

Piloting the implementation of genome-phenome analysis tools for Personalised Medicine

Sergi Beltran (CNAG-CRG)

Seminaris de Farmàcia
Barcelona, October 2nd 2018

cnag



```
O|O:123:123,123 O|O:123:123,123 O|1:123:123,123 O|1:49:52,
123,123 O|O:123:123,123 O|O:123:123,123 O|O:123:123,123 O
O|O:123:123,123 O|O:123:123,123 O|O:123:123,123 O|O:52:12
123,123 O|1:123:123,123
O|O:123:123,123 1|O:123:123,123:56;0.0852854;21;19 O
O|O:123:123,123 O|O:83:83,123 O|1:43:123,43 O|O:123:12
123,123 1|O:68:68,123 O|O:123:123,123 O|O:123:123,123 O
O|O:51:123,51 O|O:43:43,123 O|O:87:123,87 O|O:114:12
123,123 1|O:37:37,123 O|O:123:123,123 O|O:123:123,123 O
O|O:123:123,123 1|O:123:123,123
O|O:123:123,123 O|O:123:123,123:59;0.102882;5;3 O|O:113:1
123,123 O|O:123:123,123 O|O:123:123,123 O|O:76:105,76 O
O|1:123:123,123 O|O:76:76,123 O|O:123:123,123 O|O:123:1
123,123 O|O:123:123,123 O|O:123:123,123 1|O:123:123,123
O|O:123:123,123 1|O:123:123,123 O|1:106:123,106
:123,123 O|O:113:123,113
Q1,HQ2 O|O:123:1
```

```
ro@ns indelcalling]
syntax error
ro@ns indelcalling] $
--help' for more information.
ro@ns indelcalling] $ cp /scratch/devel/fcastro/data/1000genomes/indelcalling/CEU* .
ro@ns indelcalling] $
ro@ns indelcalling] $ cp /scratch/devel/fcastro/data/1000genomes/indelcalling/README_* .
ro@ns indelcalling] $ ls
SRP000031.2010_03.indels.genotypes.vcf.gz CEU.SRP000031.2010_03.indels.genotypes.vcf.gz.tbi CEU
ro@ns indelcalling] $ cp /scratch/devel/fcastro/data/1000genomes/indelcalling/CEU* .
ro@ns indelcalling] $ pwd
/devel/fcastro/COPY_temp/indelcalling
ro@ns indelcalling] $ cd /scratch/
```

CNAG Overview

The CNAG-CRG is a non-profit organization integrated in the CRG. It is funded by the Spanish Ministry of Economy and Competitiveness and the Catalan Government.

Our Mission

To carry out projects in genome analysis that will lead to significant improvements in people's health and quality of life, in collaboration with the Catalan, Spanish, European and International research and clinical community.

CNAG-CRG

Director: Ivo Gut

80 people, 60 bioinformaticians

ICTS – Infrastructure for OMICS Technologies (IOT)

The IOT includes the CNAG-CRG and the COS (Centre for Omics Sciences).



cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico



CNAG-CRG: Relevant partnerships

- ✓ Member of the **Global Alliance for Genomics and Health (GA4GH)**
- ✓ Participation, through the National Bioinformatics Institute (INB), in ELIXIR, the European bioinformatics infrastructure.
- ✓ Participation in the International Human Epigenome Consortium (IHEC)
- ✓ Participation in the International Cancer Genome Consortium (ICGC)
- ✓ Participation in the International Rare Diseases Research Consortium (IRDiRC)



The CNAG's Genomehenge 2018



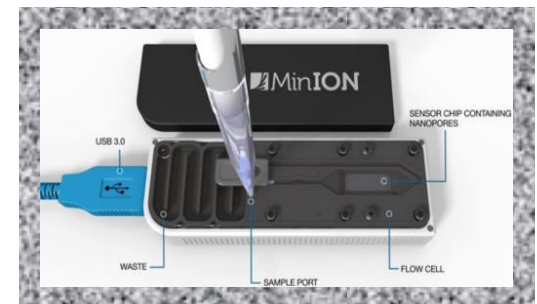
Sequencing

- 1 Illumina NovaSeq6000
- 2 Illumina HiSeq4000
- 4 Illumina HiSeq2000
- 1 Illumina MiSeq
- 3 Oxford Nanopores MinIons



Computing

- 3852 cores
- 3.7 PB disk + 3 PB tape
- 35,5 TB RAM
- Barcelona SuperComputing Center - 10 x 10 Gb/s



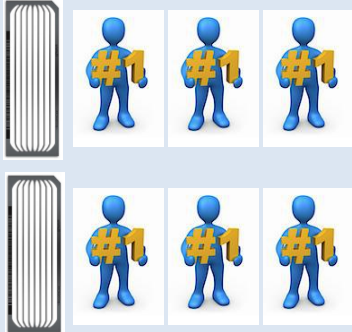
Sequencing Production

MiSeq



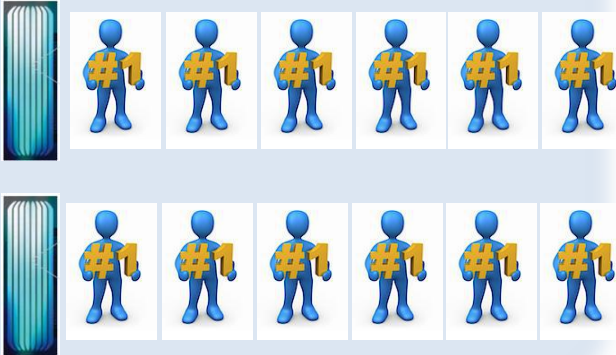
4.5x Coverage
2x300bp
50 hours

HiSeq2000



30x Coverage
2x100bp
11 days

HiSeq4000



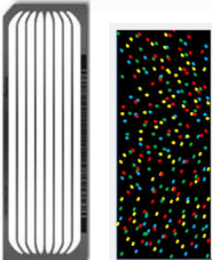
30x Coverage
2x150bp
3 days

NovaSeq6000 (high-output run mode)

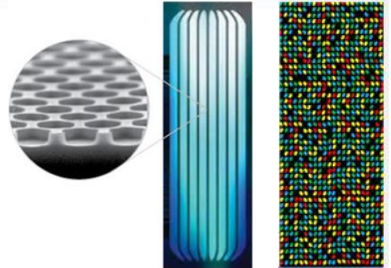


30x Coverage
2x150bp
44 hours

Non-Patterned Flow cell



Patterned Flow cell



CNAG's Workflow



BIOREPOSITORY



LABORATORY



SEQUENCING



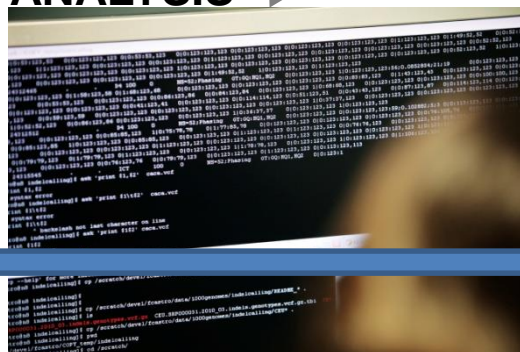
Subproject	Status	Deadline (Countdown)	Details by Sample	Details by Library	Analysis	Sample Type	Application	Expected Capture Protocol	Numbers of Samples	Ready for Sampleprep / Sequence	Sample Reception Dates	Aliquots	Libraries pass / total Barcodes	Libraries with a pass Library	Samples with a pass Library	Flowcells & Dates	Samples above Coverage	Samples above pmr by Sample	Data Transferred		
□	..11	Open (Experimental) (26 days)	Details by Sample Sample Aggregate View	Details by Library Library Aggregate View	BAG	gDNA	ExomeCapture-Seq	Agilent Human All Exon SIMM v4		yes	E061 2012-07-02 E058 948C E055 863C E059 863C E056 864C E060 E057 E054	861C 3/4 948C 863C 863C 864C	3							0 sent, of 4 to be seq	
□	..01	Open (Experimental) (27 days)	Details by Sample Sample Aggregate View	Details by Library Library Aggregate View	BAG	gDNA	WG-Seq			yes	H367 2012-07-03 H368 H368 H367	580C 2/2 581C	580C 581C	4							0 sent, of 2 to be seq
□	..03	Open (Experimental) (28 days)	Details by Sample Sample Aggregate View	Details by Library Library Aggregate View	No	Exome library	ExomeCapture-Seq	Nimblegen SeqCap 64 Mb v3		yes	H349 2012-06-18 H355 H346 H352 H343 H359 H340 H356 H337	80 / 80 I231 I229 I230 I227 I228 H342 H344 H346 H347	80							80 sent, of 80 to be seq	

LIMS

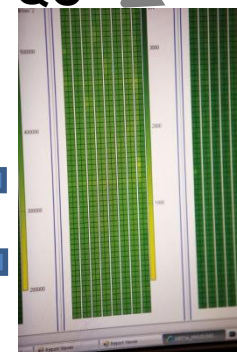
TRANSFER



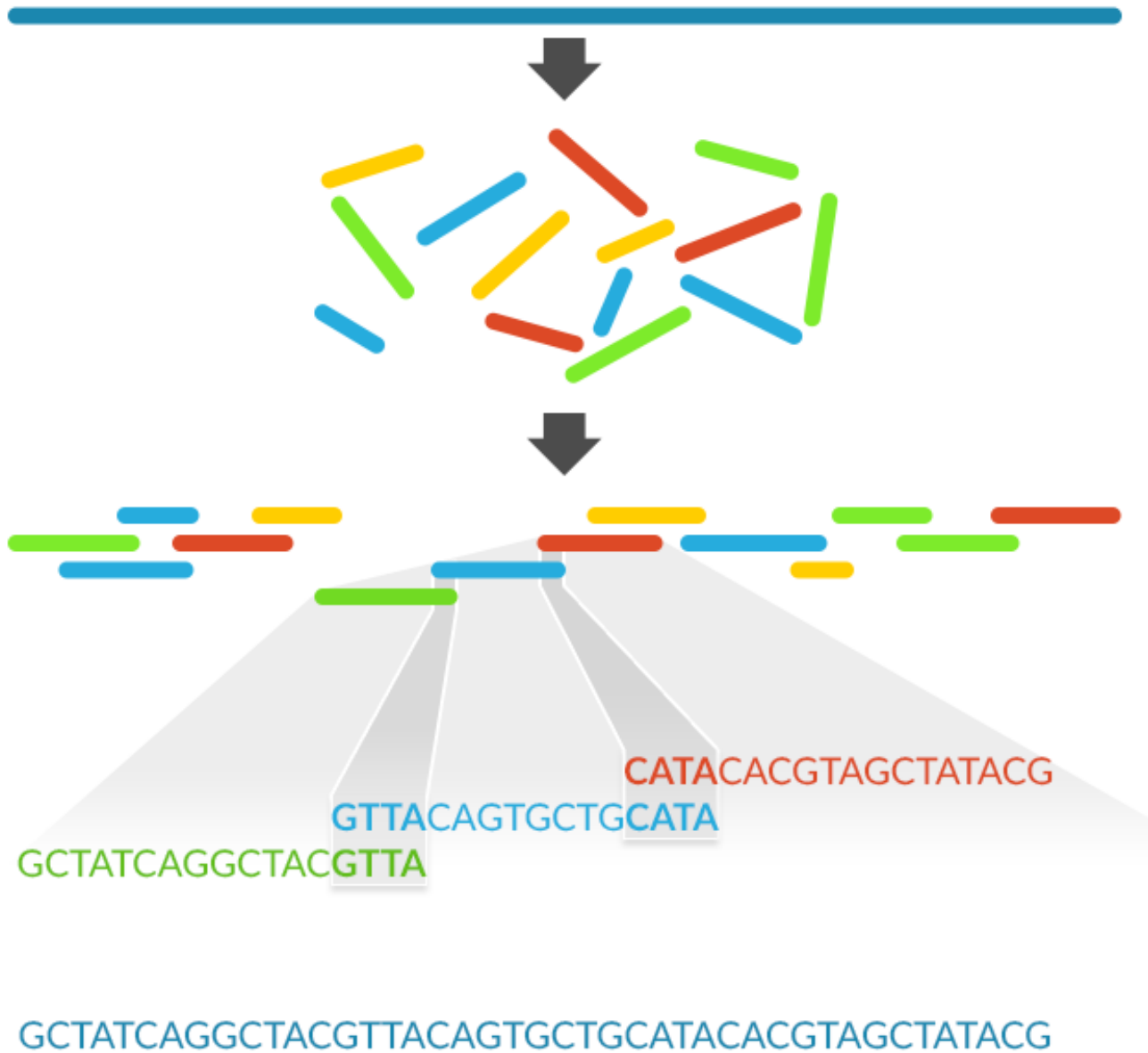
ANALYSIS

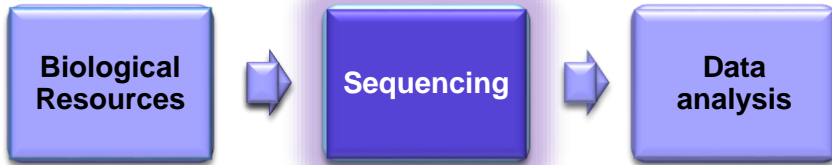


QC



Genome Sequencing





Marta Gut
Head of Sequencing

1 Library Preparation



Fragment DNA
Repair ends
Add A overhang
Ligate adapters
Purify

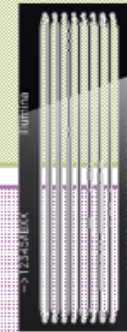


Julie Blanc, Sample Preparation manager

2 Cluster Generation



Hybridize to flow cell
Extend hybridized template
Perform bridge amplification
Prepare flow cell for sequencing



Katja Kahlem, Production manager

3 Sequencing



Perform sequencing
Generate base calls

1 Library Preparation



Fragment DNA
Repair ends
Add A overhang
Ligate adapters
Purify

Fragmentation

Repair Ends

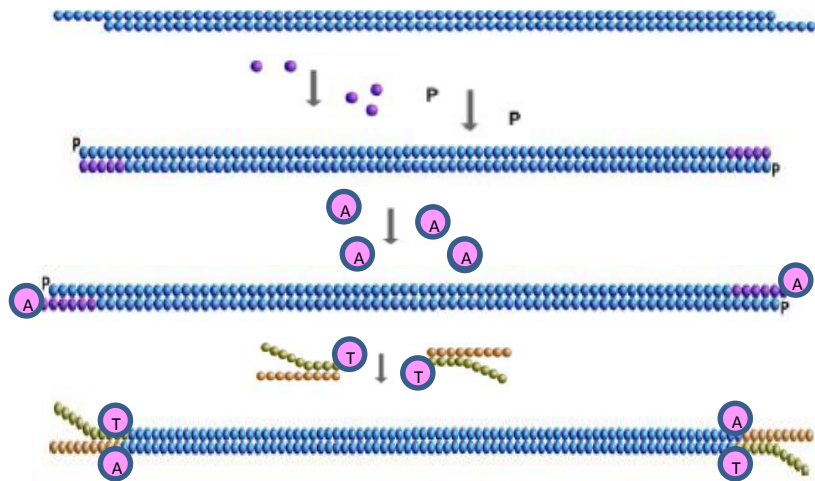
Add an "A" to the 3' ends

Ligate Adapters

Size Selection

QC

DNA fragments



1 Library Preparation

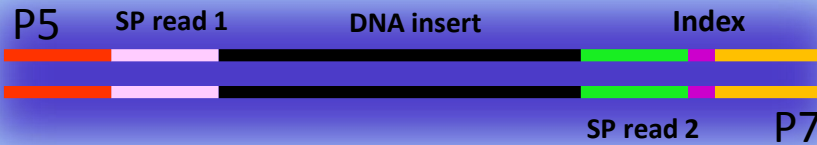


- Fragment DNA
- Repair ends
- Add A overhang
- Ligate adapters
- Purify

WG_BS_Seq

mRNA Seq

Target capture



WG_Seq

...and many more

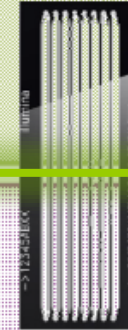
smallRNA_Seq

ChIP_Seq

2 Cluster Generation



Hybridize to flow cell
Extend hybridized template
Perform bridge amplification
Prepare flow cell for sequencing



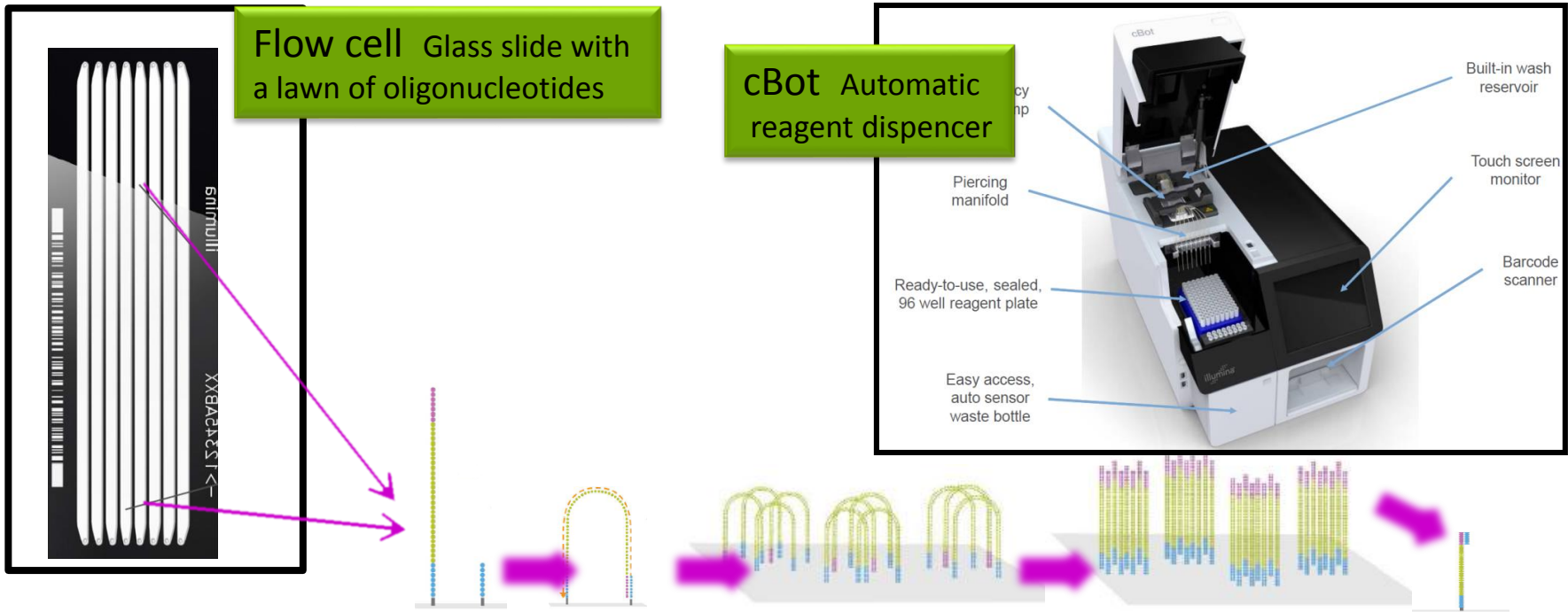
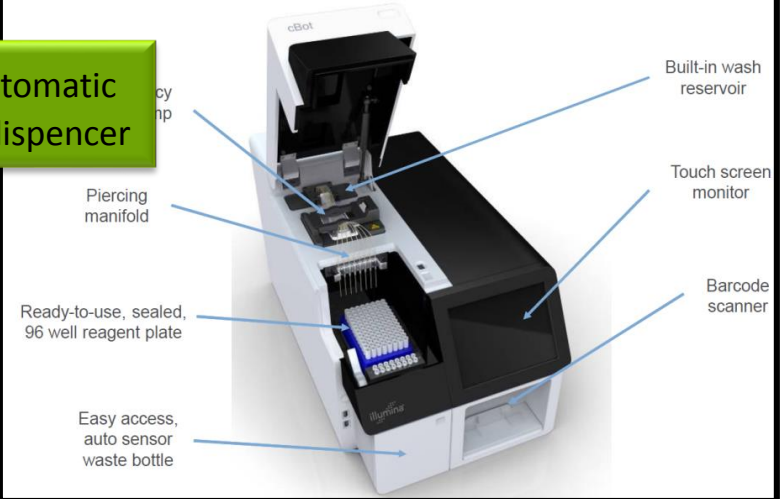
3 Sequencing



Perform sequencing
Generate base calls

Flow cell Glass slide with a lawn of oligonucleotides

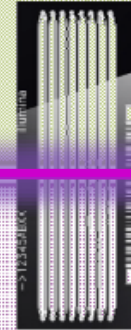
cBot Automatic reagent dispenser



2 Cluster Generation



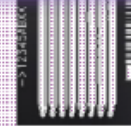
Hybridize to flow cell
Extend hybridized template
Perform bridge amplification
Prepare flow cell for sequencing



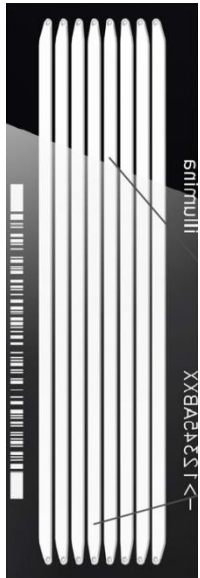
3 Sequencing



Perform sequencing
Generate base calls



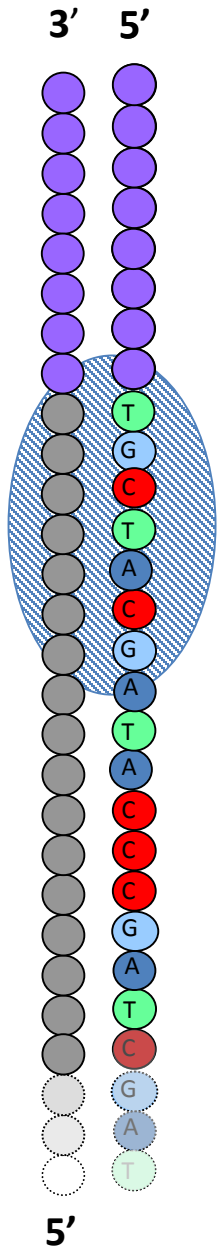
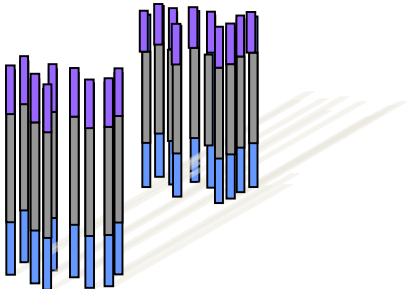
Flow cell Glass slide with a lawn of oligonucleotides and sequencing library



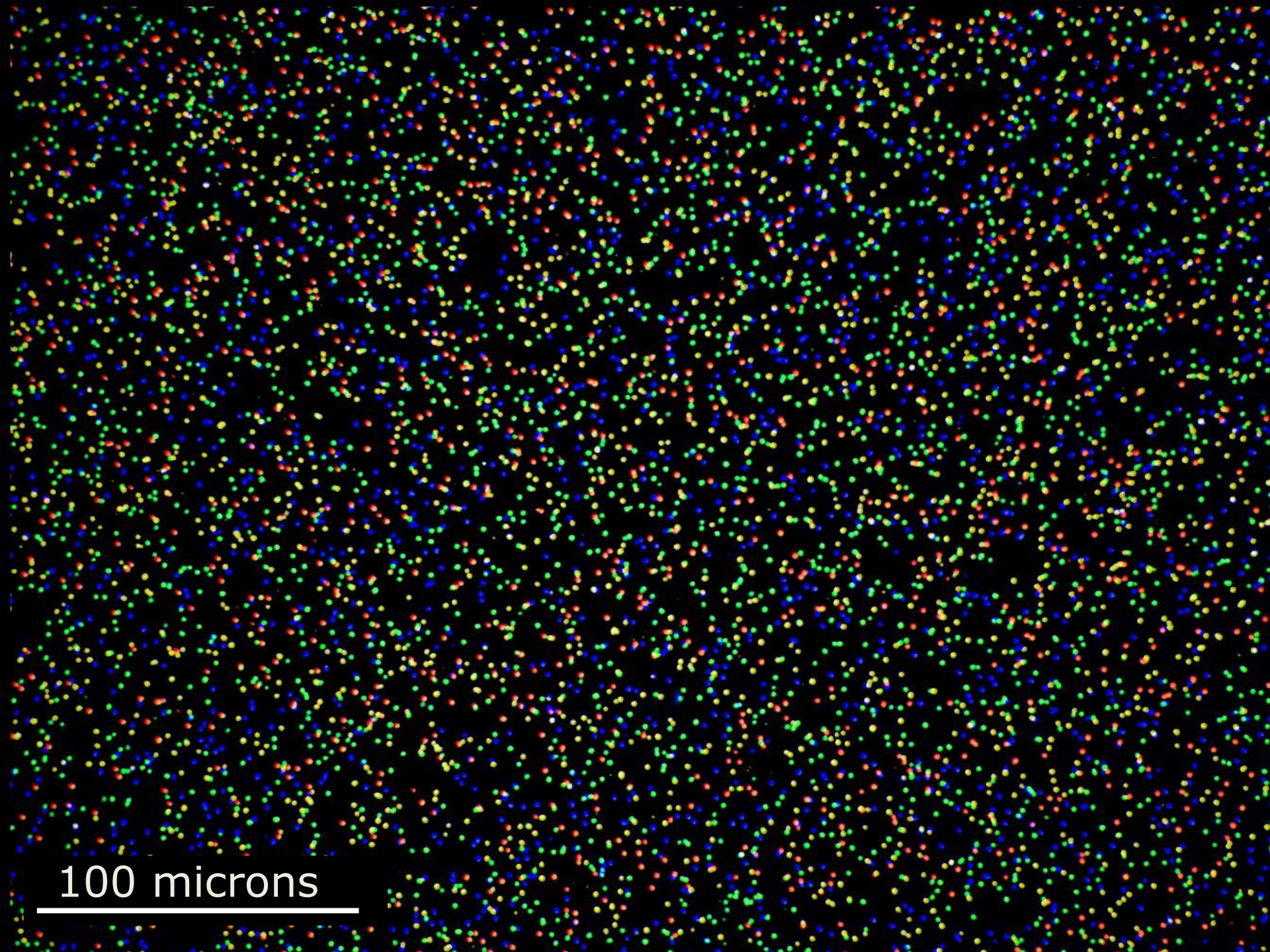
HiSeq2000 – the sequencer



Sequencing-by-synthesis (SBS)

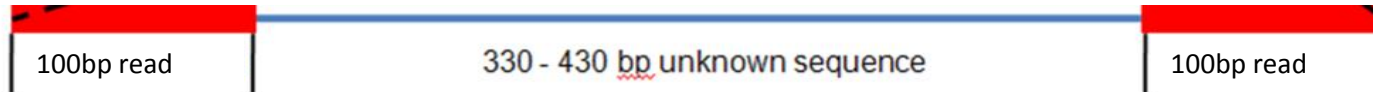
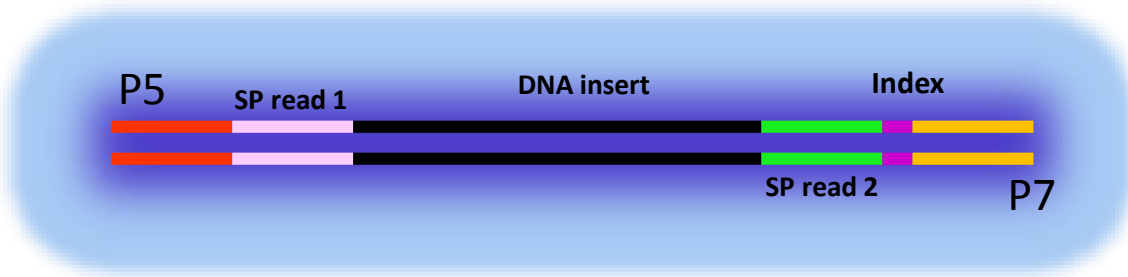


- All four labelled nucleotides in one reaction
- High accuracy
- Base-by-base sequencing
- No problems with homopolymer repeats



100 microns

Sequencing Output: FASTQ files



@SEQ_ID

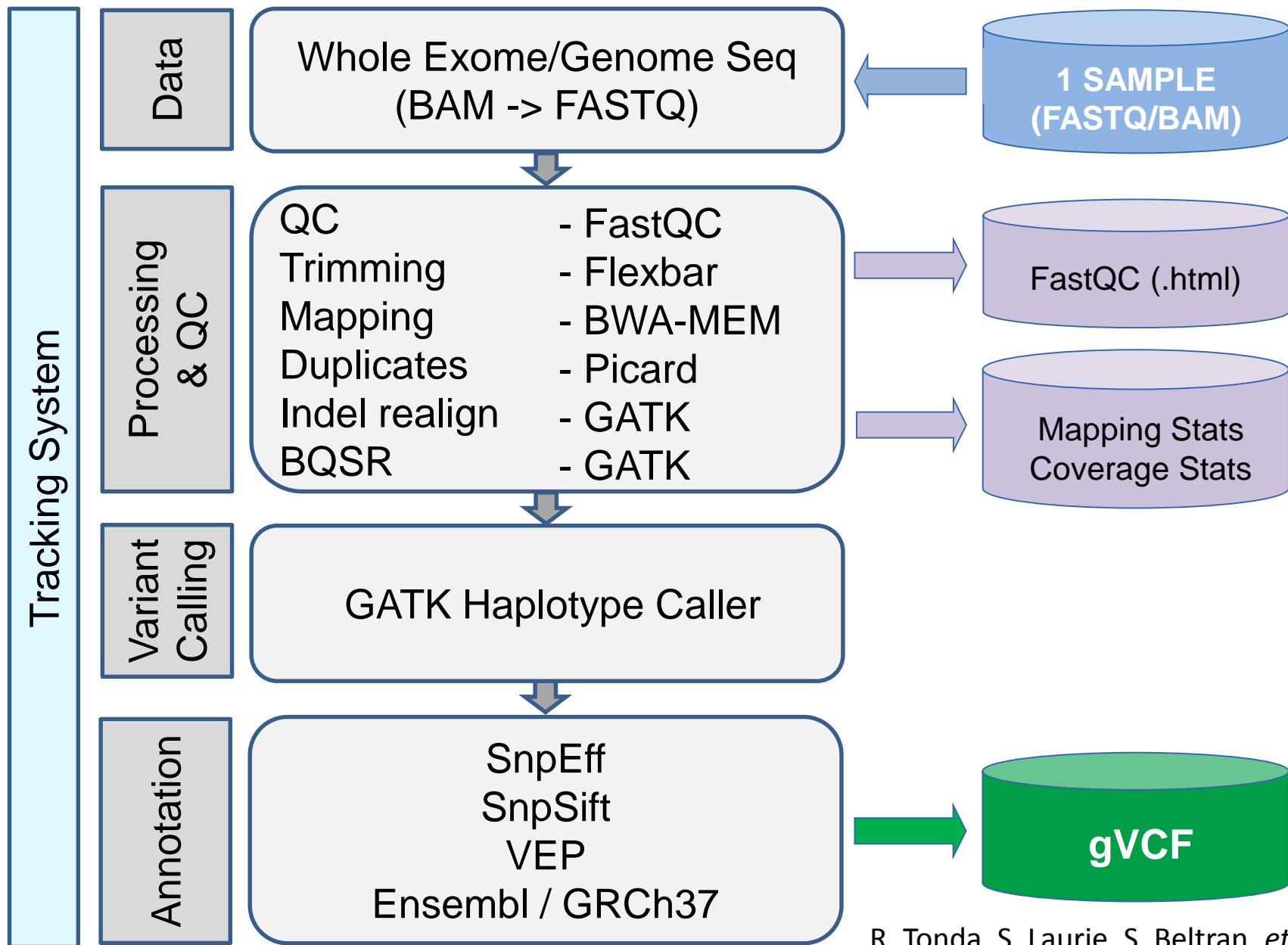
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT

+

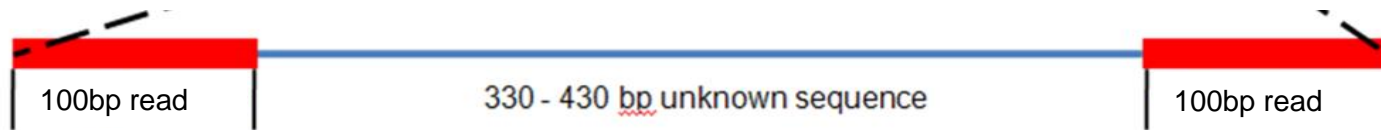
! ' ' * ((((* * * +)) % % % + +) (% % % %) . 1 * * * - + * ' ') * * 5 5 C C F > > > > > > C C C C C C C 6 5

- Developed by the Wellcome Trust Sanger Institute
- Usually, each sequence (read) is split in 4 rows
- Sequence identifiers, description and quality encoding can be different

RD-Connect Genomics variant calling pipeline

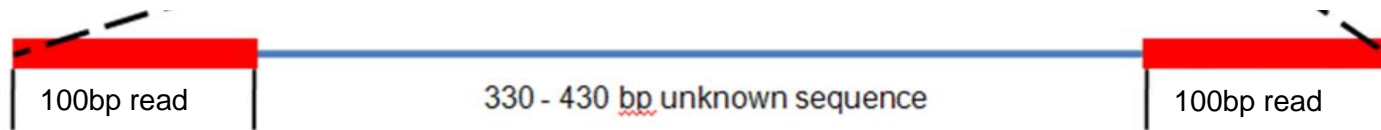



Mapping to reference genome

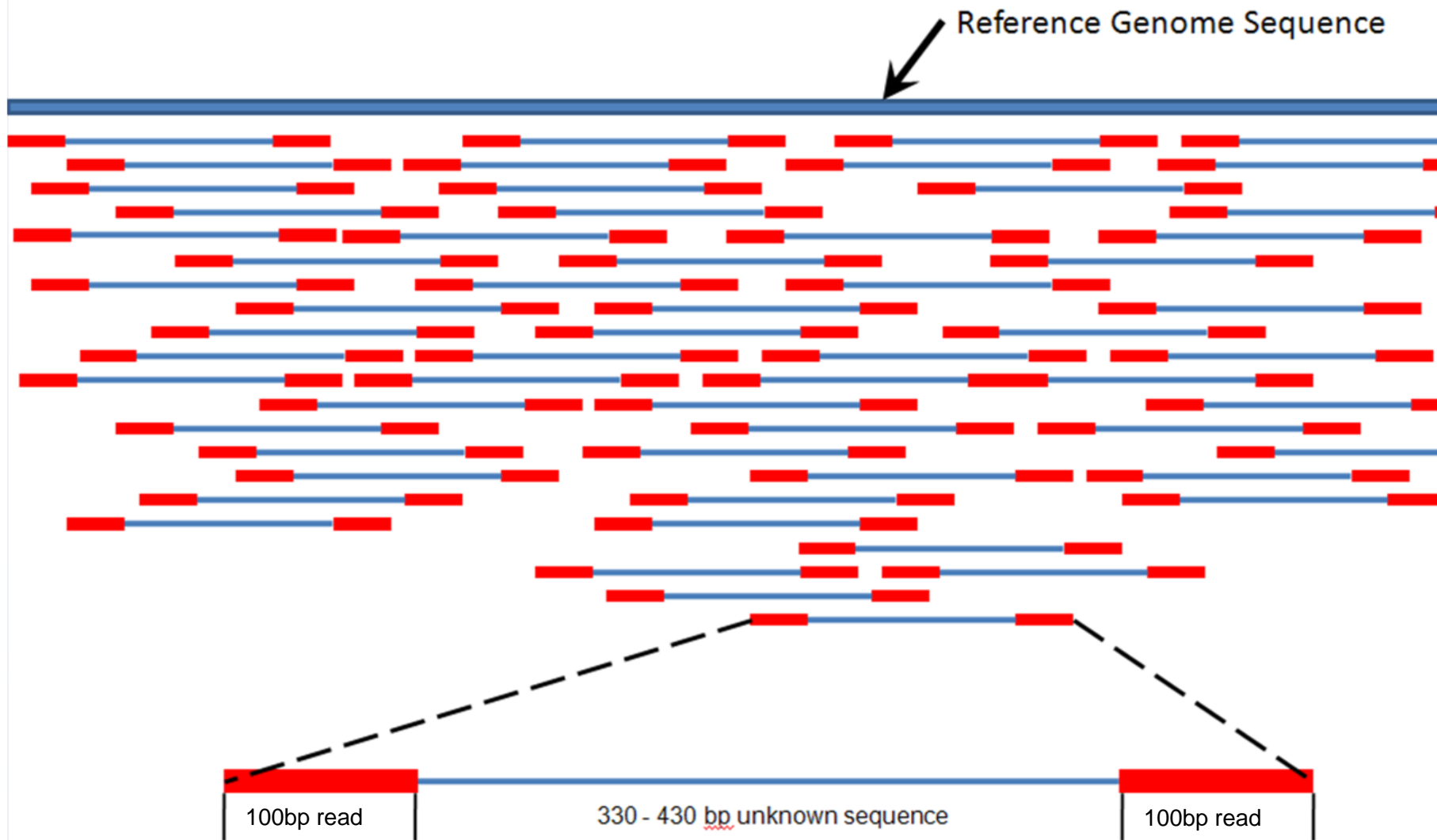


Mapping to reference genome

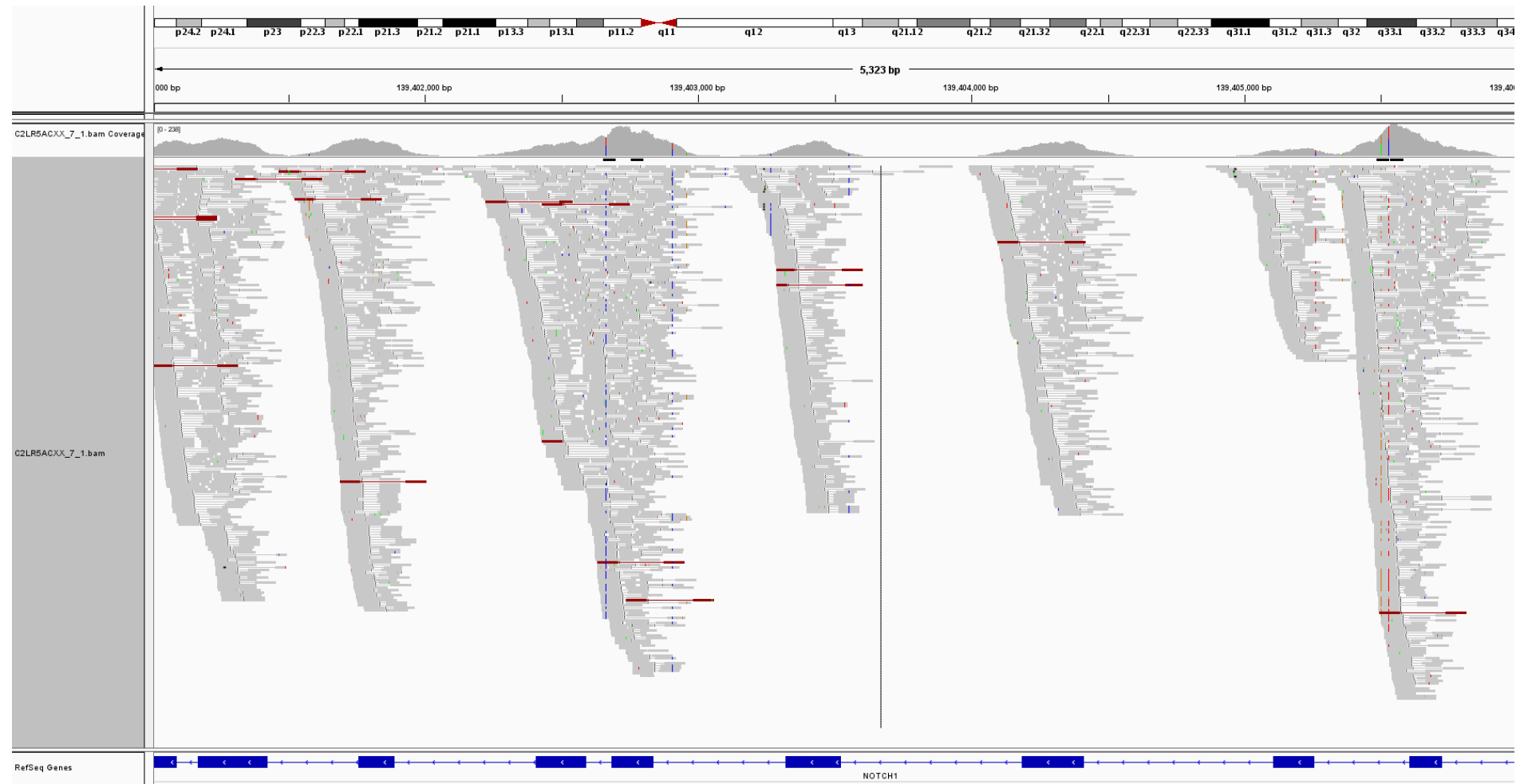
Reference Genome Sequence



Mapping to reference genome



Mapping: Exome sequence example



Alignments are stored in a **BAM file**, which is the binary version of **SAM** (Sequence/Alignment Map) format

Identification of genetic differences in comparison to a reference

Reference
(haploid)

TGGACCATCTGGTTGAGCATGTGGGGGTCAACTCCCACATTCCCAGGGAGCCCCCGG

The true
diploid
genome of
the sample

TGG**A**CCATCTGGTTGAGCAT**T**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**G**CCCCCGG
 TGG**A**CCATCTGGTTGAGCA**C**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG

ref/ref 0/0 homozygous reference ref/alt 0/1 heterozygous alt/alt 1/1 homozygous alternative ref/alt 0/1 heterozygous

Aligned
sequencing data
derived from the
sample

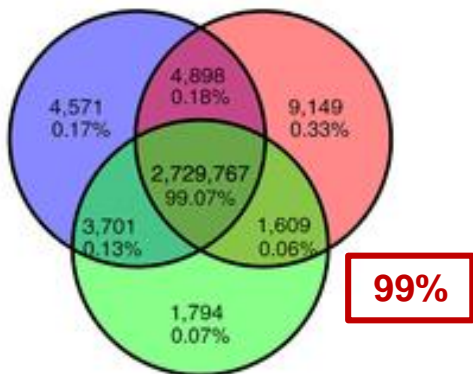
TGG**A**CCATCTGGTTGAGCAT**T**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG
 TGG**A**CCATCTGGTTGAGCA**C**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG
 TGG**A**CCATCTGGTTGAGCA**C**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG
 TGG**A**CCATCTGGTTGAGCAT**T**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG
 TGG**A**CCATCTGGTTGAGCAT**T**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG
 TGG**A**CCATCTGGTTGAGCA**C**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**G**CCCCCGG
 TGG**A**CCATCTGGTTGAGCAT**T**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG
 ATCTGGTTGAGCA**C**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG
 GGTTGAGCAT**T**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**G**CCCCCGG
 GTTGAGCA**C**GTGGGGGTCAACT**T**CCACATTCCCAGGGAG**C**CCCCCGG

ref/ref 0/0 homozygous reference 0% alternative allele ref/alt heterozygous 50% alternative allele alt/alt 1/1 homozygous alternative 100% alternative allele ? ? ? 20% alternative allele

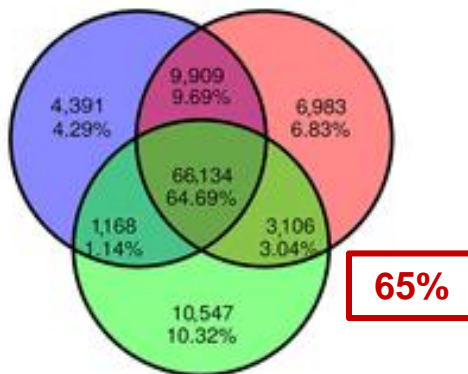
Benchmarking genomics pipelines

Reliably
Callable

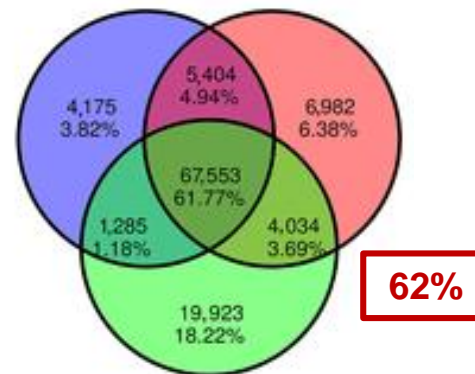
SNVs: NIST reliable



Dels: NIST reliable

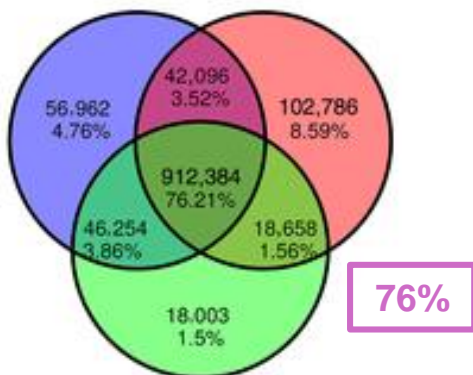


Ins: NIST reliable

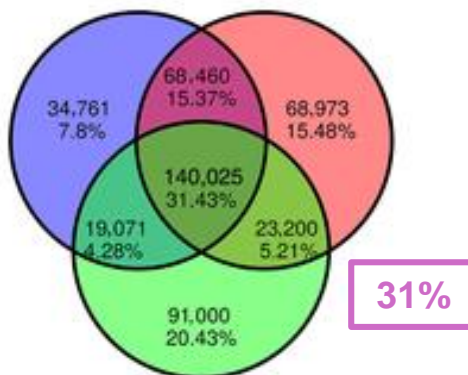


Not
Reliably
Callable

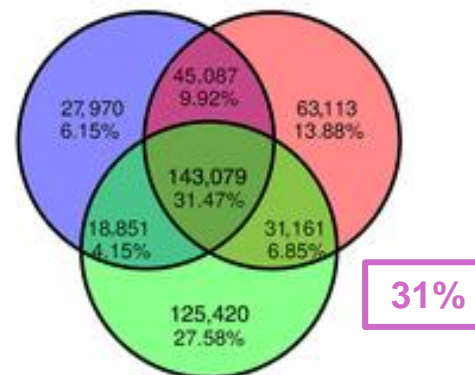
SNVs: NIST non-reliable



Dels: NIST non-reliable



Ins: NIST non-reliable



● FreeBayes ● HaplotypeCaller ● SAMtools

Rare diseases and genomics



RARE DISEASES



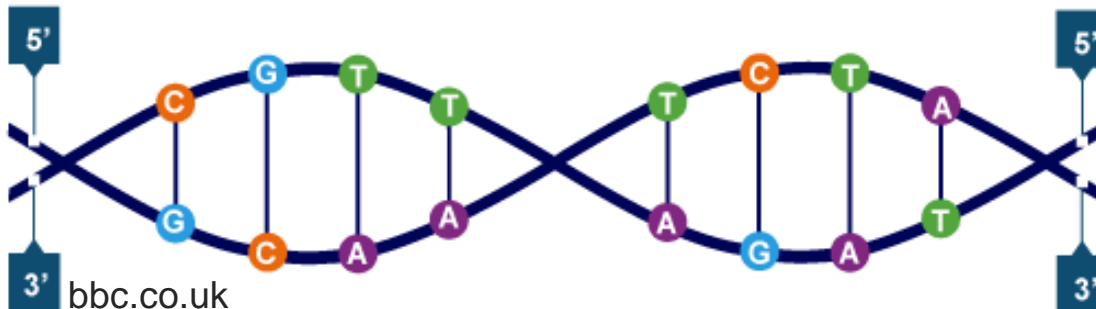
7% OF THE POPULATION ARE AFFECTED BY RARE DISEASES

THE EU CLASSSES A DISEASE AS 'RARE' WHEN **LESS THAN 1 IN 2000 SUFFER**



OVER 7000 DISEASES

- OFTEN CHRONIC AND LIFE-THREATENING
- 80% OF GENETIC ORIGIN



Human Genome:
3,200,000,000 bp
2 x 23 chromosomes

Each of us has
3.000.000 variants

Rare diseases and impact of reaching a molecular diagnosis

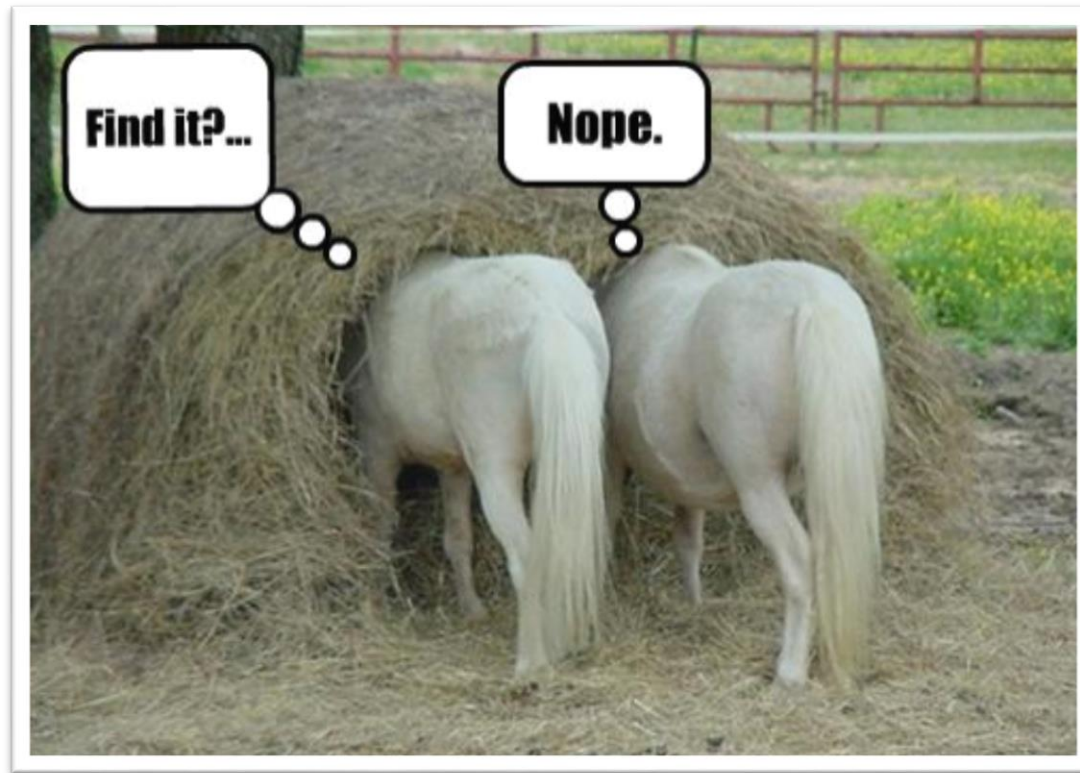
- **High phenotypic and genetic heterogeneity:**
 - **Challenge for diagnosis** mean average : 4.8 years.
- Importance of reaching a molecular diagnosis:
 - Treatment options
 - Curative (very few at this time)
 - Prevent progression: diet restriction metabolic disorders (e.g galactosemia)
 - Symptomatic and functional therapies
 - Progression and prognosis of the disease
 - Genetic counselling (prenatal / pre-implantational options)
 - Reduce patient and family anxiety



The diagnostic and gene discovery challenge

25

How do I find the pathogenic mutation within 3.000.000?



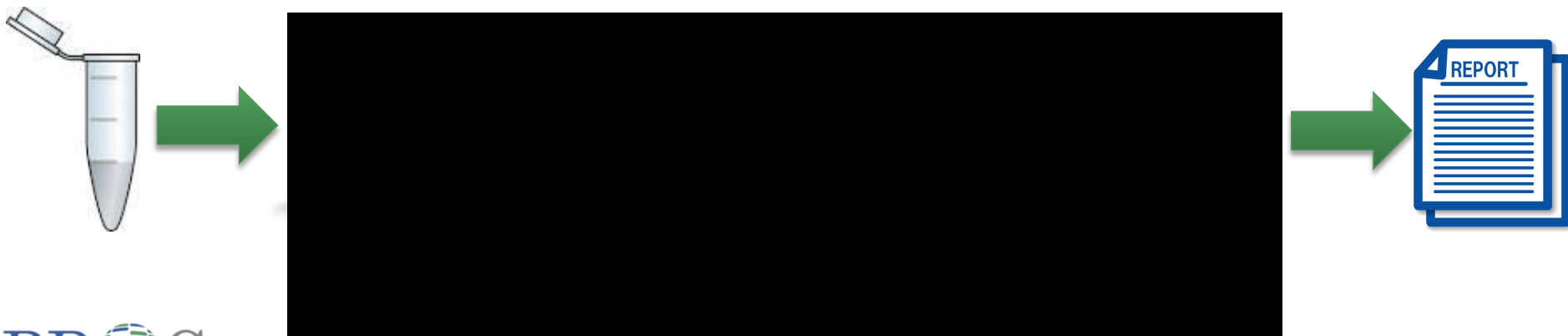


Interpretation is still difficult

26

Molecular diagnostics in NGS era

Sample in → Diagnosis out?





Interpretation is still difficult

27

Molecular diagnostics in NGS era

Sample in ~~→~~ Diagnosis out?





RD-Connect Platform (platform.rd-connect.eu)

28



Home

Genome-Phenome
Analysis Platform

PhenoTips

Registry and Biobank
Finder

Biosample Catalogue

An integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research

Welcome to the central platform for access to data submitted by RD-Connect's partner projects. The online Genome-Phenome Analysis Platform is now open for submissions from all users. Our automated registration system will come online shortly, but if you would like to access the interface now please email platform@rd-connect.eu and we will contact you to request the information we need to set you up on the system.

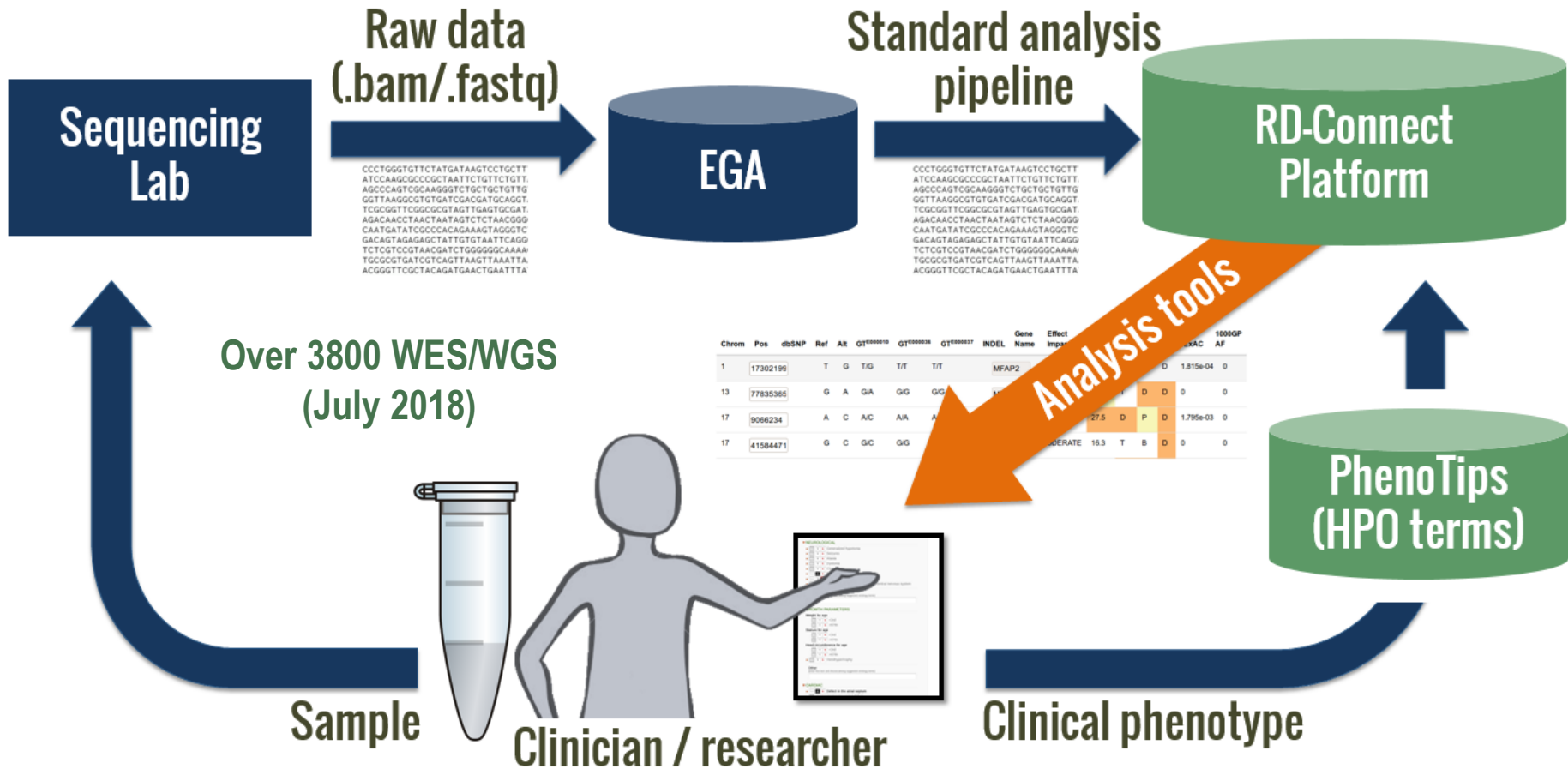
Get started today



RD-Connect Genome-Phenome Analysis Platform (GPAP)



29



Ethics Committee Approval, GDPR Compliance, Code of Conduct, Data Access Committee, Security audited, User Activity logged



RD-Connect uses PhenoTips for collating phenotypic data

30

NEUROLOGICAL

- NA Y N Generalized hypotonia
- NA Y N Seizures
- NA Y N Ataxia
- NA Y N Dystonia
- NA Y N Chorea
- NA Y N Spasticity
- NA Y N **Spinal dysraphism**
- NA Y N Morphological abnormality of the central nervous system

Other
(enter free text and choose among suggested ontology terms)

GROWTH PARAMETERS

Weight for age

- NA Y N <3rd
- NA Y N >97th

Stature for age

- NA Y N <3rd
- NA Y N >97th

Head circumference for age

- NA Y N <3rd
- NA Y N >97th

- NA Y N Hemihypertrophy

Other
(enter free text and choose among suggested ontology terms)

CARDIAC

- NA Y N Defect in the atrial septum

Onset

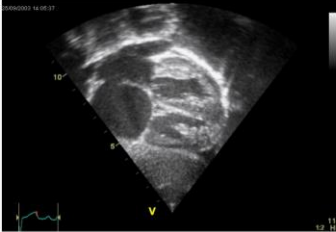
- Congenital onset
 - Embryonal onset
 - Fetal onset
 - Neonatal onset
 - Infantile onset
- Juvenile onset
- Adult onset
 - Young adult onset
 - Middle age onset
 - Late onset

Pace of progression:

- Unknown
- Nonprogressive disorder
- Slow progression
- Progressive disorder
- Rapidly progressive
- Variable progression rate

Comments:
No complications

Image / photo (optional):



Medical report (optional):

None available

Deep phenotyping in PhenoTips (Brudno *et al.*) achieved using the Human Phenotype Ontology (HPO – Robinson, Köhler *et al.*)

Diseases classified using the Orphanet Rare Disease Ontology and OMIM identifiers

Phenotype

- Basics >
- Diagnosis >
- Clinical symptoms >

Information from PhenoTips can be sent directly to other tools within the platform (e.g. Exomiser, MME)

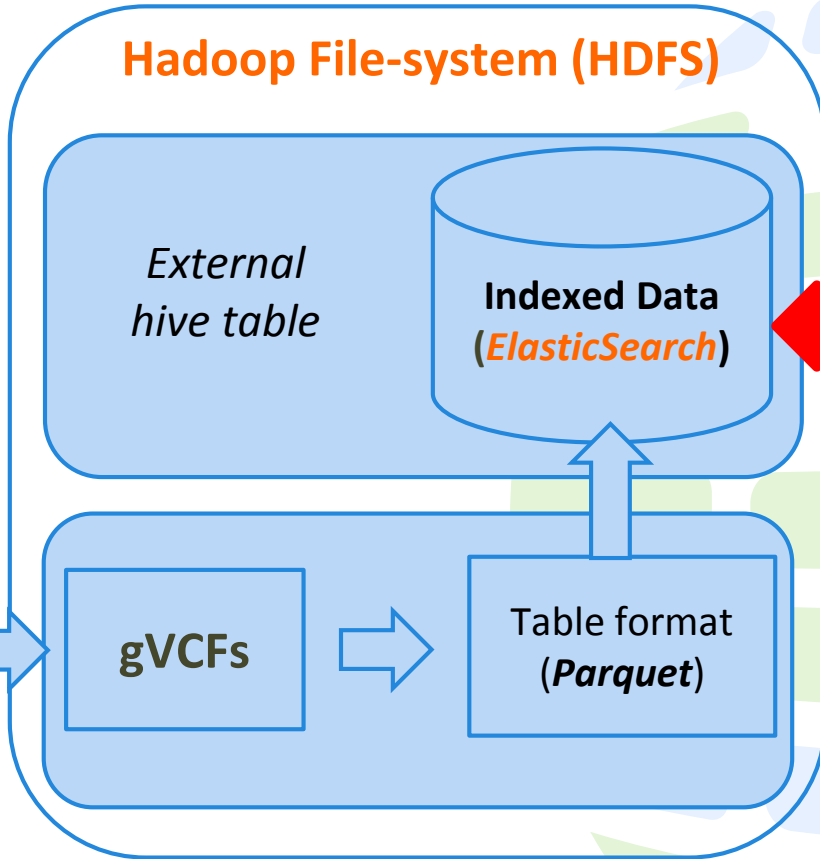


Big-data platform architecture

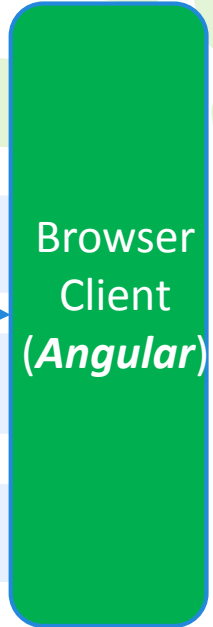
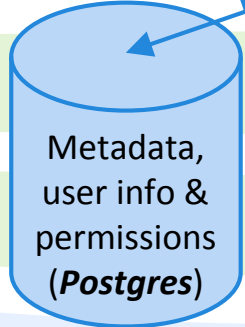
31



Authorised Access



Real-time Queries



D. Piscia, A. Papakonstantinou., C. Luengo, D. Picó, S. Laurie, I. Martínez, F. Camacho, J. Protasio, I. Gut, S. Beltran (CNAG-CRG)
JM Fernández, S. Capella, A. Valencia (BSC)

RD-Connect Genome- Phenome Analysis Platform

32

RDConnect

GENOMICS

ABOUT WELCOME TEST {PLATFORM V1.6.2, DATASET PLAY20180607} FAQ LOGOUT

Filters **PRESET FILTERS** **RESET** **SHARE** **RUN QUERY**

Sample Selection ⓘ

Select individual Samples **+** or search across all ⓘ (accessible: 15, own: 0, shared: 0, visible to all: 15)

Variant Type ⓘ

Population ⓘ

SNV Effect Prediction ⓘ

Genes, Disorders and Phenotypes

Position Specific filters and Runs Of Homozygosity

Samples Functional Predictive Population Pathways Protein interaction Diseasecard Candidate Links ALFA

RD-Connect ID	Participant ID	GT	GQ	DP	AAF
---------------	----------------	----	----	----	-----

Phenotype Analysis status Variants () Exomiser

Phenotips id	External id	Clinical status*	Inheritance	Consanguinity	Genes	Family	Pedigree	Relatives	OMIM disorder	ORDO disorder	HPO terms
--------------	-------------	------------------	-------------	---------------	-------	--------	----------	-----------	---------------	---------------	-----------

1. SAMPLES

2. FILTERS

3. RESULTS



Genomic platform: samples, inheritance and quality parameters

33

Sample Selection ?

Select individual Samples + or search across all ? (accessible: 15, own: 0, shared: 0, visible to all: 15)

? Compound het.

Affected	Experiment ID	Phenotips	MME	REF/REF	REF/ALT	ALT/ALT	Min Depth	Min Genotype Quality	Min Alternate Allele Freq	Max Alternate Allele Freq	
<input checked="" type="checkbox"/>	Case1C P0007498	P	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	10	30	0.2	0.8	×
<input type="checkbox"/>	Case1F P0007499	P	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	10	30	0.2	0.8	×
<input type="checkbox"/>	Case1M P0007500	P	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	10	30	0.2	0.8	×

Genotypes – mode of inheritance

Quality parameters

Filters **PRESET FILTERS** **RESET** **SHARE** **RUN QUERY**

Sample Selection

Select individual Samples or search across all (accessible: 15, own: 0, shared: 0, visible to all: 15)

Variant Type

Population

SNV Effect Prediction

Genes, Disorders and Phenotypes

Position Specific filters and Runs Of Homozygosity

Samples Functional Predictive Population Pathways Protein interaction Diseasecard Candidate Links ALFA

RD-Connect ID Participant ID GT GQ DP AAF

Phenotype Analysis status Variants () Exomiser

Phenotips External Clinical OMIM ORDO HPO
id id Gender status* Inheritance Consanguinity Genes Family Pedigree Relatives disorder disorder terms

1. SAMPLES

2. FILTERS

3. RESULTS



Genomic platform: filtering steps

35

Variant Type i

As defined by SnpEFF - see [Effect prediction details](#) section for a detailed explanation

Variant Class

- High
- Moderate
- Low
- Modifier

ClinVar Classification

- Pathogenic
- Likely pathogenic
- Variant of uncertain significance
- Conflicting interpretations
- Drug response
- Any

Variant Type

- SNV
- INDEL

Transcript Biotype

- Protein_coding
- RNA
- Other

Tagged Variants

- Selected samples
- Any samples

Hover the cursor over the text to see tool tips that explains the field

Population i

SNV Effect Prediction i

Genes, Disorders and Phenotypes

Position Specific filters and Runs Of Homozygosity

RD-Connect genome phenome analysis platform:

36

RDConnect

GENOMICS

ABOUT WELCOME TEST {PLATFORM V1.6.2, DATASET PLAY20180607} FAQ LOGOUT

Filters **PRESET FILTERS** **RESET** **SHARE** **RUN QUERY**

Sample Selection ⓘ

Select individual Samples **+** or search across all ⓘ (accessible: 15, own: 0, shared: 0, visible to all: 15)

Variant Type ⓘ

Population ⓘ

SNV Effect Prediction ⓘ

Genes, Disorders and Phenotypes

Position Specific filters and Runs Of Homozygosity

Samples Functional Predictive Population Pathways Protein interaction Diseasecard Candidate Links ALFA

RD-Connect ID	Participant ID	GT	GQ	DP	AAF
---------------	----------------	----	----	----	-----

Phenotype Analysis status Variants () Exomiser

Phenotyps id	External id	Clinical status*	Inheritance	Consanguinity	Genes	Family	Pedigree	Relatives	OMIM disorder	ORDO disorder	HPO terms
--------------	-------------	------------------	-------------	---------------	-------	--------	----------	-----------	---------------	---------------	-----------

1. SAMPLES

2. FILTERS

3. RESULTS



Genomic platform: results section

37

Phenotype Analysis status Variants (22) Exomiser

First Previous 1 Next Last

Chr	Position	dbSNP	Ref	Alt	Candidate	GT Case1C	GT Case1F	GT Case1M	INDEL	Gene	Effect Impact	ClinVar	CADD	SIFT	PP2	MT	ExAC	1000GP AF	gnomAD AF	Internal Freq
1	91859888		GT	G	0 TAG	GT/G	GT/GT	GT/GT	☑	HFM1	HIGH		< 20				NA	NA	NA	0.166667
			G	C	0 TAG	G/C	G/G	G/G		OMIM	MODERATE		23.1	D	P	N	NA	NA	NA	0.166667
			G	A	0 TAG	G/A	G/G	G/G		Ensembl	MODERATE		23.4	D	B	N	NA	NA	NA	0.166667
			C	G	0 TAG	C/G	C/C	C/C		PubMed	MODERATE		< 20				NA	NA	NA	0.166667
		rs60722486	G	A	0 TAG	G/A	G/G	G/G		FARP2	MODERATE		< 20	T	B	N	0.000025	NA	0.000094	0.166667
			A	C	0 TAG	A/C	A/A	A/A		HGMDB	MODERATE		24.0	D	D	N	NA	NA	NA	0.166667
			T	A	0 TAG	T/A	T/T	T/T		MUC20	MODERATE		26.2	D	D	N	NA	NA	NA	0.166667
			T	C	0 TAG	T/C	T/T	T/T		Entrez	MODERATE		23.5	D	P	D	NA	NA	NA	0.166667
		rs73132598	G	A	0 TAG	G/A	G/G	G/G		GeneCards	MODERATE		24.8	D	D	D	0.000033	NA	0.000033	0.166667
			TAG	T	0 TAG	TAG/T	TAG/TAG	TAG/TAG	☑	COSMIC	HIGH		< 20				NA	NA	0.000482	0.166667
			G	GTTTTTTTTT..	0 TAG	G/GTTTTTTTTT..	G/G	G/G	☑	ClinVar	MODERATE		< 20				NA	NA	0.000095	0.166667
			T	G	0 TAG	T/G	T/T	T/T		ExAC	MODERATE		< 20				0.000480	NA	NA	0.166667

- Ensembl
- ExAC
- gnomAD
- UCSC
- NCBI
- DGVa
- GWAS Central
- GA4GH Beacon
- VarSome

- OMIM
- Ensembl
- PubMed
- FARP2
- HGMDB
- MUC20
- Entrez
- GeneCards
- COSMIC
- ClinVar
- ExAC
- TECTA
- GTEEx
- UNC79
- gnomAD
- GWAS Central
- ATLAS
- WikiPathways
- Open PHACTS

Links to multiple databases



Genomic platform: results section

38

Multiple tabs with detailed information for data interpretation

Gene Name	Transcript ID	Effect Impact	Consequence	Feature Type	HGVS coding	Amino Acid change	Amino Acid length	Exon Rank	CDS Position	Transcript BioType
HFM1	ENST00000370425	HIGH	frameshift_variant	transcript	c.255delA	p.Leu86Ter	1435	4/39	255/4308	protein_coding
HFM1	ENST00000427444	HIGH	frameshift_variant	transcript	c.129delA	p.Leu44Ter	196	3/4	129/591	protein_coding
HFM1	ENST00000455133	HIGH	frameshift_variant	transcript	c.255delA	p.Leu86Ter	138	4/4	255/417	protein_coding

Phenotype Analysis status Variants (22) Exomiser

First Previous 1 Next Last

Chr	Position	dbSNP	Ref	Alt	Candidate	GT Case1C	GT Case1F	GT Case1M	INDEL	Gene	Effect Impact	ClinVar	CADD	SIFT	PP2	MT	ExAC	1000GP AF	gnomAD AF	Internal Freq
1	91859888	.	GT	G	0 TAG	GT/G	GT/GT	GT/GT	<input checked="" type="checkbox"/>	HFM1	HIGH		< 20			NA	NA	NA	NA	0.166667
1	225477618	.	G	C	0 TAG	G/C	G/G	G/G		DNAH14	MODERATE		23.1	D	P	N	NA	NA	NA	0.166667
2	233349186	.	G	A	0 TAG	G/A	G/G	G/G		ECEL1	MODERATE		23.4	D	B	N	NA	NA	NA	0.166667





RD-Connect GPAP functionalities

39

- ✓ **Standard filters and annotations** (variant impact on protein, frequency in control populations, frequency within database, *in silico* predictors, gene, genomic position, etc.)
- ✓ **Filter by genes of interest** (predefined/custom lists, associated to OMIM disease, HPO symptoms or Reactome)
- ✓ Filter by variants annotated in **ClinVar database**
- ✓ Filter to regions with observed long **Runs of homozygosity** (RoH)
- ✓ Direct link to **multiple external resources** (Ensembl, UCSC, gnomAD, HGMD, Human Splicing Finder, DiseaseCards, ALFA, etc.)
- ✓ **Phenotype driven strategy for variant prioritization** (Exomiser)
- ✓ **TAG and share candidate variants** using ACMG guidelines
- ✓ **Collaborative environment:** share data and queries
- ✓ **Anonymized data discovery** through Beacon and Matchmaker exchange










Data discovery: GA4GH Beacon

40

GRCh37 ▾ 13 : 32954208 A>T Search

Response	All None
<input checked="" type="checkbox"/> Found	8
<input checked="" type="checkbox"/> Not Found	43
<input type="checkbox"/> Error	8

Organization	All None
<input type="checkbox"/> AMPLab, University of C...	
<input type="checkbox"/> BGI	
<input type="checkbox"/> BioReference Laboratories	
<input checked="" type="checkbox"/> Broad Institute	
<input type="checkbox"/> Centre for Genomic Regu...	
<input checked="" type="checkbox"/> CNAG	
<input type="checkbox"/> Curoverse	
<input type="checkbox"/> DNASTack	
<input checked="" type="checkbox"/> EMBL European Bioinfor...	
<input type="checkbox"/> Global Alliance for Geno...	
<input type="checkbox"/> Google	
<input checked="" type="checkbox"/> Institute for Systems Biol...	
<input checked="" type="checkbox"/> Mike Lin	
<input checked="" type="checkbox"/> National Center for Biote...	
<input type="checkbox"/> ...	

	EBI - 1000 Genomes Project, ... EMBL European Bioinformatics Institute	Not Found
	ExAC Broad Institute	Not Found
	ICGC - Cancer Projects Ontario Institute for Cancer Research	Not Found
	Kaviar Institute for Systems Biology	Found
	NHLBI Exome Sequence Proj... National Center for Biotechnology Information	Not Found
	RD-Connect CNAG	Not Found
		Not Found

Question:

Does any sample in your database have **variant V**?

Answer:

Yes / No



Data discovery: MatchMaker Exchange (MME, IRDiRC, GA4GH)

Question: Do you have a patient with similar phenotype and genotype as mine?



2016 BBMRI-LPC Whole Exome Sequencing Call

Sequencing the exome of 900 rare disease samples
in collaboration with EuroBioBank and RD-Connect

Objectives

- ✓ to promote the usage of biobanks for rare diseases
 - samples deposited in the EuroBioBank network
- ✓ to promote the utilization of cutting-edge next-generation sequencing technology for the identification of novel causative variants and genes
 - free-of-charge sequencing of 900 exomes
- ✓ to molecularly diagnose rare disease patients
 - analysis through the RD-Connect platform
- ✓ to promote data sharing for rare disease research to enable future discovery
 - data sharing through the EGA and RD-Connect and phenotyping with HPO



2016 BBMRI-LPC Whole Exome Sequencing Call

Sequencing the exome of 900 rare disease samples
in collaboration with EuroBioBank and RD-Connect

17 transnational projects selected

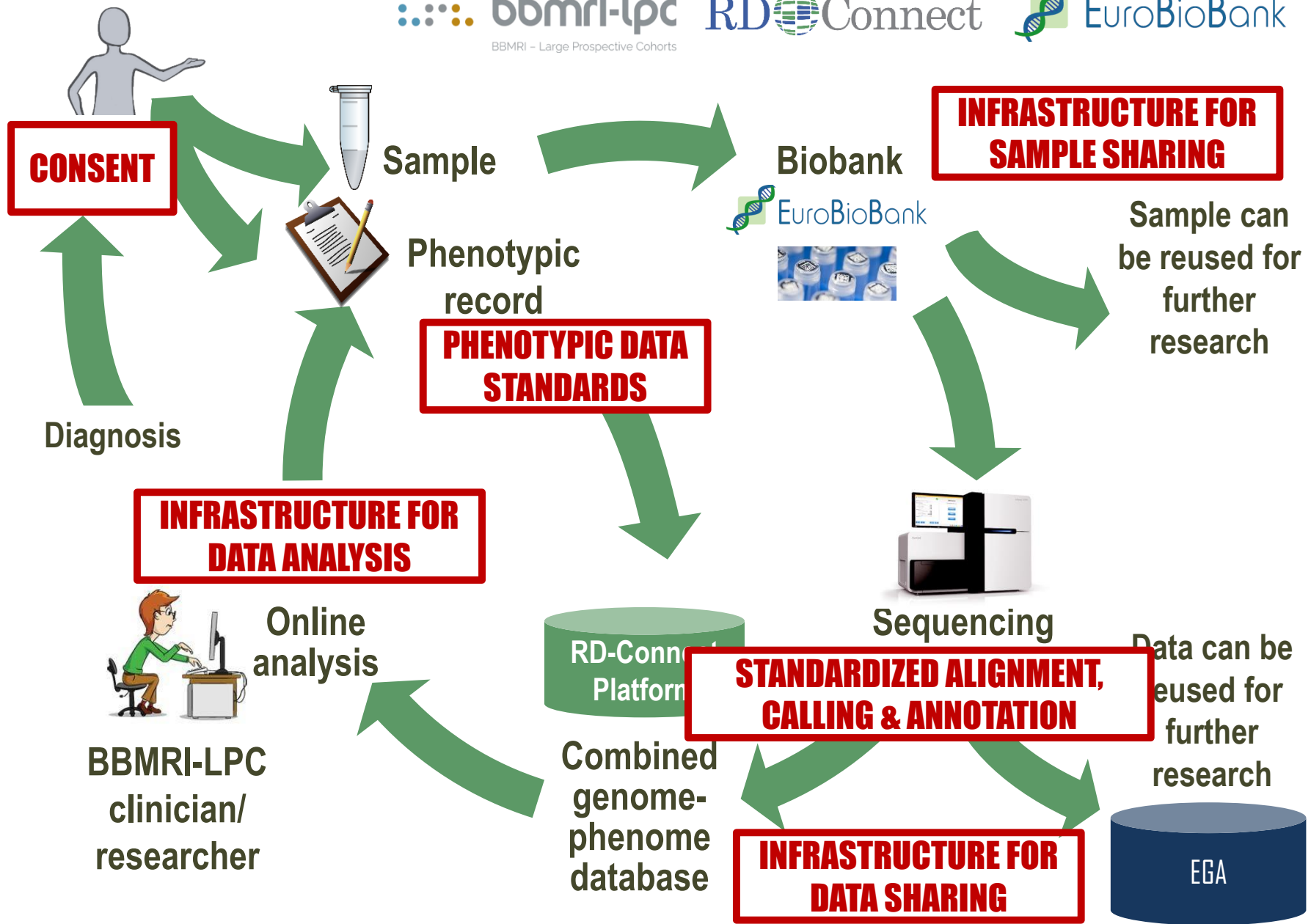
899 samples for sequencing in total

10 projects sequenced at CNAG-CRG (**545** samples)

7 projects sequenced at WTSI (**354** samples)



A paradigm for best practice in data sharing in genomic studies?





URDCAT is a research project embedded into the “Pla Estratègic de Recerca i Innovació en Salut (PERIS) 2016-2020” funded by the Health Department of the Generalitat de Catalunya.

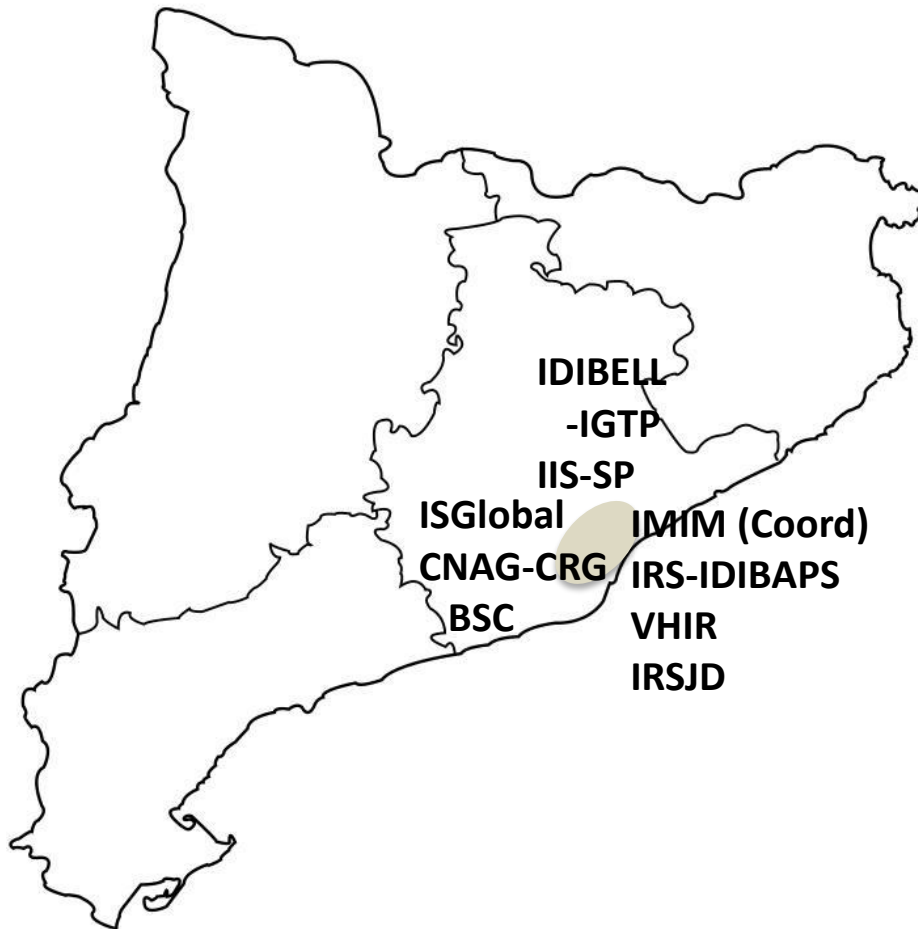
MAIN OBJECTIVE:

Enable the Catalan Health System to **provide personalised genomic medicine as a fully integrated service for patients with RDs**, initially as a **pilot project for RDs with neurologic involvement**.

Specific objectives:

- **Standardise** the process of **analysis and integration of clinical and genomic data**
- Implement a **platform for analysis of genomics** data that is suitable for clinical practice
- **Identify causative genetic variants** of undiagnosed RD patients
- **Highlight the utility of genomics**, and advance its usage as a tool of **personalised medicine** for RD diagnostics
- **Educate and train stakeholders** in the value and use of genomic data

Coordinator: Luis Pérez Jurado (IMIM)



16 Groups

- 7 Hospitals (IIS)
- 3 CERCA centers
- FEDER (Patients)

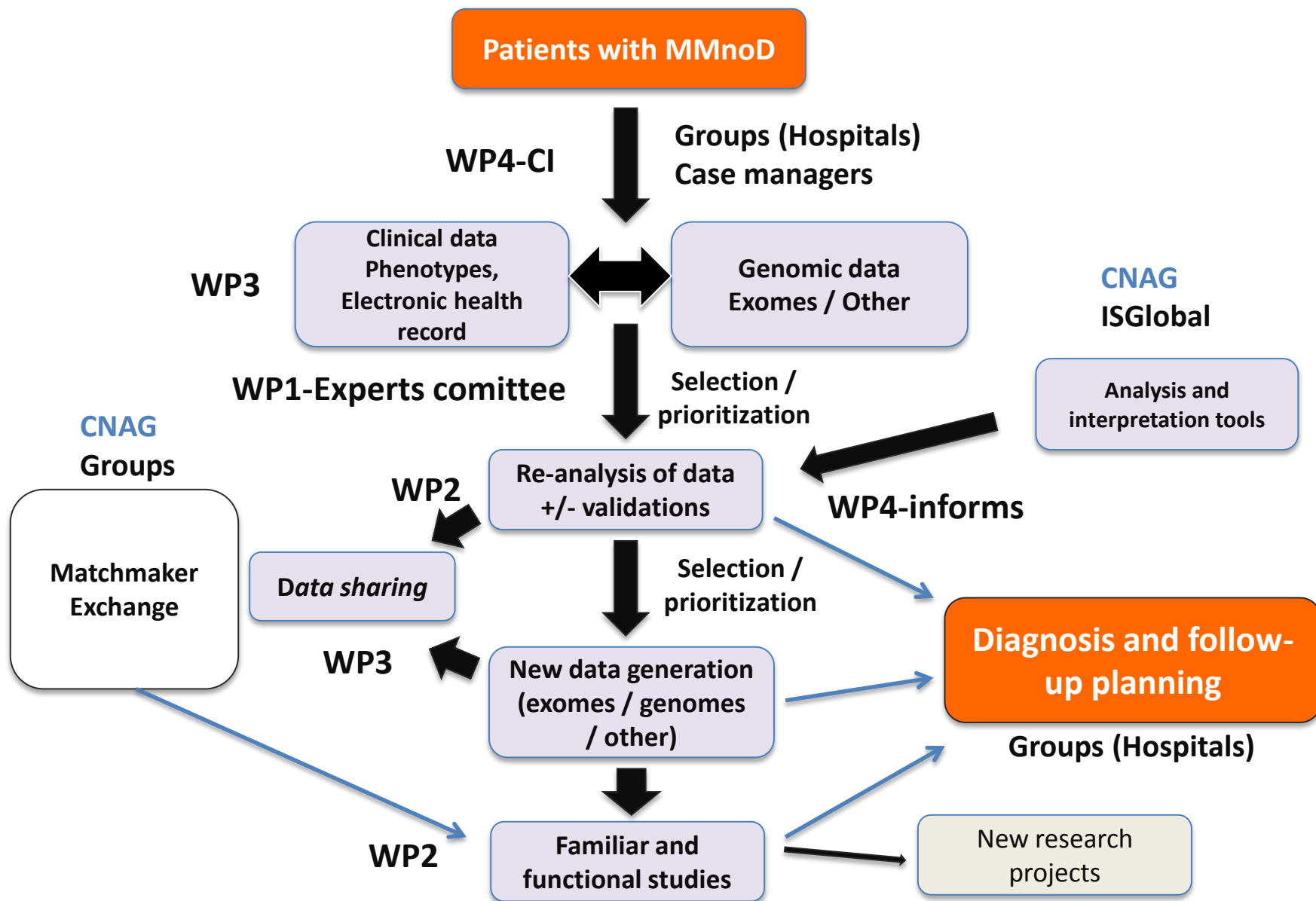
WP1. Clinical data: patient selection and characterization (Coordinator: Alfons Macaya)

WP2. Genomics: sequencing, analysis, interpretation and validation (Coordinator: Antonia Ribes)

WP3. Data analysis and Interpretation platform (RDCat) (Coordinator: Sergi Beltran)

WP4. Ethical, Legal and and Social Issues (Coordinator: Francesc Palau)

WP5. Training (Coordinator: Luis Pérez Jurado)



[Genet Med.](#) 2017 Feb;19(2):249-255. doi: 10.1038/gim.2016.190. Epub 2016 Nov 17.

Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics.

[Kalia SS](#)¹, [Adelman K](#)², [Bale SJ](#)³, [Chung WK](#)^{4,5}, [Eng C](#)⁶, [Evans JP](#)⁷, [Herman GE](#)⁸, [Hufnagel SB](#)⁹, [Klein TE](#)¹⁰, [Korf BR](#)¹¹, [McKelvey KD](#)^{12,13}, [Ormond KE](#)¹⁰, [Richards CS](#)¹⁴, [Viancos CN](#)¹⁵, [Watson M](#)¹⁶, [Martin CL](#)¹⁷, [Miller DT](#)¹⁸.

Filters ▲ PRESET FILTERS RESET SHARE RUN QUERY

Sample Selection ? ▲

Select individual Samples + or search across all ? (accessible: 0, own: 568, shared: 0, visible to all: 568)

? Compound het.

Affected	Experiment ID	Phenotips	MME	REF/REF	REF/ALT	ALT/ALT	Min Depth	Min Genotype Quality	Min Alternate Allele Freq	Max Alternate Allele Freq	
<input checked="" type="checkbox"/>	EPR532870 439		<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	20	50	0,2	0,8	✕

Variant Type ? ▼

Population ? ▼

- ACMG Medically Actionable Genes (n=59)**
- BabySeq Class A and B Genes (n=889)
- Decipher all, July2017 (n=1,692)
- Decipher confirmed, July2017 (n=1,098)
- Digenic gene list, Feb2017 (n=136)
- Imprinted-confirmed list, Feb2017 (n=80)
- Imprinted-all, Feb2017 (n=253)
- Intellectual Disability-confirmed, Nov2017 (n=10)
- Intellectual Disability-all, Nov2017 (n=1648)
- Leukodystrophy Associated (n=229)
- Medically Interpretable Genome (n=5,419)
- Mitocarta 2.0 (n=1158)
- Muscle Gene Table, July2016 (n=416)
- Primary Immune Deficiency, Nov2017 (n=387)
- Tubingen HSP Version 6 (n=140)

59 genes for which the ACMG recommends reporting as incidental or secondary findings: Kalia SS et al (2016). Genetics in Medicine

Search OMIM

Genes linked to : ✕ Remove All

Search HPO

Fetch HPOs From PhenoTips ✕ Remove All

Genes linked to :

Upload comma separated list of HGNC identifiers

CURSO MEDICINA GENÓMICA PERSONALIZADA

DIAGNÓSTICO DE ENFERMEDADES NEUROLÓGICAS NO DIAGNOSTICADAS

Clínica y Herramientas de Diagnóstico
Genómica e Interpretación de Datos
Resultados y Asesoramiento Genético

Curso dirigido a profesionales sanitarios e investigadores

Coordinadores del Curso

Dr. Luis Pérez Jurado

Dr. Francesc Palau

Dra. Antònia Ribes

Dr. Alfons Macaya

Dr. Sergi Beltrán

5, 6 y 7 de NOVIEMBRE de 2018
de 15 a 20h (Campus del Mar - UPF)

Precio del curso: 50€

Inscripciones: curso@urdc.cat

**Reconocida con 1.5 créditos de formación continuada*



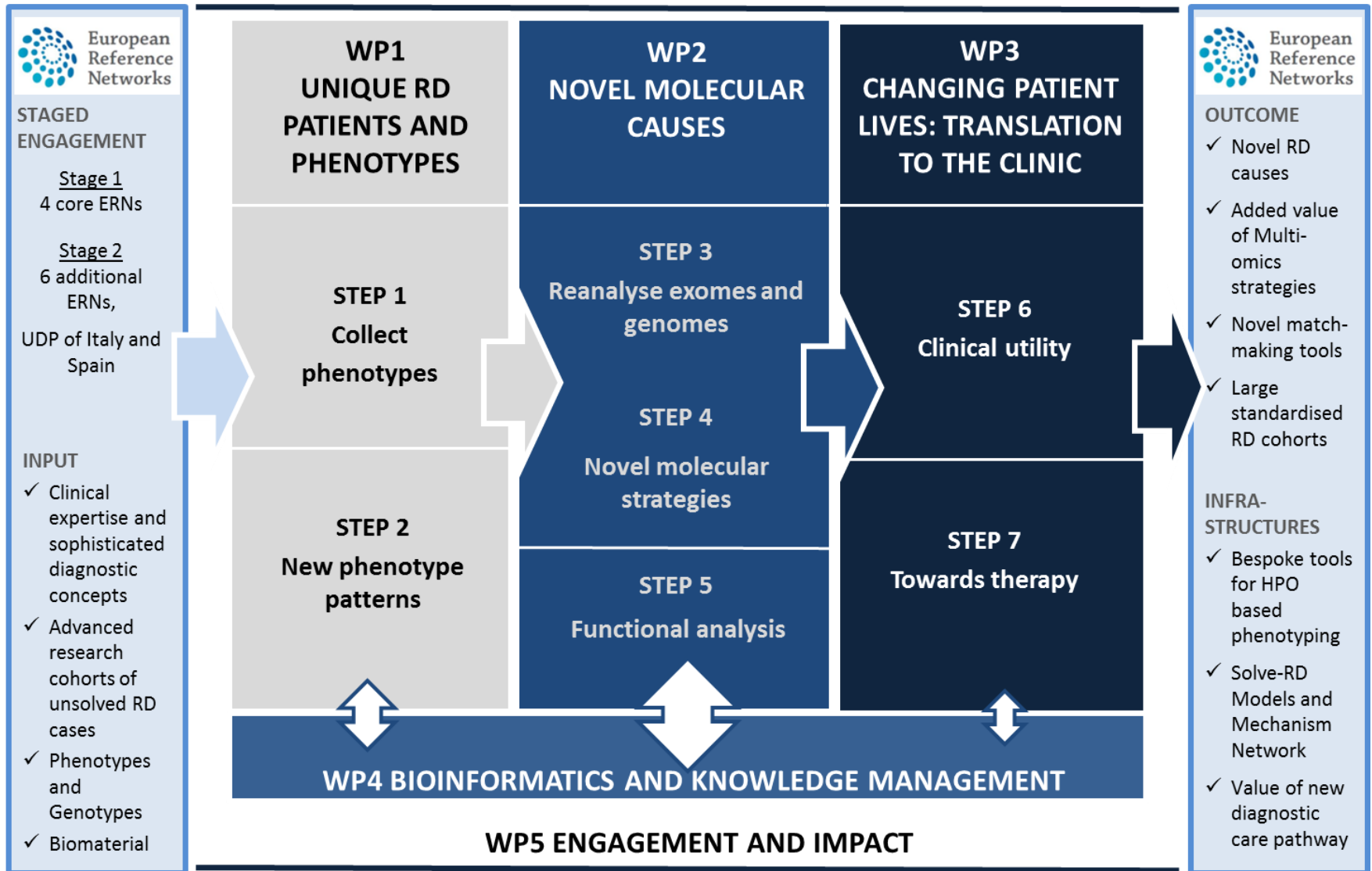
NAGEN – Proyecto Genoma Navarra

NAGEN es un proyecto piloto I+D+i promovido por el Centro de Investigación Biomédica Navarrabiomed en colaboración con la Dirección del Complejo Hospitalario de Navarra (CHN) para la implementación de la tecnología de secuenciación de Genoma Humano Completo (WGS) para su uso clínico en pacientes del CHN.

El objetivo de NAGEN es identificar las alteraciones del genoma de pacientes con más de 170 tipos de enfermedades raras o con ciertos tipos de cáncer de posible origen genético, cuya alteración genética causal no ha podido ser determinada a pesar de haberse realizado las pruebas genéticas indicadas para ello.

[Empiece hoy](#)

SolveRD



ELIXIR - the European life-science Infrastructure for Biological Information

ELIXIR consolidates Europe's **national centres, services, and core bioinformatics resources** into a single, **coordinated infrastructure**.



Platforms

Tools
Interoperability
Data
Compute
Training



Use Cases

Human Data
Marine Metagenomics
Rare Disease
Plant science



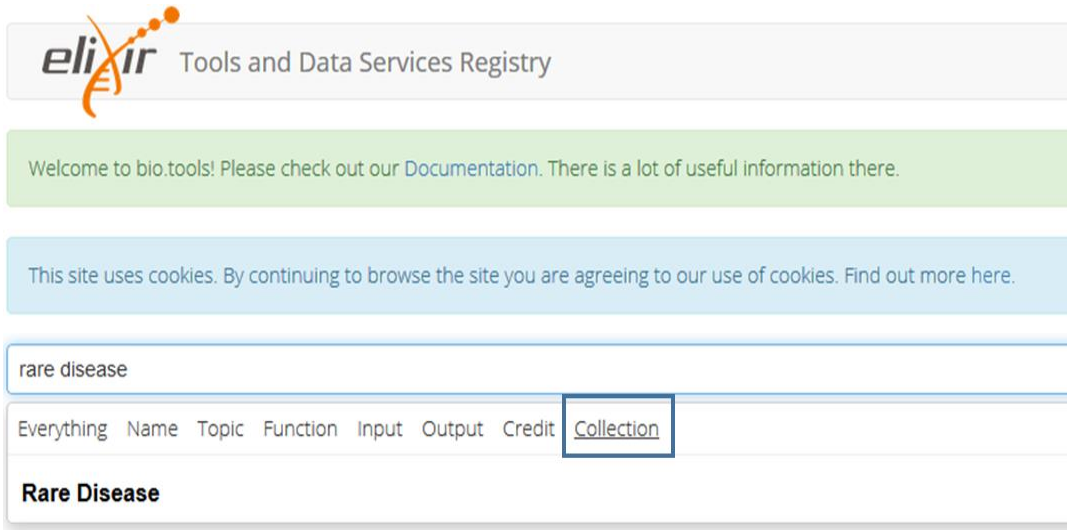
5 year project
2015-2020

ELIXIR-EXCELERATE

Fast-track ELIXIR implementation
and drive early user exploitation
across the life-sciences



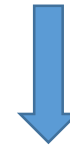
ELIXIR registry of data resources – bio.tools



The screenshot shows the bio.tools website header with the ELIXIR logo and the text "Tools and Data Services Registry". Below the header is a green welcome message: "Welcome to bio.tools! Please check out our Documentation. There is a lot of useful information there." A blue cookie consent banner follows: "This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here." Below the banner is a search bar containing the text "rare disease". Underneath the search bar is a navigation menu with tabs: "Everything", "Name", "Topic", "Function", "Input", "Output", "Credit", and "Collection". The "Collection" tab is currently selected and highlighted with a blue border. Below the navigation menu, the text "Rare Disease" is visible.



List all tools useful
for RD research



Ongoing work:
Link with benchmarking
activities (WP2)



**bio.tools
API**

Ongoing work:
RD-connect collection to pull up
to date information concerning
the tools integrated in the GPAP



ELIXIR registry of data resources

 Registry [Home](#) [Add content](#) [Docs](#) [Log in](#) [Sign up](#)

Welcome to bio.tools! Please check out our [Documentation](#). There is a lot of useful information there. ✕

This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. [Find out more here.](#) ✕

Everything: diseasecard ✕ Search 

Enter search query, e.g. "Proteomics", "Sequence alignment", "BAM".

Previous **1** Next

RESULTS SECTION

Sort by  Display as

Diseasecard: Rare disease research portal

<http://bioinformatics.ua.pt/diseasecard/> → Link to the resource

Diseasecard is an information retrieval tool for accessing and integrating genetic and medical information for health applications. Resorting to this integrated environment, clinicians are able to access and relate diseases data already available in the Internet, scattered along multiple databases.

[Data retrieval](#), [ID retrieval](#), [Query and retrieval](#), [Genotyping](#), [Diffraction data integration](#)

Brief description

Updated 13 days ago
Added 4 months ago
Topic Rare diseases, Pathology, Molecular interactions, pathways and networks, Medical informatics, Database management
Tool Type Database portal
Collection Rare Disease

Search browser: type a resource/
tool of interest

Tool name and link to
more detailed information



Diseasecard: Rare disease research portal

[Rare diseases](#) [Pathology](#) [Molecular interactions, pathways and networks](#) [Medical informatics](#) [Database management](#)

version ▾

[Database portal](#)

[Rare Disease](#)

Diseasecard is an information retrieval tool for accessing and integrating genetic and medical information for health applications. Resorting to this integrated environment, clinicians are able to access and relate diseases data already available in the Internet, scattered along multiple databases.

<http://bioinformatics.ua.pt/diseasecard/>

Links to useful
information concerning
the tool.

[Data retrieval](#) ▾
[ID retrieval](#) ▾
[Query and retrieval](#) ▾
[Genotyping](#) ▾
[Diffraction data integration](#) ▾

Reference / source (direct link
to pubmed entry)

Publications

[PUBMED](#) ▾

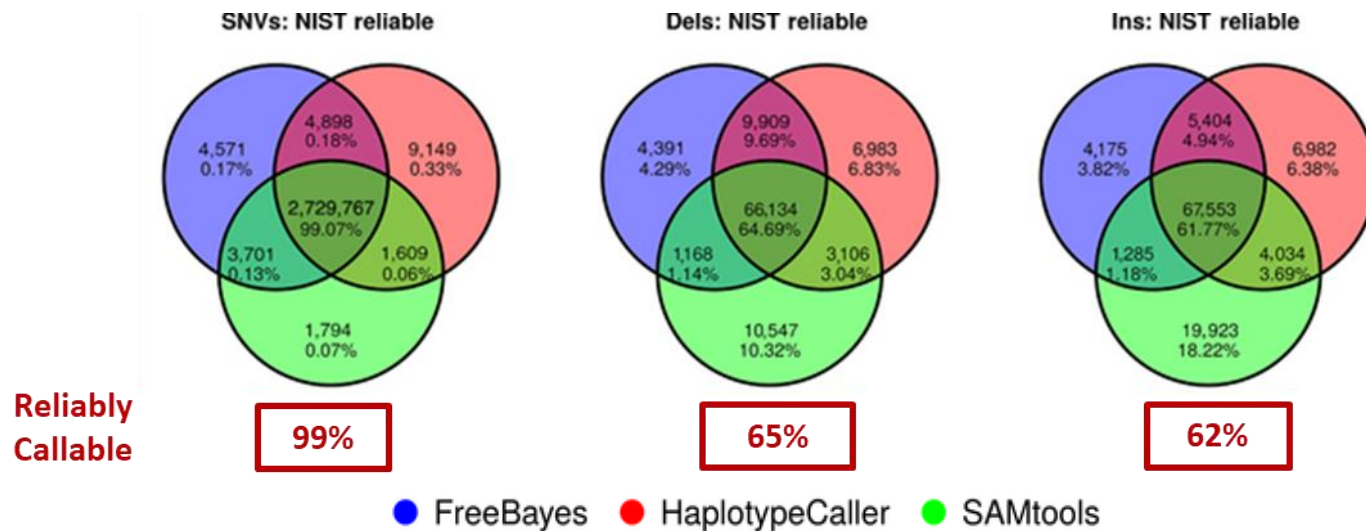
Other actions

Actions

[Request editing rights](#) ▾ [Request ownership](#) ▾



Benchmarking and standardizing genomics pipeline



Laurie *et al.* Human Mutation, 2016



USEFUL FOR THE RD COMMUNITY
-> RD-connect variant calling pipeline

- Working in collaboration with the RD communities to implement the [adequate quality reporting standards](#)
- **ELIXIR IS:** "Development of Architecture for Software Containers at ELIXIR and its use by EXCELERATE use-case communities"

ELIXIR Implementation Study: Integration of ELIXIR-IIB in ELIXIR Rare Diseases activities

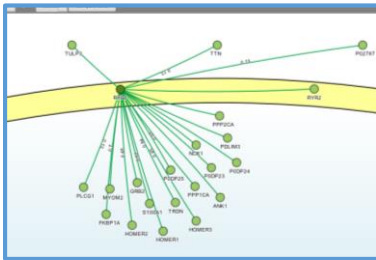


Filters: PRESET FILTERS, RESET, SHARE, RUN QUERY

Variant Type: high moderate Population: exac Genes: gene-list

Protein A (Gene name)	Protein B (Gene name)	Protein B: UniProt id	Type of interaction	Mentha score
RYR1	TRDN	Q13061	direct interaction	0.843
RYR1	FKBP1A	P62942	direct interaction	0.702
RYR1	HOMER1	Q86YM7	direct interaction	0.623

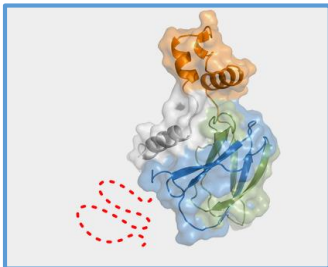
Chr	Pos	dbSNP	Ref	Alt	Candidate	QY ⁵⁰⁰⁰⁰⁰	INDEL	Gene Name	Effect impact	ClinVar	CADD	SIFT	PP2	MT	ExAC	1000GP AF	Internal Freq
12	32994058	rs147240502	A	C	TAC	A/C		PKP2	MODERATE	255/32/2	25.6	D	D	D	0.0047	0.0027	0.002009
16	15844048	rs111404182	G	A	TAC	G/A		MYH11	MODERATE	0/0	33	D	D	D	0.0005	0.0004	0.001211
19	39062815		G	C	TAC	G/C		RYR1	MODERATE		24.4	D	D	D	NA	0	0.000192



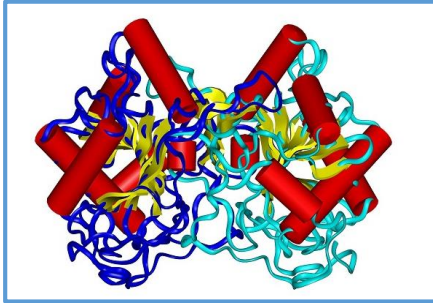
MINT / MENTHA (protein interaction databases)



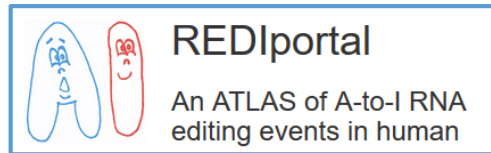
disease-gene Association database



VHLdb: A database of von Hippel-Lindau protein interactors and mutations



GALT-PROTEIN-DB 2.0



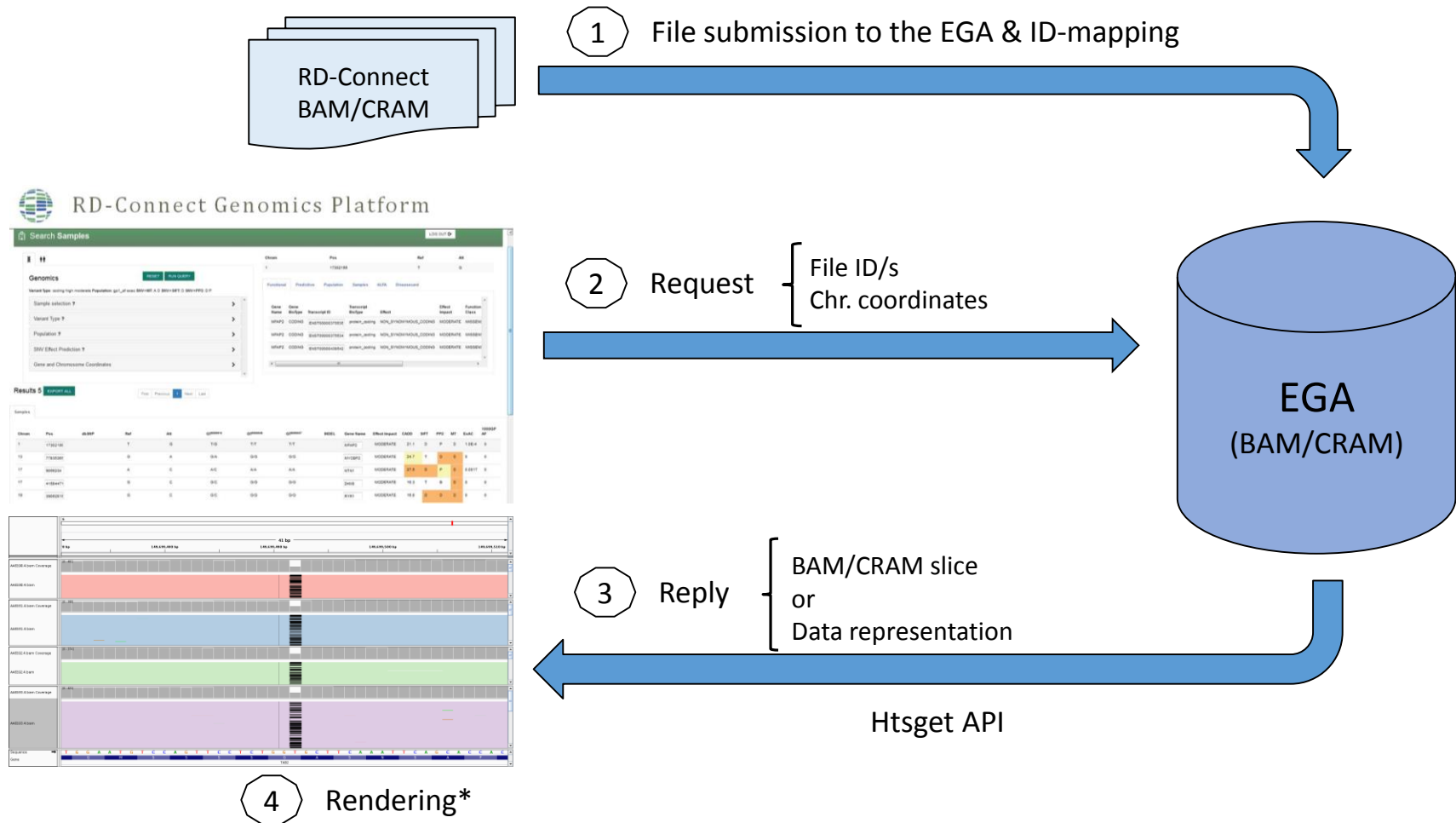
REDportal
An ATLAS of A-to-I RNA editing events in human



Human Mitochondrial DataBase



ELIXIR Implementation Study to visualize data deposited at the EGA in real-time



Secure and encrypted connections

*Broad's IGV screenshot

Adoptable by other projects and data types



Acknowledgements



WP1: Coordination
Ivo Gut / Hanns Lochmüller
(CNAG-CRG)

WP2: Patient registries
Domenica Taruscio (ISS and EPIRARE)

WP3: Biobanks
Lucia Monaco
(Fondaz. Telethon & EuroBioBank)

WP4: Bioinformatics
Christophe Béroud
(INSERM Marseille)

WP5: Unified platform
Ivo Gut (CNAG Barcelona)

WP6 Ethical/legal/social
Mats Hansson (Uppsala)

WP7: Impact/Innovation
Kate Bushby
(Newcastle and EUCERD/ EJARD)

IMIM

Luis Pérez-Jurado

CLÍNIC

Antònia Ribes

Montse Milà

Francesc Cardellach

HSJD

Francesc Palau

Rafael Artuch

VHIR

Alfons Macaya

Ramón Martí

Eduardo Tizzano

IDIBELL

Aurora Pujol

Conxi Lázaro

HSP

Pia Gallano

ISGlobal

Juan R. González

UB

Daniel Grinberg

CNAG

Sergi Beltran



Ángel Alonso

Sara Pasalodos

Josefa Salgado

Joaquín Dopazo

David Gómez

S. Beltran

S. Laurie

D. Piscia

L. Matalonga

A. Papakon.

F. Camacho

D. Picó

R. Tonda

J.R. Trotta

G. Parra

C. Luengo

J. Camps

I. Martinez

D. Ovelheiro

J. Protasio

M. Ingham

E. Casals

R. Alcántara

I. Gut

M. Gut

M. Bayès

B. Fusté

CNAG lab

CNAG admin



BBMRI – Large Prospective Cohorts

