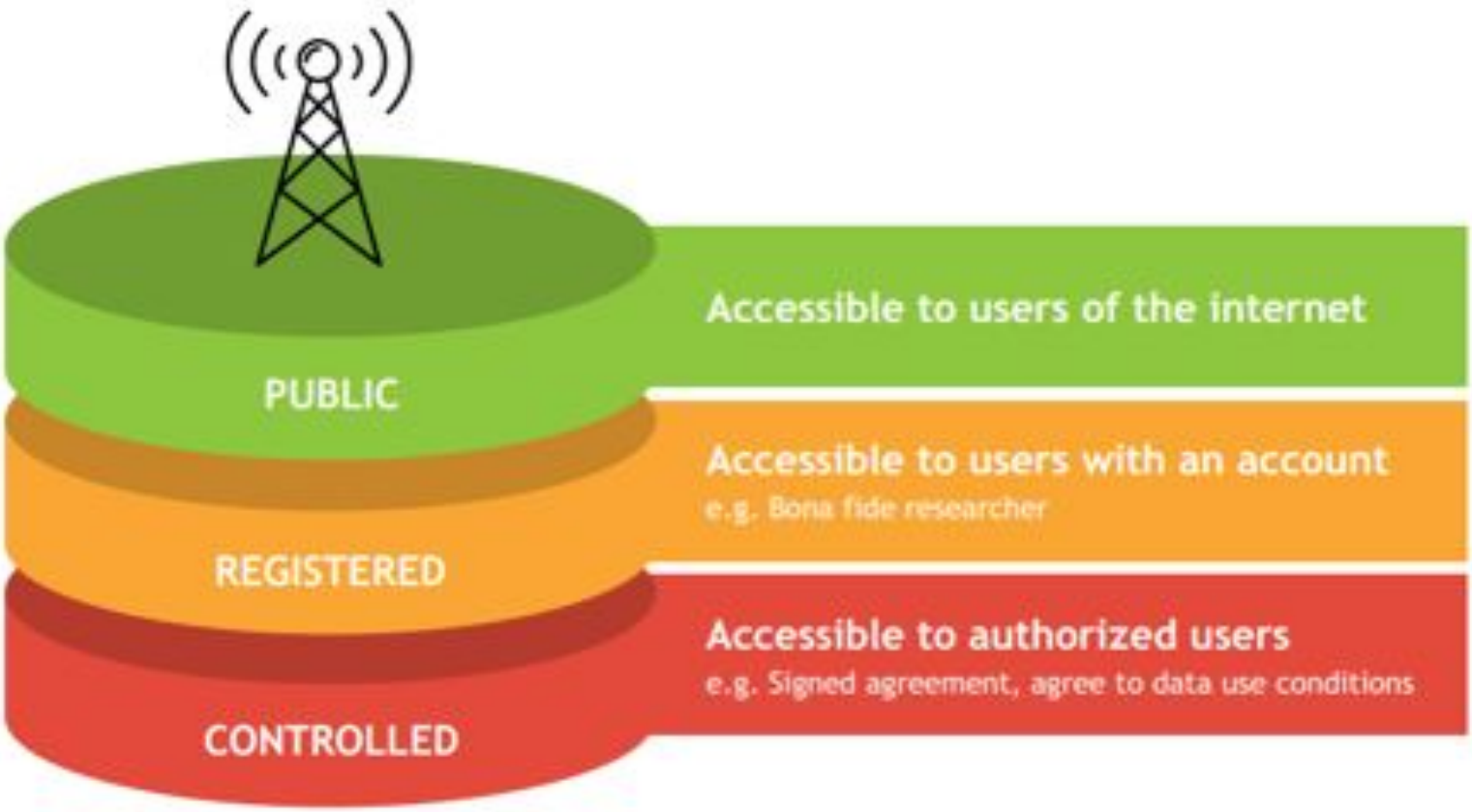
- Strong interaction with **WP4 & WP5**
- Strong interaction with **various projects** (EJP-RD, …)
- Need to interact with **ELIXIR Beacon**
- Need to take into account **data privacy**

Beacon

Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

elixir

WP6: Requirements



WP7: Dissemination

Lead	Michael Baudis (**ELIXIR-CH**), Victoria Dominguez Del Angel (**ELIXIR-FR**), Gary Saunders (**ELIXIR-Hub**)
Members	**ELIXIR-Hub** (Kathi Lauer), **ES** (Joaquin Dopazo), **HU** (Attila Gyenesei), **NO** (Eivind Hovig, Pubudu Samarakoon), **SI** (Brane Leskosek), **UK** (Krzysztof Poterlowicz)
Delivery	**M1-M24**

WP7: Main objectives

- Stakeholders (Bioinformaticians, biomedical communities and researchers)
- Delivering training and provide training materials based on global adoption of tools and guidelines
- Evaluation of tools and services

- Solution:
  - Detection the CNV (Bioinformaticians), TrD events)
  - Participation to meetings
  - Capacity-building training events :
    - Tools
    - Data FAIRification
    - Data Interpretation to access to knowledge base system

WP7: Milestones & Deliverables

Milestones:

- **M7.1** Participation to international meetings to promote the h-CNV community (M12, M24)
- **M7.2** Organisation and participation to international bio hackathons/Jamborees/Capacity building events dedicated to CNV (M18)

Deliverables:

- **D7.1** Creation of Jamborees to gather experts' point of view on the various objectives and related tasks and developments. (M12, M24) bioinformaticians and developers
- **D7.2** Creation of regular hackathons to ensure smooth developments and benchmarks by various ELIXIR Nodes. (M12, M24) Bioinformaticians and developers
- **D7.3** Promotion of the ELIXIR h-CNV community through participation to international meetings, such as GA4GH. A contact has already been established with the Human Genome Variation Society (HGVS). (M12, M24) Biomedical communities
- **D7.4** Creation of regular Capacity building events to ensure knowledge dissemination across ELIXIR Nodes, biomedical communities and Industry. (M12, M24)
- **D7.5** Set Up documentation about best practices, training material and guidelines on CNV interpretation (M18-M24) Biomedical communities
-

WP7: Milestones & Deliverables

Milestones:

- **M7.1** Participation to international meetings to promote the h-CNV community (M12, M24)
- **M7.2** Organisation and participation to international bio hackathons/Jamborees/Capacity building events dedicated to CNV (M18)

WP7: Milestones & Deliverables

Milestones:

- **M7.1** Participation to international meetings to promote the h-CNV community (M12, M24)
- **M7.2** Organisation and participation to international bio hackathons/Jamborees/Capacity building events dedicated to CNV (M18)
- Topic : Copy Number Variation Beacon (David, Gary and Michael)

  https://www.biohackathon-europe.org/

WP7: Milestones & Deliverables

Milestones:

- **M7.1** Participation to international meetings to promote the h-CNV community (M12, M24)
- **M7.2** Organisation and participation to international bio hackathons/Jamborees/Capacity building events dedicated to CNV (M18)

Deliverables:

- **D7.1** Creation of Jamborees to gather experts' point of view on the various objectives and related tasks and developments. (M12, M24)
- **D7.2** Creation of regular hackathons to ensure smooth developments and benchmarks by various ELIXIR Nodes. (M12, M24)
- **D7.3** Promotion of the ELIXIR h-CNV community through participation to international meetings, such as GA4GH. A contact has already been established with the Human Genome Variation Society (HGVS). (M12, M24)
- **D7.4** Creation of regular Capacity building events to ensure knowledge dissemination across ELIXIR Nodes. (M12, M24)
- **D7.5** Set Up documentation about best practices and guidelines on CNV interpretation (M18-M24)
-

WP7: Milestones & Deliverables

Milestones:

- **M7.1** Participation to international meetings to promote the h-CNV community (M12, M24)
- **M7.2** Organisation and participation to international bio hackathons/Jamborees/Capacity building events dedicated to CNV (M18)

Deliverables:

- **D7.1** Creation of Jamborees to gather experts' point of view on the various objectives and related tasks and developments. (M12, M24)
- **D7.2** Creation of regular hackathons to ensure smooth developments and benchmarks by various ELIXIR Nodes. (M12, M24)
- **D7.3** Promotion of the ELIXIR h-CNV community through participation to international meetings, such as GA4GH. A contact has already been established with the Human Genome Variation Society (HGVS). (M12, M24)
- **D7.4** Creation of regular Capacity building events to ensure knowledge dissemination across ELIXIR Nodes. (M12, M24)
- **D7.5** Set Up documentation about best practices and guidelines on CNV interpretation (M18-M24)
-

WP7: Requirements

- Strong interaction with all WP
- Interaction with **ELIXIR Training (TeSS)**
- Synergy with other **ELIXIR platforms (Data, Tools, Compute)**
- Need to reach external international initiatives and scientific societies (EJP-RD, HGVS, GA4GH,…)

Collaborative Discussion

Topics to discuss

- human Copy Number Variation priorities
- Possible bottlenecks?
- Needs for ELIXIR - how can we ensure that services are designed to serve end-users?

Topics to discuss

- How to reach ELIXIR-Platforms / communities
- We have needs that Platforms/communities can provide
- How to work together?

Workshop round up: summary, actions, and next steps

Workshop round up - Actions

- Regular h-CNV conference calls to follow up on WP advances and bottlenecks
  - *Regular timeslot need to be identified*
- Circulate the latest version of the F1000 whitepaper, with responses to comments from HoNs
- Work with ELIXIR platforms / communities to achieve our first objectives and work towards the new ones
- Work on the h-CNV proposal accepted to the next BioHackathon 2019 in Paris
  - *Main project is to work on beacon for CNV data repositories*
  - *What do we get at the end through this effort?*
- To do things we need resources: Identify new funding opportunities and apply individually and with partners (support letters…)
  - *We would like to apply for an EJP-RD proposal next year. We need to get prepared*

Thank you

www.elixir-europe.org

CNV software listing

- CNV detection software curation based on published articles
- Initial software list from the dbstar project (ELIXIR-Fr)
- Currently 245 software listed
- List of items collected :
  - Name of the resource,
  - Reference, Publication year,
  - recently cited (year + ref),
  - last update,
  - Running Operating system (Mac, Linux, Windows)
  - Programming language (R, Java, Python, C++, …)
  - Input format (BAM, BED, pileup, GTF, Counts)
  - Output (VCF, tabulated file, CSV)
  - detection methodology (HMM, Bin GC content, log coverage ratio, …)
  - test dataset (public dataset, simulated dataset)
  - Reference dataset requirement,
  - Experiment type compatibility (WGS, WES, panel, long/short reads, single cell)