# MGvizPM: Precision Medicine web reports made 'siimple'

David Gómez-Peregrina<sup>1,2,3,8</sup>, José M. Juanes<sup>1,3,5,7,8</sup>, Eva Panadero<sup>2</sup>, Santiago Gala<sup>2</sup>, Vicente Arnau<sup>5,9</sup>, MGViZ Felipe J. Chaves<sup>3,4,7</sup>, Ana Bárbara García-García<sup>3,4</sup>, Pablo Marín-García<sup>1,2,5,6,8</sup>



www.mgviz.org





VNIVERSITAT SUCSIC Institute for Integrative Systems Biology

E ÖVALÈNCIA CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Institute for Integrative Systems Biology

I<sup>2</sup> SysBio

1. Medical Genomics Visualization Group mgviz.org

2. Kanteron Systems S.L.

HL7 standard protocols.

3. Unidad de Genómica y Diagnóstico Genético UGDG - INCLIVA Valencia

4. CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM) 5. Institute for Integrative Systems Biology I2SysBio,

8. The Siimple Open-Source Organization siimple.xyz 9. Departamento de Informática, Escuela Técnica Superior de Ingeniería

6 Departamento de matemáticas y Ciencias Experimentales, Universidad

Medical Genomics Visualization Group (MGviz), Siimple OSS, Seqplexing and Kanteron Systems have jointly developed NGS data analysis workflows that create automatic technical reports for precision medicine with fully integrated QC and LIMS procedures. Our genetic and pharmacogenetic data can be easily integrated in HIS systems and use

We have developed a full suit of open source tools in Python, R and MERN stack for clinical bioinformatics as a service. These tools include serving variant annotation, interactive selection tools, reannotation and automatic clinical reports generation.

Católica de Valencia UCV,

7. Sequencing Multiplex S.L.

(ETSE), Universidad de Valencia.

We are doing trials to deploy this service in a cloud platform for creating a service for analyzing customized NGS gene panels and exomes in a clinical context for diabetes, cancer and mental disabilities.

# MGviz NGS pipelines QC/QA **PROCESS** FastQ **FASTQ** Validator (raw) Seqtk FastQC **Qual Trim FASTQ** (processed) BWA BAM (raw) **BAMclipper SAMtools** Picard Other... Mark **Duplicates** Artifact Cleaner Picard Bedtools BAM Coverage (processed) **Variant Caller** Multi Variant Calling: VCF **VCEnsemble** (raw) Norm Decomp **BCFtools** Stats **BCFtools** Filter VCF (processed) Annovar/VEP VCF (annotated)

Pipeline automation following Common Workflow Language (CWL) implementations like NextFlow. Also, there are modules to implement pipelines with Snakemake or whatever workflow manager and tools the customer desires.

## Summary

Clinical genomics is a reality, but better tools for creating reports without great effort and easy to interpret actionable variants are needed.

Here we present a professional platform for generating semi-automatic genetic clinical reports. Consistency and clarity of reports are needed in order to translate new clinical genomics insight into personalized healthcare and this tool helps clinicians and genetic counselors to create these reports.

## References

- https://doi.org/10.1093/bib/bbx144
- https://doi.org/10.1038/s41375-017-0007-7
- 3. <a href="http://www.nature.com/doifinder/10.1038/gim.20">http://www.nature.com/doifinder/10.1038/gim.20</a> 15.30
- 4. <a href="https://dx.doi.org/10.1016%2Fj.jmoldx.2016.10.002">https://dx.doi.org/10.1016%2Fj.jmoldx.2016.10.002</a>

#### MGvizPM workflow:

- 1. Upload SE/PE FASTQ files.
- 2. Select the configuration parameters to run a NGS pipeline.
- 3. As the samples are analyzed, the results can be consulted.
- 4. QC plots can be visualized and are interactive to explore the results.
- 5. Filter actionable and variants of interest, annotate additional variant information (stored in the user history), consult the integrated data from different reference databases and select the variants to report.
- 6. Fill the report to generate the clinical-technical report, download the DOC and modify as desired.

#### **Customized NGS pipelines:**

- → Optimized for each experimental design and sequencing library protocol.
- → Standard and customized QC/QA steps.
- → Atomized pipelines to facilitate extension.
- → Scalable for production.

Seqplexing TP53

Seqplexing TP53

**€**FastQC Report

Per base sequence qua

Per sequence quality scores

Sequence Length Distribution Sequence Duplication Levels

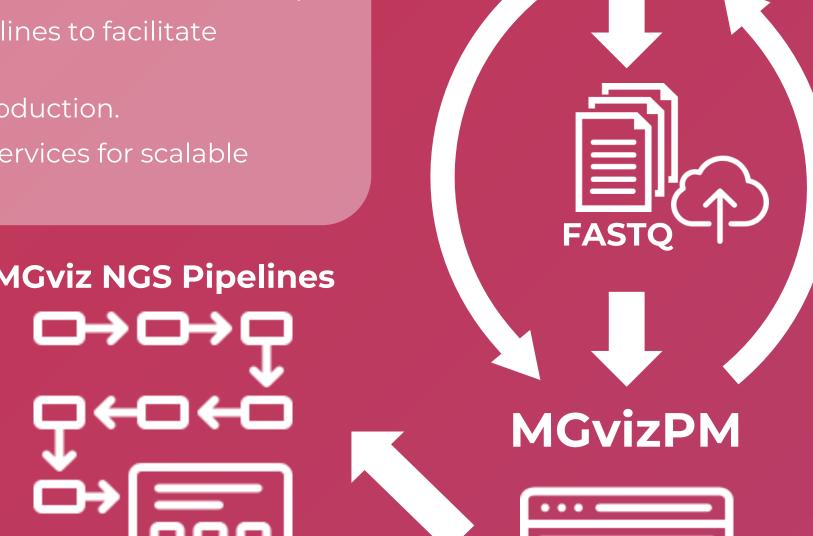
Overrepresented sequences

Per base sequence content Per sequence GC content Per base N content

Home / Analysis Overview

test3pmg

→ Docker microservices for scalable deployment.

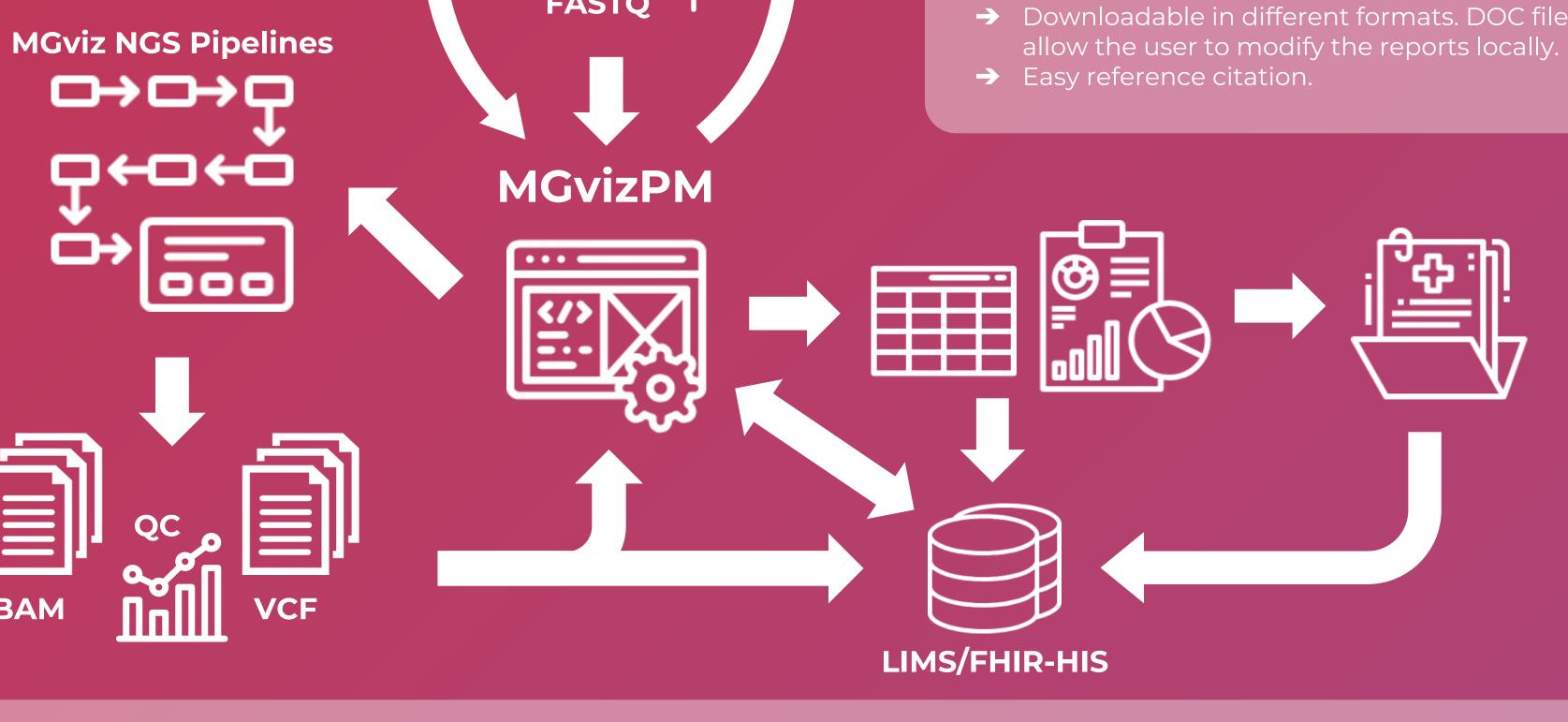


Variant prioritization:

- → Variant annotation includes clinical, evolutionary and pharmacogenetics annotation for prioritization.
- > Extended clinical information of each variant include databases such as COSMIC, ClinVar, InterVar (and other genic/project specific databases).
- → Variant calling QC metrics are included to discard possible false-positive variants.
- → All the information of the samples and the decisions made by the user are saved in a LIMS system for reproducibility, audits and access to the history. Data storage is compatible with FHIR-HIS system integration.

#### Automatic report generation:

- → Fast and intuitive access to clinical database information.
- → Customized templating.
- → Embedded QC results.
- → Downloadable in different formats. DOC files



#### **USE CASE:** OGBS-TP53 http://tp53report.seqplexing.com:3001



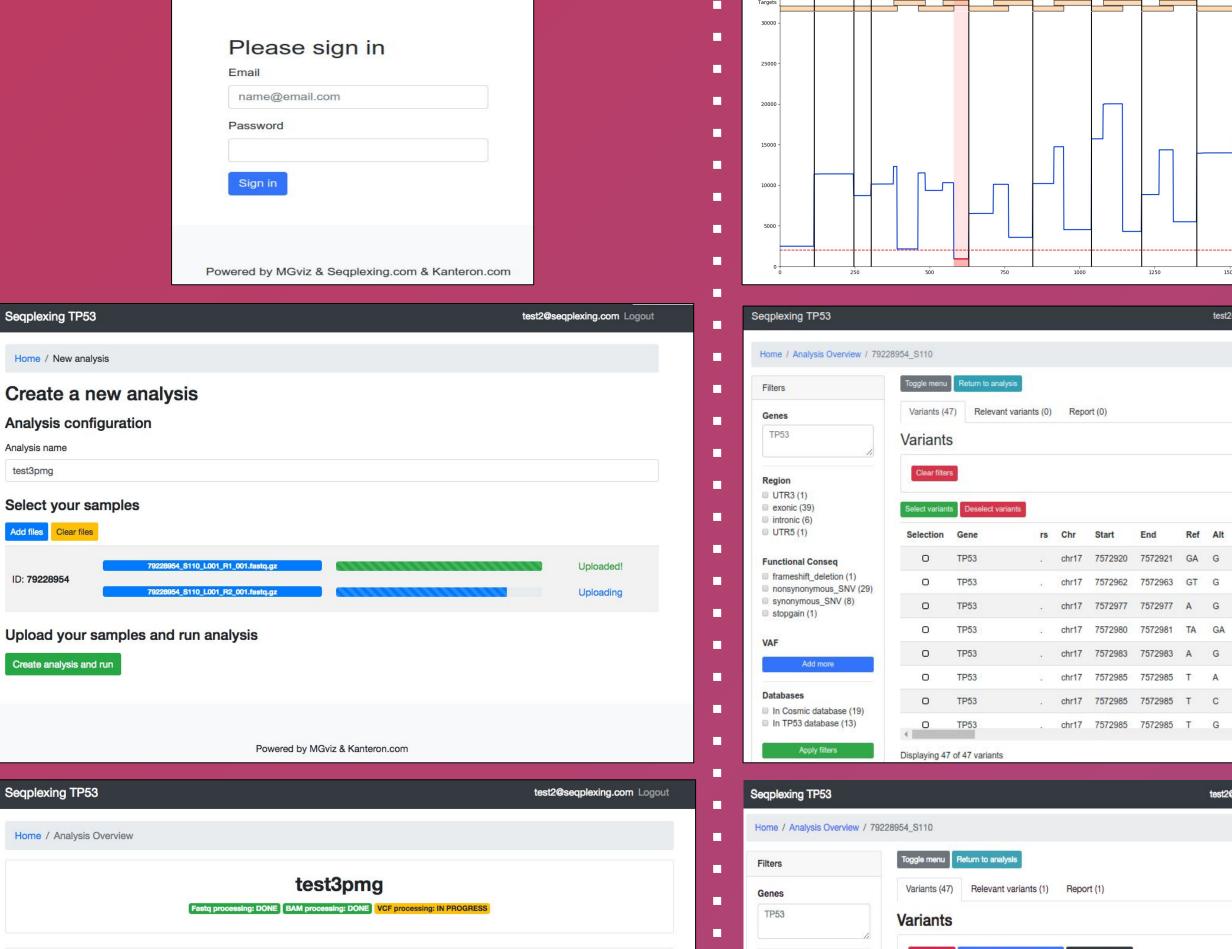
Variants (47) Relevant variants (1) Report (1)

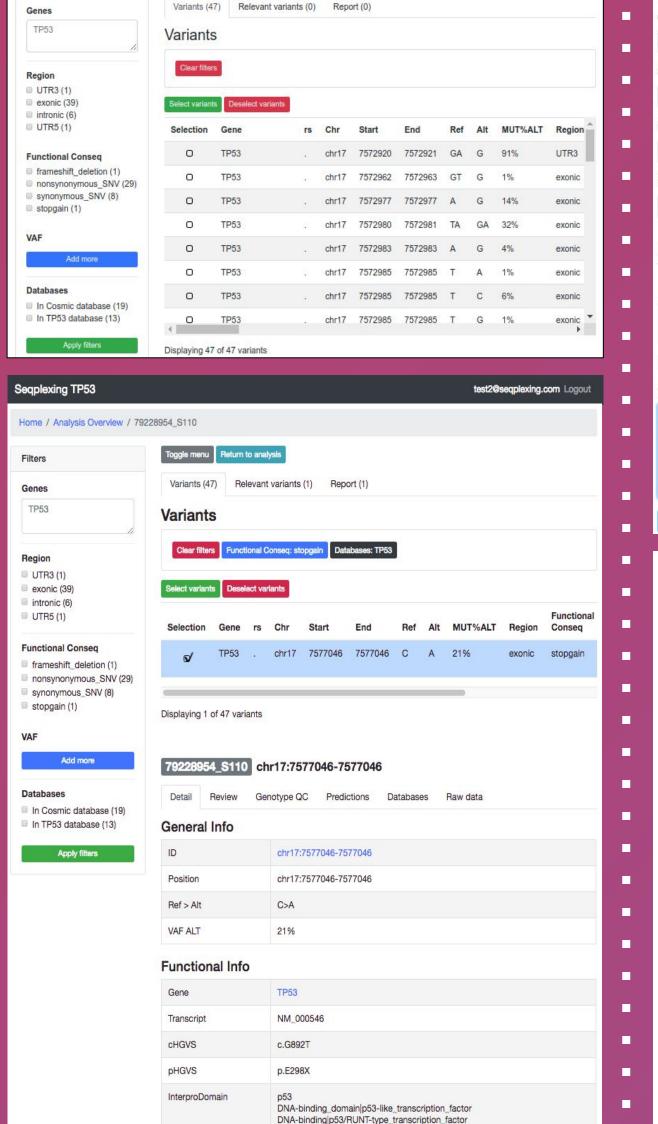
Toggle menu Return to analysis

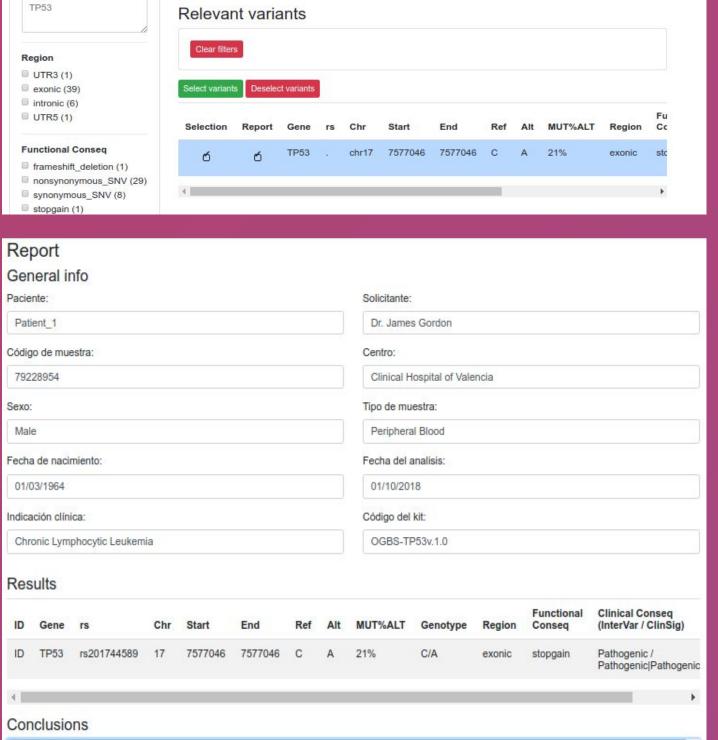
test2@seqplexing.com Logo

Seqplexing TP53

Home / Analysis Overview / 79228954\_S110







resistance.	700 - V - C. P. P. C. P. P. C. P. C. P. C. P. C. P. P. P. C. P. P. P. C. P.	2008 - 18 C - 19 Marie (1900 - 190 - 1906 - 190	s and an independent work to be a second of the second		is related to chemo-therapy and immun	
Generate and d	ownload report					
Laboratory X	Mutational A	nalysis of the TP53 G	eeqplexing	Laboratory X  Appendix  General Info		eeqplexii
Performed by Requested by				ID	rs201744589	
Dr. James Gordon			Hospital 2		17:7577046-7577046	
Address 2		L More and the Co	Address 2		C>A	
		, marcos 2			C/A	
Patient Name:	Mr. Patient1	Date of sample collection:	2018-09-24	Genotype VAF ALT	21%	
	38300		9000.0000000000000000000000000000000000			
Diagnosis:	CLL	Date sample delivery:	2018-09-25	Functional Info		
Type of material:	Peripheral blood / Genomic DNA	Result issue:	2018-10-01	Gene	TP53	
				Transcript	NM_001276761	
- Analytical method: The entire coding region of the TP53 gene including splice sites was analyzed using small amplicon targeted next-generation sequencing. Genomic DNA was amplified using high-fidelity Polymerase with primers specific for TP53 exons 1-11 (Sequencing Multiplex S.L kit OGBS-TP53v1.0). The library was prepared from small amplicons with Nexteral adapters added in a second PCR step (following OGBS-TP53v1.0 specifications). Sequencing was performed using MiSeq Reagent Kit v2 (300 cycles) on the Illumina MiSeq instrument. For variant detection, MozivNGS-reporter platform v0.9 (MGviz.org) was used. Reads were aligned to the human reference genome (GRCh37.d5) and variants were annotated using ANNOVAR. Coverage: minimum: 3047, mean: 12173 per base.				c.HGVS	c.G775T	
				p.HGVS	p.E259X	
				Interpro_domain	p53 DNA-binding_domain p53-like_transcription_factor DNA-binding p53/RUNT-type_transcription_factor DNA-binding_domain	
				Cosmic		
Detection limit: >0.2% variant allele frequency.     Reference sequences: NM_001276761. Mutation nomenclature follows the Human Genome Variation.				ID	COSM3723940	
Society (HGVS) guidelines.					COSM121080 COSM1646820 COSM10710	
Result  Number of reported mutations: 1				Occurrence	3(salivary_gland) 3(prostate) 6(breast) 3(akin) 2(pancreas) 3(ovary) 3(urinary_tract) 3(ivar) 1(large_intestine) 2(NS)	
1) TP53: c.G775T, p.Giu259*  - Mutation type: Stop gain mutation - Allelic proportion (%VAF): 21% - Clinical consequence: Pathogenic						
Interpretation  Stop gain variant in exon 8 within the DNA-binding domain of TP53. According to the locus specific databases IARC-TP53 website, COSMIC and Clinvar, this is a truncating mutation related to poor outcome. Additionally, it is related to chemotherapy and immunotherapy resistance.					33(Jung) 1(stomach) 15(upper_aerodigestive_tract) 7(oesophagus)	
III I						
Disclaimer For research use of	nly. Data is provided as it is	s reported at the genetic research datab	ases (see appendix for detailed			
list and versions). Seq	plexing is not responsible for	r the accuracy of the information provide or oncologist. Neutral polymorphism and	d by those databases.			
	pe provided upon request.	s sociogist. Necesal polymorphism and	evinger variants are not included			
Authorized signatures				Coverage QC		
			AMPERION SYSTEMS		Powered by:  MGvlz.s	MANUFACH SYST

top gain variant in exon 5 within the DNA-binding domain of TP53. According to the locus specific databases (JARC, TP53 website), COSMIC and