BIOINFORMATICS DATA ANALYSIS

PORFOLIO AT VIVIA BIOTECH
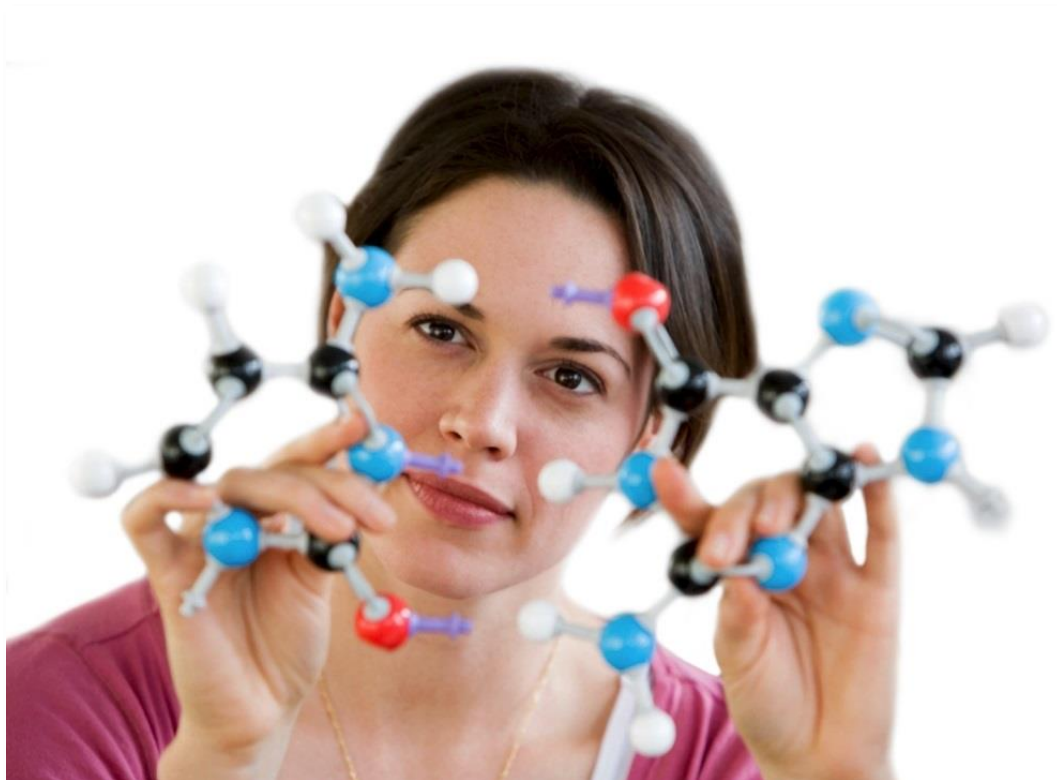
Table of contents

1. Introduction

1.1. Objective

MOLECULAR BIOLOGY DATA IS BECOMING MORE AND MORE RELEVANT IN DRUG DEVELOPMENT AND PRECISION MEDICINE STUDIES. MOST OF THE PROJECTS NOWADAYS ARE CONCEIVED TO INCLUDE IT AS A SUBSTANTIAL PART.

VIVIA BIOTECH IS DEFINITIVELY AWARE OF THIS FACT AND THUS WE HAVE BEEN INCLUDING, DURING THE LAST YEARS, A SET OF EXPERIMENTAL PROCEDURES TO COMPLEMENT OUR PORFOLIO OF INMUNOPHENOTYPING AND PHARMACOLOGICAL STUDIES.

RELEVANCE AND IMPORTANCE OF MOLECULAR BIOLOGY DATA GOES IN PARALLEL WITH THE EXTREMELY HIGH COMPLEXITY IN THE DESIGN, EXECUTION AND INTERPRETATION OF RESULTS.

WE ARE DELIGHTED TO ANNOUNCE THAT FROM THIS MOMENT ON, WE ARE ALSO OFFERING A LIST OF SERVICES IN THE DATA ANALYSIS GENERATED FROM THS TYPE OF ASSAYS. WE HAVE BEEN WORKING TO SET UP A QUALIFIED TEAM IN ORDER TO PROVIDE YOU WITH:

- *CONSULTANT SERVICES TO ASSIST IN EXPERIMENT DESIGN AND SELECTION OF DATA ANALYSIS PROCEDURES AND RESULTS VISUALIZATION MODES.*
- *EXECUTION OF ANALYSIS AND DELIVERY OF RESULTS ACCORDING TO EXISTING STANDARD PACKAGES OR PRE-DEFINED CUSTOM PROCEDURES.*

THIS DOCUMENT AIMS TO PROVIDE YOU WITH A VISION OF THE PRESENT AND FUTURE OF BIOINFORMATIC DATA ANALYSIS AT VIVIA AND A COMPREHENSIVE UNDERSTANDING OF ALL THE NEW OPTIONS THAT WE MAKE AVAILABLE FOR THE ANALYSIS OF YOUR DATA.

1.2. Vivia Services: Evolution and Extension

- *PharmaFlow platform*

Vivia Biotech provides precision medicine solutions for patients, clinicians and pharmaceutical companies, through a high technology system applied for ex vivo testing that preserve Native Environment (NE) of cells populations. Vivia's PharmaFlow platform allows to perform complete pharmacological characterizations of *ex vivo* drugs effects using patient's samples. Such ex vivo drug profiling studies are carried out by integrating immunophenotyping and flow cytometry technologies together with robust statistical and modelling analysis and the use of effective graphical and visualization tools. In parallel to these studies, Vivia can streamline clinical development by generating predictive *ex vivo* native environment assays and identifying reliable biomarkers.

- *Innovation and Development*

Success in precision medicine strongly depends on integrating multiple variables coming out from functional assays as well as high quality genetic and molecular data from well annotated patient cohorts. We focus on coupling pharmacology and molecular biology with comprehensive clinical data.

For this reason, Vivia has gradually updated the catalogue of available instrumental techniques and molecular biology assays in addition to extended applications in flow cytometry. A differential critical aspect of Vivia's service to pharma companies is the dedication and resources put on innovating and developing new applications. This drives to new assay types and data analysis technologies that we make available to our customers to provide enhanced studies to complement and correlate the ex vivo drug profiling.

- *New challenges in Bioinformatic area*

Recently Vivia has strengthened the bioinformatic area by recruiting a PhD expert in molecular biology with a master in bioinformatics with experience and knowledge in clinical research, particularly in Multiple Myeloma. In addition, we also count with a senior Bioinformatician from the Spanish National Cancer Center (CNIO) with more than 20 years of experience and a high number of excellent publications in the field.

We present herein the new portfolio of products for genetic and molecular data analysis that we are offering, either associated or independent to those already available in phenotyping and pharmacological studies. The conjunction of the expertise already achieved at Vivia in drug

profiling and precision medicine analysis together with the knowledge and enthusiasm from the new incoming members, has resulted in a highly motivated, well balanced and qualified team of professionals. We´re ready to take on the challenge of performing the integrated bioinformatic analysis of your data with the highest quality.

1.1. Key aspects of bioinformatic analysis at Vivia

Good results analysis, enriched with graphs, diagrams and tables, may help a lot our customers to get a better interpretation of the results generated with their compounds, and is undoubtedly, a great value of our services. We always have focused on this goal, dedicating great efforts in fitting customer's needs and implementing all requirements in a timely manner.

In addition, expertise reached though the results analysis and the fluent collaboration with our customers, has allowed us to provide valuable assistance in experimental design of new projects, which is a critical aspect to achieve better, meaningful results.

To make that possible, it is essential to count, maintain and coordinate the following aspects and put them all together in every project developed at Vivia to ensure the highest quality in the analysis. So far, we have considered them intensively and now with the service extension we plan to emphasize on them even more.

1.1.1. IT and software development

We have wide expertise in information technologies and software development required to set up and maintain systems that support precision medicine services. Multiple applications are configured and extensively used on a daily base. Among them, we regularly use LIMS system as well as sophisticated statistical tools, such as R, NONMEM, SAS or Statistica.

We commonly develop scripts and programs for different purposes using R studio, Linux scripts, Iron Python, VBA and PLSQL. All our data is stored in Oracle databases that help us to meet our requirements in several areas like performance, scalability, back-ups and security. Our IT infrastructure include dedicated servers, networks and workstations that offer a high performance, safe and reliable environment. Moreover, Microsoft Sharepoint platform provide several powerful tools that improve work performance and quality such as documents and projects management.

1.1.2. Graphical and visualizations tools

As pointed above, results graphs and visualizations are a key aspect of project analysis. For this purpose, we have been using Spotfire from Tibco (https://www.tibco.com/solutions/analytics-pharma-biotech), a well-known application, especially in pharma and biotech companies, that is at the same time, a very powerful, scalable and flexible tool to run complex analytical procedures with excellent performance and computational efficacy, and on the other hand a very friendly and intuitive interface to assist scientist in data (or big data) analysis, results interpretation and decision-making processes.

Recently, we have acquired a license of Spotfire Web Server, a tool that will allow to offer dynamic results reports to our customers with the possibility to interact with graphs and visualizations through a web player application. We plan to roll it out before the end of the year. In addition, we are evaluating the PerkinElmer OmicsOffice suite for Spotfire, specifically designed for NGS data analysis (http://www.perkinelmer.com/product/omicsoffice-omicsoffice).

1.1.3. Results interpretation and discussion

Results interpretation and discussion is definitively a key aspect of the Vivia services. We have realized over the last years, that a deeper comprehension of the projects as well as the results generated, turn back into higher quality of the analysis and better feedback from our customers.

We dedicate a lot of efforts for this purpose and follow up with it as an intense team work, and thus, we regularly set internal meetings at different points in projects life time, since initial phases with experiments design, to final analysis of results and report generation. Once project reports are sent, another round in results analysis starts where customers feedback and requirements for new analysis or editions are always welcome.

2. Services for phenotyping and pharmacological studies

Below is a schematic summary of the type of results analysis developed at Vivia during the last years as part of the services provided to pharma companies in drug profiling studies. The list is very dynamic, changing often and always supported by our customers' feedback.

- PHENOTYPING AND EXPRESSION LEVELS
    - Quantitative phenotypic analysis of cellular subpopulations according to expression of markers identified by flow cytometry.
    - Estimation of expression level of markers identified by flow cytometry
- PHARMACOLOGICAL EFFECT
    - Analysis of pharmacological cytotoxic effect of single drugs over cell populations and subpopulations identified by flow cytometry.
        - Dose-response curve fitting
        - Pharmacodynamic parameters estimation
        - Categorical analysis according to different variables (experimental conditions, subpopulations identified by different markers, etc.)
        - Comparative analysis among drugs tested under same experimental conditions.
    - Analysis of antiproliferative effect of single drugs over populations identified by flow cytometry in cell proliferation assays.
        - Same analysis as in cytotoxic effect assays.
    - Characterization of population pharmacological effects of drugs (cytotoxic or antiproliferative) by the development and application of Pharmacodynamic Mixed-Effects Non-linear population models. Estimation of population typical parameters and associated inter-patient variability together with model quality parameters and errors estimations.
    - Synergy measurement. Estimation of the level of synergy in the pharmacological effect (cytotoxic or antiproliferative) observed in combination experiments using two drugs.
        - Empirical estimation using Combination Index. Color maps for estimation of optimal synergistic dose-ratios.
        - Modelling using Interaction Surface approach. 3D Surface plots

3. Services for Molecular biology assays

3.1. Available Services

We believe that a success personalized medicine approach needs to simultaneously collect and integrate biological data from different scientific sources fed by the combination of various experimental technologies. For this reason, we have worked hardly not only to integrate in our services portfolio the conventional molecular diagnostic test, but also to include the most powerful molecular biology technique, Next Generation Sequencing (NGS).

We provide an integral solution for NGS experimental procedures that comprises: DNA-seq (WGS, WES, Targeted sequencing, BS-seq, ChIP-seq), and RNA-seq (WTS, mRNA-seq, miRNA-seq, RNA Targeted sequencing, Single-cell RNA-seq, Quantseq 3' mRNA-seq, RIP-seq).

3.1.1. Conventional molecular assays

- *DNA/RNA isolation*

   Extraction of DNA and RNA is the basic method used in molecular biology and they are the primary step for many downstream applications. We have an optimized method to isolate DNA and RNA, both manually and automatically, with high concentration and integrity.

- *DNA/RNA quantitation and quality check*

   We provide different quantification methods to accurately measure DNA and RNA concentration, including fluorometric and microfluidic quantitation by Qubit 5.0 and Agilent 2100 bioanalyzer respectively. The last option also allows the evaluation of RNA integrity.

- *Absolute and relative gene expression quantification by RT-PCR*

   This methodology allows to determine the number of target cDNA molecules by comparison with DNA standards using a calibration curve or determine the relative amount by comparing to the amount of control reference gene.

- *Verification of DNA mutations by capillary electrophoresis*

   Sanger sequencing is still the gold standard technique to validate NGS results due to its simplicity and reliability. We offer a complete workflow for validating variants detected by NGS technology.

3.1.2. NGS assays

We offer extensive NGS assay options to provide researchers with a full project customization to get valuable data according to each particular interest. A great variety of library kits compatible with the two most used sequencing platforms, Ion Torrent and Illumina, are available. We also offer a complete consulting service to guide our customers through the experimental design and the bioinformatic analysis.

- *DNA-sequencing approaches*

Whole Genome Sequencing (WGS)

This method is a comprehensive approach for analyzing entire genomes. WGS is able to determine the full genome sequence of any species, including the mitochondrial DNA, which is widely known to be involved in many inherited and genetic disease, as is the case of cancer.

This method allows a comprehensive view of the genome, particularly interesting for the identification of single nucleotide variants, insertions, deletions, copy number variants, and other larger structural aberrations involved in abnormal gene expression patterns or altered signaling mechanism.

Whole Exome Sequencing

WES is the NGS technique for sequencing the coding region of the genome. This approach enables to focus the resources on the genes that most likely lead to phenotypic changes, providing a high coverage of the 95% of the exons which harbor the majority of genetic variants associated with human diseases. This methodology provides huge amount of information at an affordable price, including the identification of single nucleotide variants, insertion, deletion and copy number variation.

Deep Targeted DNA sequencing

This methodology offers the possibility of sequencing specific regions of the genome in a rapid and cost-effective way. Focusing on critical genes or regions of interest enables to achieve higher read depth and so, identifying genetic alteration only present in minor subclones or cell subpopulations (<5%).

Single cell DNA sequencing

Single-cell RNA sequencing has emerged as an indispensable tool for revealing unknown cellular feature and deepen our understanding of biological systems. Genetic changes such as single mutation or copy number variation may be explored at the level of individual cells.

Thus, a very small number of cells is required, making this method useful for the genetic profiling of resistant clones or circulating tumor cells.

- ***RNA-sequencing approaches***

Whole Transcriptome Sequencing (WTS)

WTS provides essential insights into genome organization, expression and regulation. This methodology allows for the study of both coding and multiple types of non-coding RNA. Abundant RNA species can be removed with specific kits (Ribo-Zero).

Messenger RNA sequencing (mRNA-Seq)

mRNA-Seq provides transcriptional profiling of coding regions of the genome. It can be used to determine the structure of genes, splicing patterns and other post transcriptional modifications. The information extracted from a mRNA-Seq experiment also allows the detection of rare and novel fusion transcripts, the detection of mutations in coding regions, and measure gene and transcript abundance.

Non-coding RNA-sequencing

miRNA-Seq allows to study non-coding molecules responsible for the negative regulation of gene expression. Since the role of this small molecules was elucidated, its demand is growing rapidly.

RNA Targeted sequencing

This approach allows to focus exclusively on certain transcripts and provide increasing sensitivity in the regions of interest. This method improves the performance and the dynamic range compared to RNA-seq and reduces the cost, optimizing conditions when a limited number of genes need to be evaluated.

Single Cell RNA Sequencing (scRNA-Seq)

scRNA-Seq *has become one of the most interesting approaches to study the uniqueness of each cell and uncover information undetectable by conventional RNA sequencing. Expression changes and minor structural aberration may be explored at the level of individual cells. A very small number of cells is required, making this method useful for the genetic profiling of resistant clones or circulating tumor cells.*

Quantseq 3' mRNA-Seq:

Genome-wide analysis of gene expression based on 3' read mapping and 3' UTR quantification. QuantSeq provides a cost-efficient RNA-Seq approach that can be suitable for experiments with low quality RNA and FFPE samples.

RIP-seq

Allows to study RNA-binding proteins (RBPs), extracted through immunoprecipitation (RIP), to characterize all RNAs that were found to be bound in vivo by a given RBP.

- **Epigenomic NGS assays**

Chromatin Immunoprecipitation assay (ChIP-seq)

ChIP-seq is a powerful method to study the molecular interactions between proteins and DNA, such as transcription factors. This methodology combines a chromatin immunoprecipitation assay with NGS, overcoming the limitation of ChIP-ChiP by performing a genome-wide analysis. Elucidating protein:DNA interactions may help to understand important aspects regarding gene expression regulation.

DNA Bisulfite sequencing (BS-Seq)

BS-seq is considered a gold-standard technology for single-nucleotide-resolution DNA methylation detection. A large number of studies have demonstrated that DNA methylation plays an important role in regulating several physiological and pathological processes. Altered patterns of DNA methylation impacts various cellular processes involving DNA stability and gene expression, representing an attractive diagnostic and therapeutic target.

3.2.   Data analysis services

3.2.1.   Objective

The analysis of the data generated from NGS experiment is typically the bottleneck for researchers and clinical workers, due to its strong complexity and huge size.

The lack of published guidance leads to high degree of variability in data analysis, which sometimes generates inaccurate results. Additionally, a considerable number of bioinformatic tools are available to process and analyze raw sequencing data coming from NGS experiments, and the selection of the most suitable ones can directly impact the quality of the results. Therefore, bioinformatic pipelines are required to combine the best tools in a proper fashion to perform an efficient and robust analysis of the data.

In this regard, we have adopted widely tested bioinformatic pipelines to analyse NGS data from different platforms and experimental approaches, which implement the best practices described in the literature. We offer two different pipelines depending on the assay type, DNA-seq or RNA-seq. Each pipeline is especially designed to meet the computational operations needed for each NGS experiment.

3.2.2.   Bioinformatic pipelines adopted by Vivia

The bioinformatic pipelines that we have adopted implement the best practices and recommendations described in the literature, to ensure optimal NGS testing quality. In addition, we can adapt these pipelines to any experimental approach.

We offer two different versions depending on the assay type, DNA-seq or RNA-seq. Both pipelines have been previously used in papers published in high impact journals:

- *DNA-seq pipeline:*

  - *Ruiz Y et al. Mutational screening of newly diagnosed multiple myeloma patients by deep targeted sequencing.* **Haematologica***. 2018 Jun 28. PMID: 29954938.*
  - *Lynch CJ et al. The RNA Polymerase II Factor RPAP1 Is Critical for Mediator-Driven Transcription and Cell Identity.* **Cell Rep***. 2018 Jan 9;22(2):396-410.*
  - *Cuadrado A et al. The contribution of cohesin-SA1 to gene expression and chromatin architecture in two murine tissues.* **Nucleic Acids Res***. 2015 Mar 31;43(6):3056-67.*

- *Menezes J et al. CSF3R T618I co-occurs with mutations of splicing and epigenetic genes and with a new PIM3 truncated fusion gene in chronic neutrophilic leukemia. **Blood Cancer J**. 2013 Nov 8;3:e158.*

- *Menezes J et al. Exome sequencing reveals novel and recurrent mutations with clinical impact in blastic plasmacytoid dendritic cell neoplasm. **Leukemia**. 2014 Apr;28(4):823-9.*

- *Calvete O et al. A mutation in the POT1 gene is responsible for cardiac angiosarcoma in TP53-negative Li-Fraumeni-like families. **Nat Communications**. 2015 Sep 25; 6:8383.*

- ***RNA-seq pipeline:***

  - *Forsthuber A et al. CXCL5 as regulator of neutrophil function in cutaneous melanoma. **The Journal of investigative dermatology**. 2018.*

  - *Valiño-Rivas L et al. TWEAK increases CD74 expression and sensitizes to DDT proinflammatory actions in tubular cells. **PloS one**. 2018; 13(6): e0199391.*

  - *Garcia-Carpizo V et al. CREBBP/EP300 bromodomains are critical to sustain the GATA1/MYC regulatory axis in proliferation. **Epigenetics & chromatin**. 2018; 11(1):30.*

  - *Oldrini B et al. Somatic genome editing with the RCAS-TVA-CRISPR-Cas9 system for precision tumor modeling. **Nature communications**. 2018; 9(1):1466.*

  - *Metehan Cifdaloz et al. Saa3 is a key mediator of the protumorigenic properties of cancer-associated fibroblasts in pancreatic tumors. **PNAS** January 19, 2018.*

  - *. Systems analysis identifies melanoma-enriched pro-oncogenic networks controlled by the RNA binding protein CELF1**. Nature Communications** 8: 2249 (2017).*

  - *Montero JJ et al. Telomeric RNAs are essential to maintain telomeres**. Nat Communications**. 2016; 7: 12534.*

  - *Simón-Carrasco L et al. Inactivation of Capicua in adult mice causes T-cell lymphoblastic lymphoma. **Genes & development**. 2017; 31(14):1456-1468.*

  - *Olmeda D. et al. Whole-body imaging of lymphovascular niches identifies pre-metastatic roles of midkine. **Nature**. 2017; 546(7660):676-680.*

3.2.3.   First line analysis

We provide an integral solution for NGS experimental procedures that, on the one hand, covers the more common bioinformatic demands, and that, on the other hand, adds more specific downstream analyses aimed to answer more detailed questions.

Our pipeline explicitly support data from WES and WTS, since they are the most widely used NGS approaches, although it is also suitable for others sequencing strategies such as whole genome sequencing or targeted DNA and RNA sequencing by performing minor modifications.

- *Analysis performed/provided by our DNA-pipeline:*

  - *QC analysis*
  - *Read filtering*
  - *Read mapping*
  - *Variant and InDels calling (with/without germline fraction)*
  - *Variant and Indels annotation and priorization.*
  - *Copy Number variation identification (with/without germline fraction) and annotation.*
  - *Other structural variation detection (subject to experimental design)*
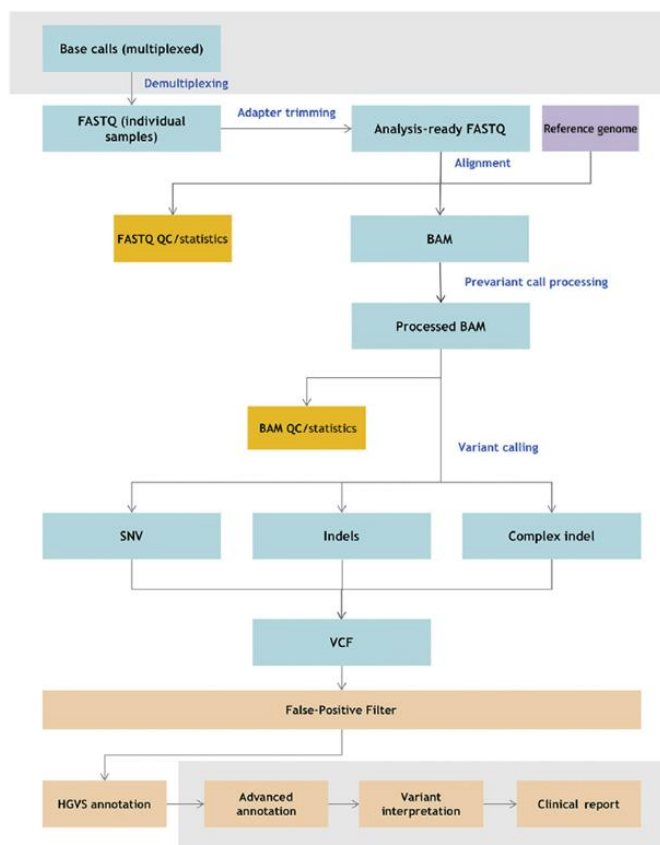

Figure 1. The figure illustrates the typical bioinformatics pipeline used for processing DNA-NGS data. Somak Roy et al J Mol Diagn 2018, 20: 4e27

- *Analysis performed/provided by our RNA-pipeline:*

  - *QC analysis (Sequencing quality check + contamination check)*
  - *Read filtering*
  - *Read mapping*
  - *Detection of gene and transcript abundance*
  - *Differential gene expression*
  - *Gene set enrichment analysis (GSEA) of gene signatures that can represent molecular pathways, oncogenic signatures, immunologic signatures, GO terms, transcription factors, etc).*
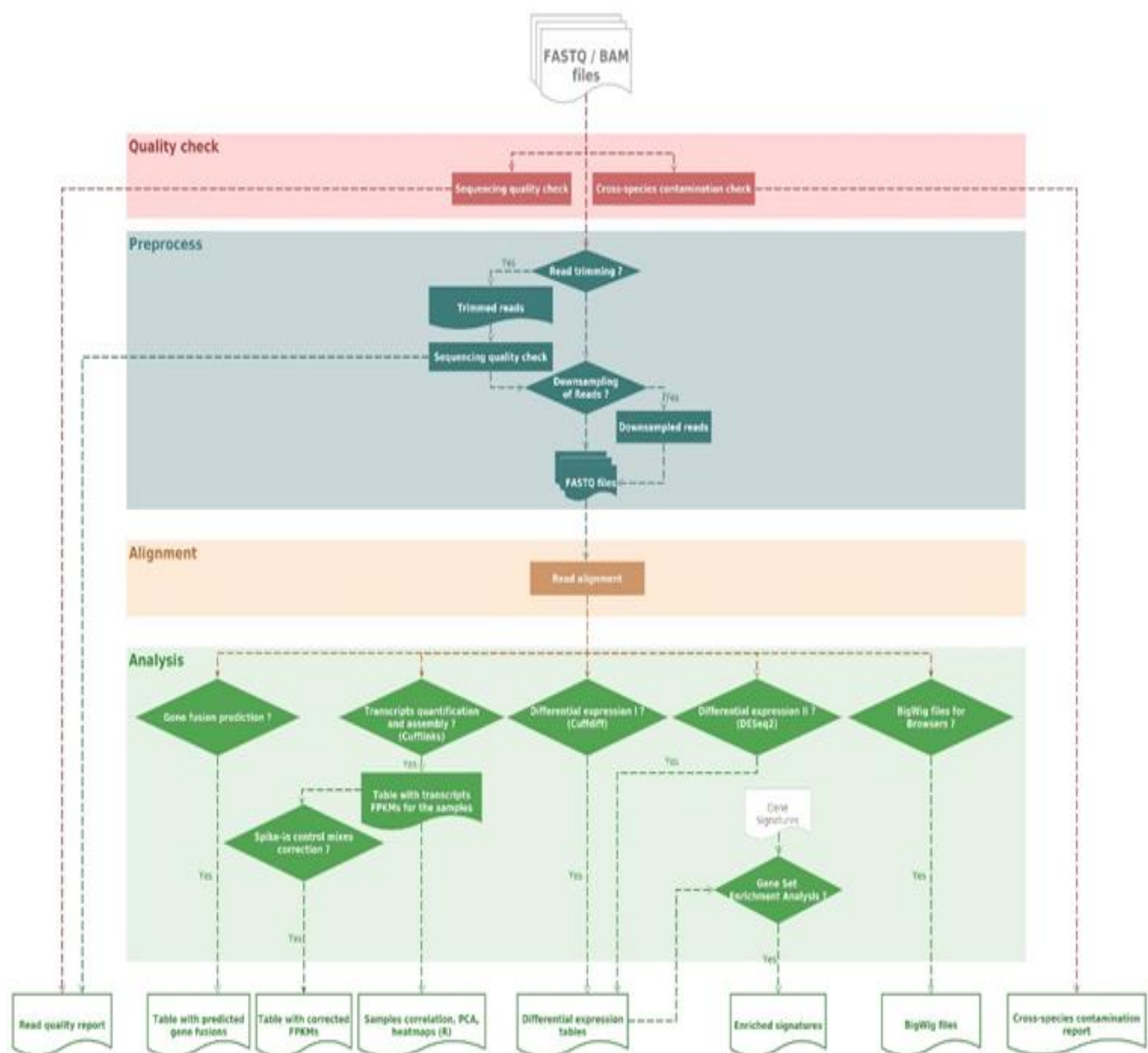

Figure 2. The figure illustrates the typical bioinformatics pipeline used for processing RNA-NGS data.

3.2.4.   Downstream analysis

Several bioinformatic databases that provide biological information generated from large scale biological experiments are publicly available. Depending on the case, the information is extracted from cohorts of patients with cancer, or from a battery of tumor cell lines. Information regarding the mutational landscape of genes and their expression patterns can be obtained from these databases. Furthermore, searching for available drugs able to target cancer genes and so prioritize cancer treatments is possible too. A list of the available services and the kind of biological information that we can provide is detailed below.

- *Evidence sources for precision oncology.*

Find drugs associated with specific genomic variants, evidence levels and cancer types for specific genes. http://depo-dinglab.ddns.net/

Gene	Alteration ▲	Variant Type	Cancer Type	Drug Therapy	Evidence Level	Effect	Sources
KRAS	-	CNA	COADREAD	BRAF inhibitor + MEK inhibitor/ anti-EGFR mAb ⓘ	Case Reports	Resistant	ENA 2014 (abstr 428)
KRAS	p.K117N	missense	COADREAD	cetuximab/panitumumab (anti-EGFR antibodies) ⓘ	Clinical Trials	Resistant	26952655 ↗
KRAS	12	missense	ALL	MEK inhibitors ⓘ	Preclinical	Sensitive	18701506 ↗
KRAS	12	missense	CESC	MEK inhibitors ⓘ	Preclinical	Sensitive	22169769 ↗
KRAS	12	missense	CHOL	MEK inhibitors ⓘ	Clinical Trials	Sensitive	23391555 ↗
KRAS	12	missense	COADREAD	ERK inhibitors ⓘ	Preclinical	Sensitive	23614898 ↗
KRAS	12	missense	COADREAD	panitumumab ⓘ	FDA Approved	Resistant	24024839 ↗
KRAS	12	missense	GIST	imatinib ⓘ	Case Reports	Resistant	24687822 ↗
KRAS	12	missense	LIHC	sorafenib + MEK inhibitor ⓘ	Clinical Trials	Sensitive	25294897 ↗
KRAS	12	missense	MM	PI3K pathway inhibitors + MEK inhibitors ⓘ	Preclinical	Sensitive	22985491 ↗
KRAS	12	missense	NSCLC	PI3K pathway inhibitors + MEK inhibitors ⓘ	Clinical Trials	Sensitive	25516890 ↗
KRAS	12	missense	OV	PI3K pathway inhibitors + MEK inhibitors ⓘ	Clinical Trials	Sensitive	25500057 ↗, 25516890 ↗
KRAS	12	missense	PAAD	gemcitabine + MEK inhibitors ⓘ	Clinical Trials	Sensitive	23583440 ↗
KRAS	12	missense	PAAD	PI3K pathway inhibitors + MEK inhibitor ⓘ	Clinical Trials	Resistant	ASCO 2015 (abstr 4119)
KRAS	12	missense	STAD	anti-EGFR mAbs ⓘ	Preclinical	Resistant	22614881 ↗, 22290393 ↗
KRAS	12	missense	UCEC	PI3K pathway inhibitors + MEK inhibitors ⓘ	Preclinical	Sensitive	21984976 ↗, 22662154 ↗

- *Prioritize anti-cancer drug treatments according to individual genomic data.*

By exploring the largest database of drug-target associations available, from well-known targeted therapies to preclinical drugs, we can prioritize multiple druggable alterations in genomically complex tumors.

- *Explore genetic dependencies identified in loss-of-function screens in panels of tumor cell lines*

http://www.cancergd.org/

Genes whose function is selectively essential in the presence of cancer-associated genetic aberrations, represent promising targets for the development of precision therapeutics. Integrating genotypic profiling with large-scale loss-of-function genetic screens in tumor cell lines, allows to identify such genetic dependencies. A genetic dependency is identified when there is a statistical association between the presence of a particular mutation and increased sensitivity to the inhibition of a specific gene.

Genetic dependency example with the KRAS gene:

Dependency ⇕	P-value ⇕	Effect size (%) ⇕	ΔScore ⇕	Study ⇕	Experiment Type ⇕	Multiple Hit ⇕	String Interaction ⇕	Inhibitors ⇕
Search	<0.05 ⇕	>= 65.0	< 0.0		⇕	⇕	⇕	⇕
KRAS	6×10^{-27}	95.9	-1.78	McDonald(2017)	shRNA	Yes	Highest	
KRAS	1×10^{-23}	91.5	-4.34	Tsherniak(2017)	shRNA	Yes	Highest	
KRAS	2×10^{-23}	96.1	-0.61	Meyers(2017)	CRISPR	Yes	Highest	
HERPUD2	5×10^{-10}	75.5	-0.61	Tsherniak(2017)	shRNA	Yes		
CDKL1	5×10^{-9}	74.0	-0.41	Tsherniak(2017)	shRNA	Yes		
CFLAR	2×10^{-8}	75.7	-0.35	Meyers(2017)	CRISPR			
ADSL	2×10^{-8}	73.9	-0.50	McDonald(2017)	shRNA			
ALDH1L1	2×10^{-8}	72.9	-0.64	Tsherniak(2017)	shRNA	Yes		
CD3D	3×10^{-8}	72.8	-1.10	Tsherniak(2017)	shRNA	Yes		
OR5K3	4×10^{-8}	74.9	-0.08	Meyers(2017)	CRISPR			
IFRD1	1×10^{-7}	81.9	-1.07	Tsherniak(2017)	shRNA	Yes		
RAPGEF3	1×10^{-7}	72.2	-0.38	McDonald(2017)	shRNA			
NT5C1B	2×10^{-7}	90.4	-1.58	Marcotte(2012)	shRNA			
E2F4	2×10^{-7}	71.2	-0.41	Tsherniak(2017)	shRNA	Yes		
TRPM7	2×10^{-7}	71.9	-0.55	McDonald(2017)	shRNA	Yes		
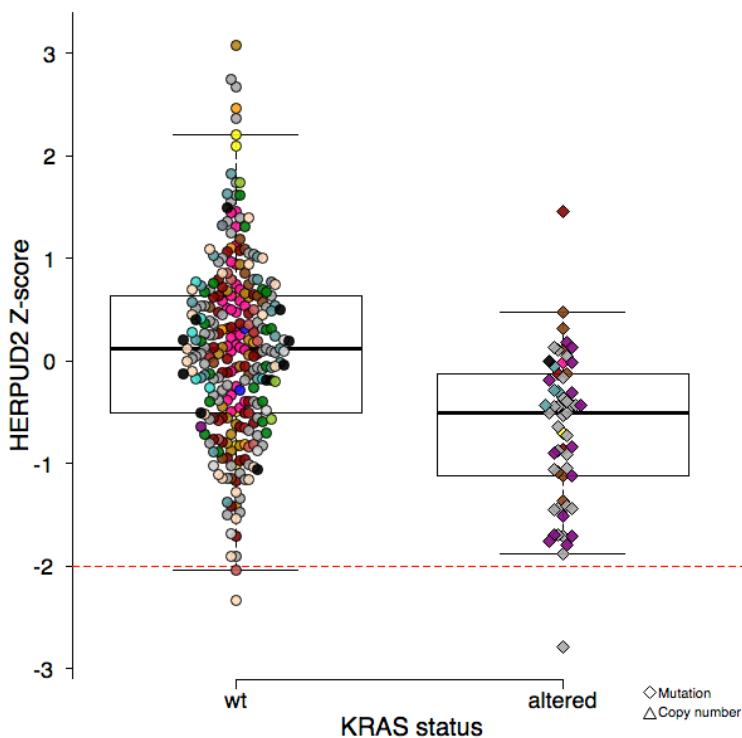
Show	Tissue	Cell lines		
		Wild type	Altered	Total
☑	Blood & Lymph	29	3	32
☑	Bone	6	1	7
☑	Breast	31	1	32
☑	Cervix	2	0	2
☑	CNS	19	0	19
☑	Endometrium	5	0	5
☑	Esophagus	19	0	19
☑	Head & Neck	5	0	5
☑	Kidney	9	0	9
☑	Large Intestine	11	7	18
☑	Liver	4	0	4
☑	Lung	64	27	91
☑	Ovary	19	4	23
☑	Pancreas	1	14	15
☑	Pleura	2	0	2
☑	Prostate	6	0	6
☑	Skin	22	0	22
☑	Soft tissue	6	0	6
☑	Stomach	11	1	12
☑	Urinary tract	3	0	3
All None	Totals:	274	58	332

Download boxplot as: [PNG image] [CSV file]

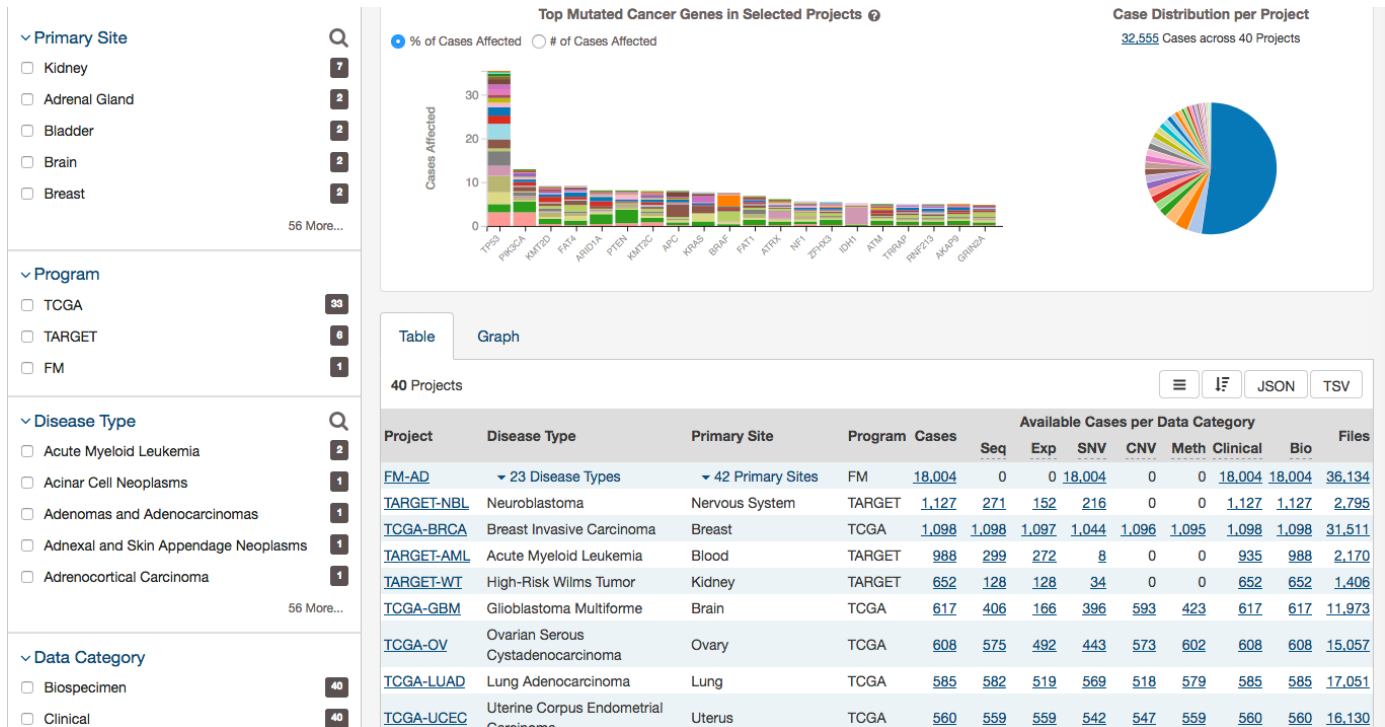Download legend for: [Selected tissues] [All tissues]

KRAS altered cell lines have an increased dependency upon **HERPUD2**

(p=5×10^{-10} | Effect size=75.5% | ΔScore=-0.61 | Tissues=Pan cancer | Source=Tsherniak(2017) | ExpType=shRNA)
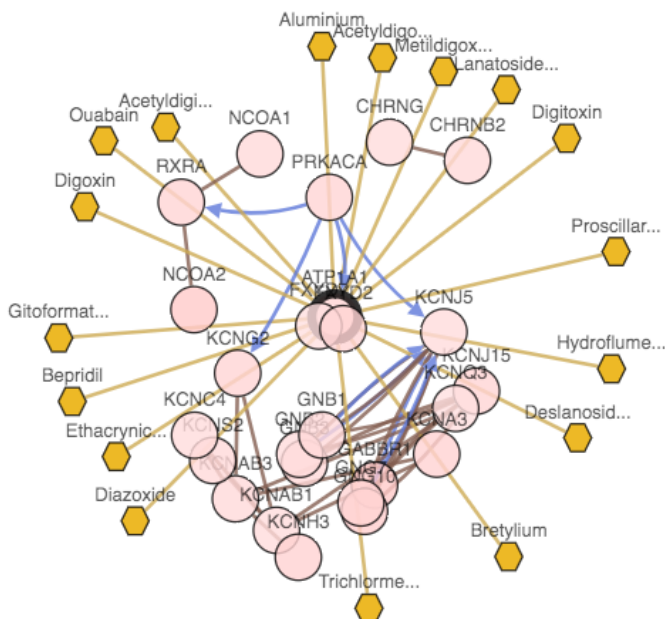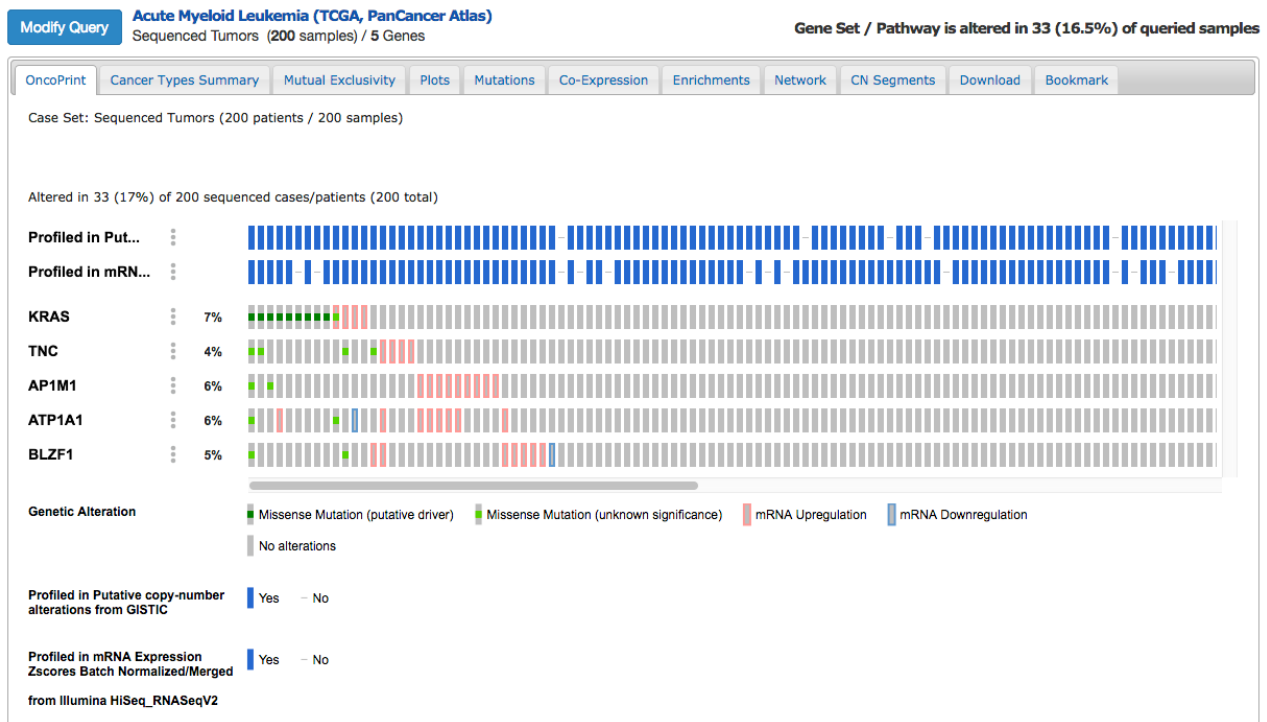
- *Explore data available from experiments performed on cohorts of patiens from TCGA (The Cancer Genome Atlas) across 33 types of cancer*

We can provide information about mutations, putative copy-number alterations, expression and co-expression for genes through the different cohorts of patients studied.

https://portal.gdc.cancer.gov/projects

http://www.cbioportal.org/index.do

- ***Database with biological information from more than 1000 tumor cell lines***

https://software.broadinstitute.org/software/cprg/?q=node/11

By mining a database that comprises more than 1000 tumor cell lines, we can provide relevant biological aspects of particular genes across the cell lines, that allows you to determine: (i) Gene expression based on RNA-seq experiments, (ii) Copy number variation based on Affymetrix assays, (iii) Gene essentiality based on genome-scale RNAi and CRISPR-Cas9 genetic perturbations, and (iv) Gene Promoter methylation based on RRBS-seq experiments.
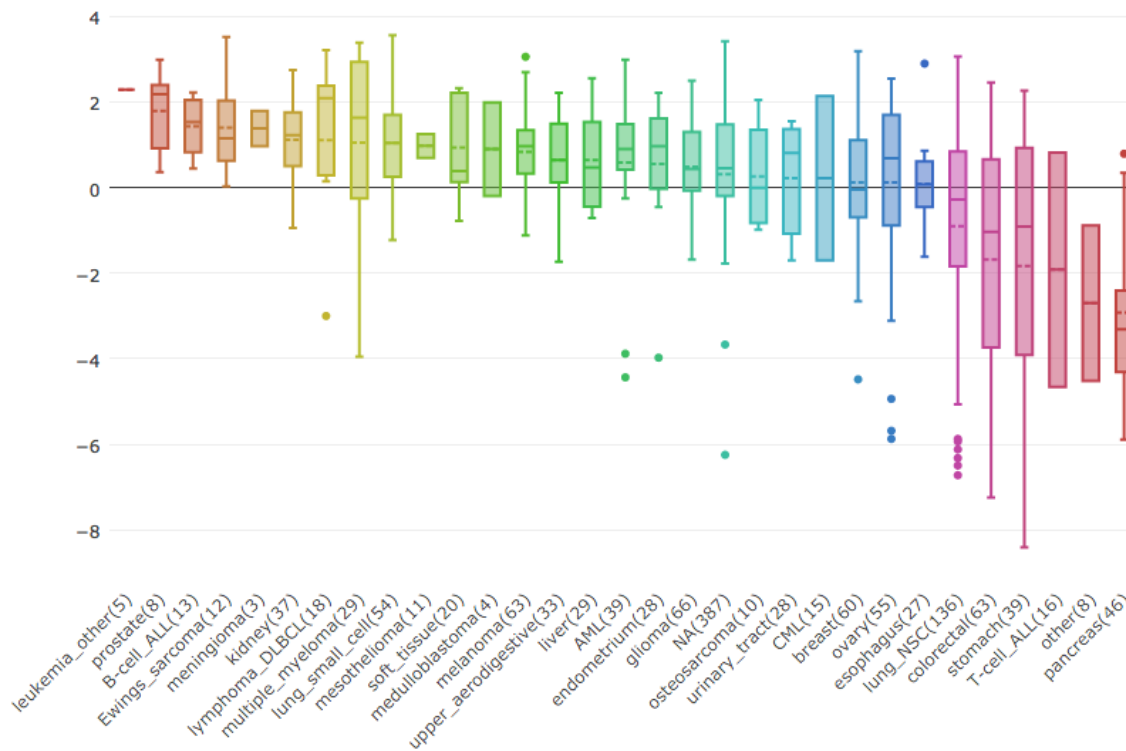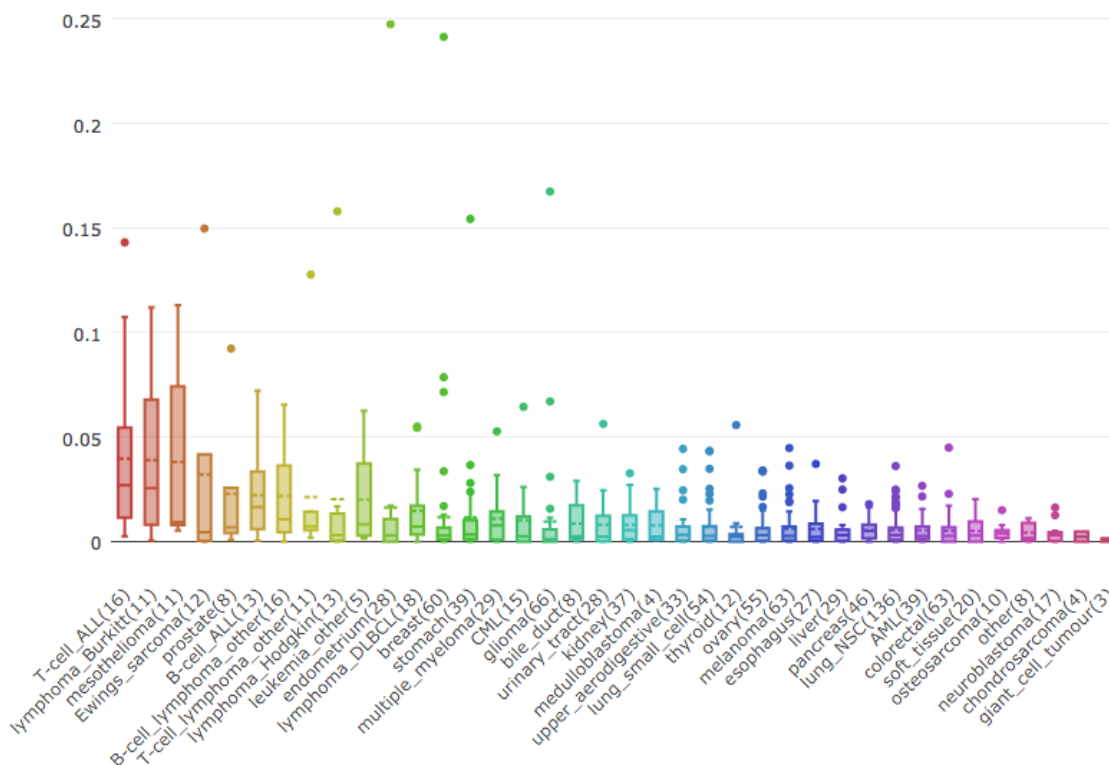
mRNA expression (RNAseq): KRAS
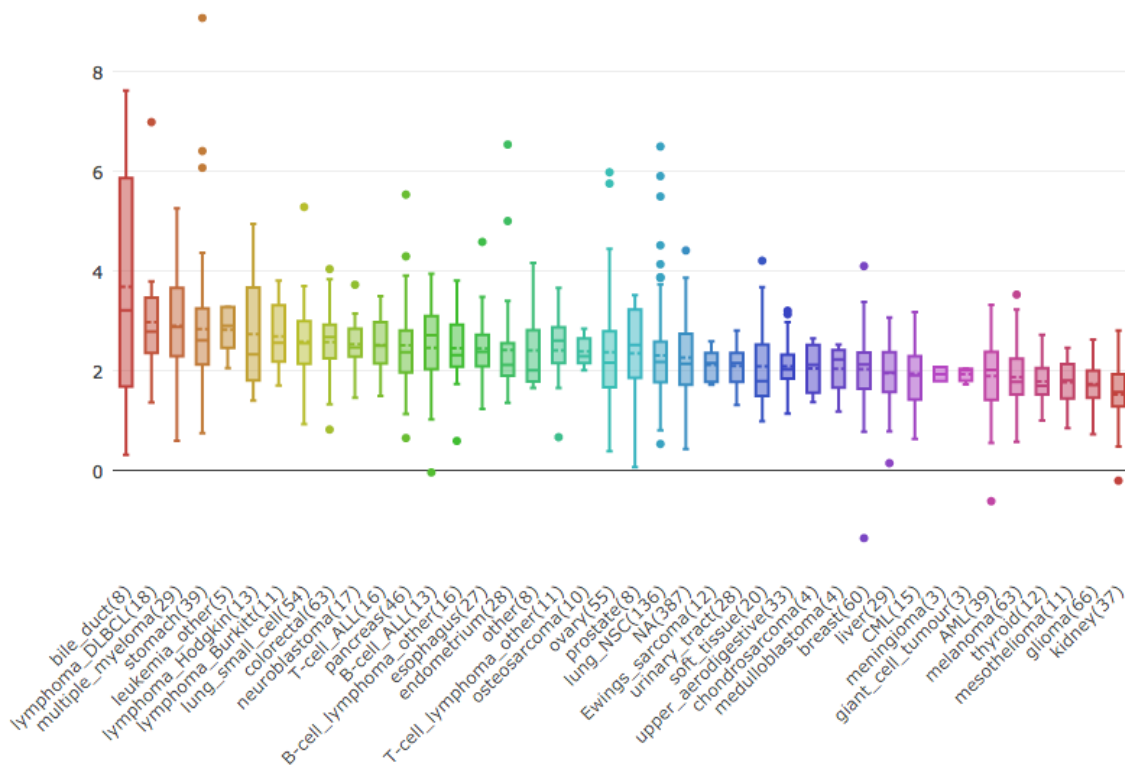
Copy Number: KRAS
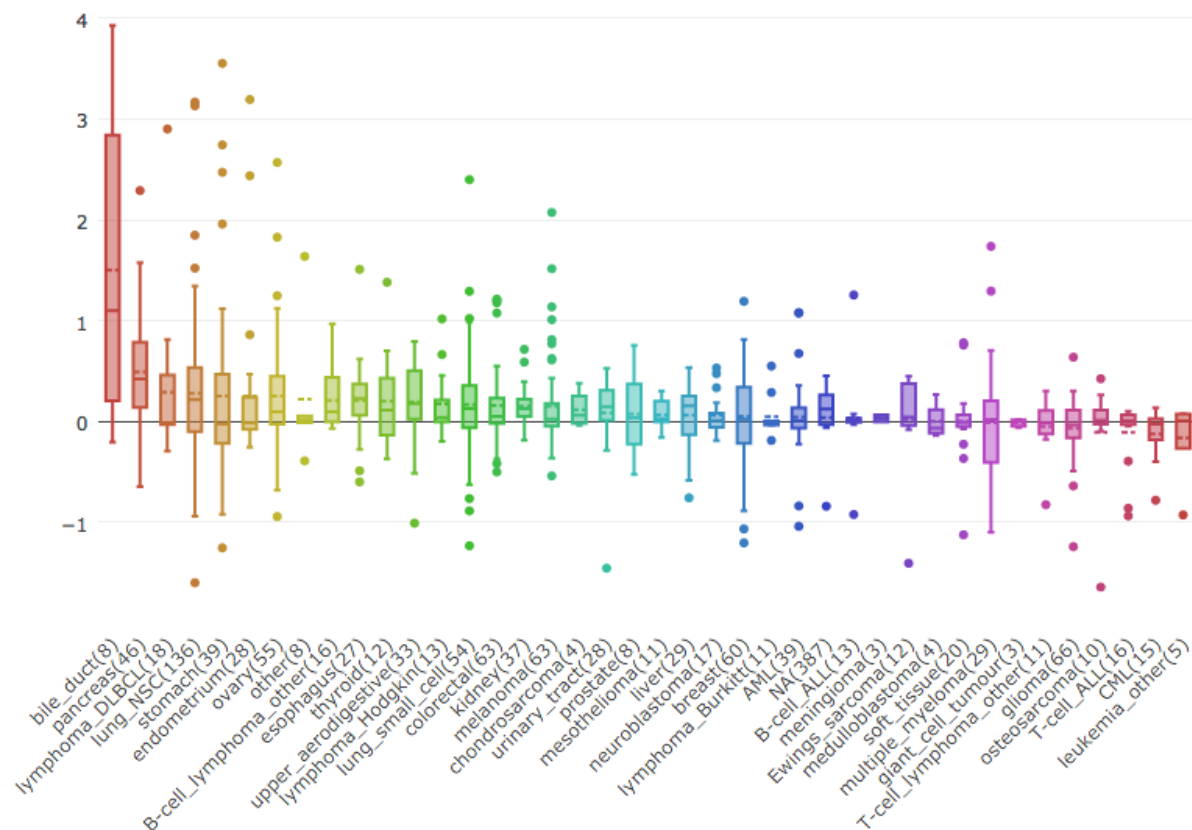
shRNA knockdown: KRAS

DNA methylation (RRBS): KRAS

- *Compare the results of your NGS experiment with similar experiments available from public repositories.*

Sometimes it can be interesting to verify the biological hypothesis driven by your own NGS experiments and compare it to data from similar experiments when they are available elsewhere. We can provide this ad-hoc comparisons by mining and analyzing available public datasets.

- *Data visualization and comparison*

Upload and view your NGS data in conjunction with an innumerable amount of data from public experiments. This can be easily done through a genome browser that concentrates a large number of biological data tracks with information obtained from experiments performed by research groups worldwide, or by computational analysis and predictions. Putting the information from your own experiment in the same genome browser, will allow you to detect additional biological details that could have impact on your experiment and drive your conclusions.
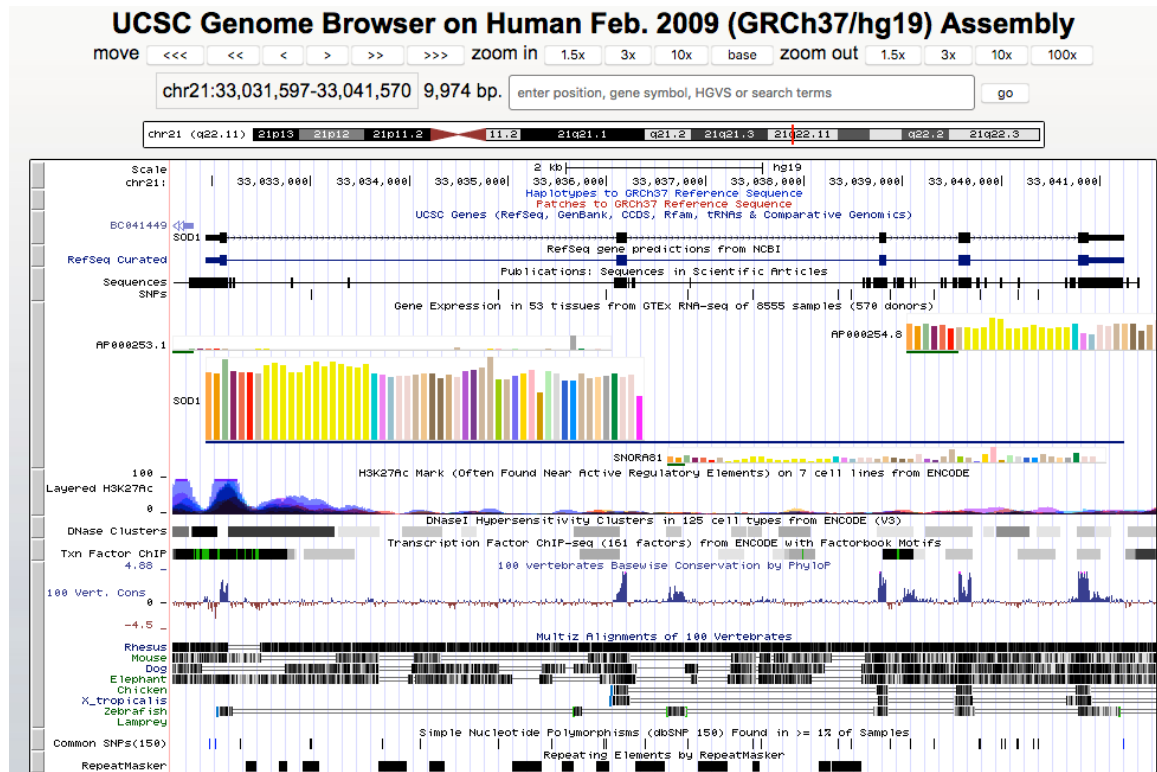
4. Expertise and team members

As pointed above, the initiative to extend our offer for data analysis services has been possible after gathering a cohesive and well-balanced team of professionals with expertise and knowledge to afford this new challenge. Below is briefly summarized the professional profile of each member involved in molecular biology and bioinformatic area. More detailed CV aspects of the two new members who will be more directly responsible of the molecular biology data analysis are described further on.

- **Joan Ballesteros**. PhD. Chairman & Chief Scientific Officer at Vivia Biotech. Expert in pharmacology and structural biology. In 1999 he founded a Biotech company in San Diego, California called Novasite Pharmaceuticals as Chief Scientific Officer. In 2006, he took the technology platform developed in Novasite to launch Vivia Biotech in Spain as Chairman and CSO.

- **Joaquin Martinez**. MD, PhD Medical Director at Vivia Biotech, Head of the Hematology Department at Hospital 12 Octubre and PI of Hematological Malignances Research group at Spanish National Oncology Research Center (CNIO). Professor of Medicine, Universidad Complutense de Madrid. Postdoctoral Research Fellowship within the Cellular Therapy Program at the Princess Margaret Hospital, University of Toronto, Canada. Dr Martínez López directs the optimization of new molecular diagnostic and prognostic tests for different molecular alterations observed in haematological malignancies looking for new biomarker. His most recent efforts are focus in minimal residual disease and in novel therapeutics for multiple myeloma and leukemia.

- **Julian Gorrochategui**. M.Sc. VP Bioinformatics and Information Technologies at Vivia Biotech. Expert in compound profiling and data analysis in pharmacology, drug discovery, precision medicine, in vitro diagnosis devices (IVD) and clinical studies. Microbiologist and Systems Analyst at Merck Research Laboratories (1991-2008); Director of Bioinformatics at Vivia Biotech (2008-18).

- **Jose Luis Rojas**. M Sc. Bioinformatic analyst at Vivia Biotech. Master in Bioinformatics. Expert in software development and information technologies. Experience in compound profiling and laboratory automation (Merck Reseach Laboratories 2005-2008 and Institute of Neurobiology Severo Ochoa 2009-2011) as well as Bioinformatic analysis. (CNIO 2011-12). Responsible at Vivia since 2012 for pharmacodynamic population modelling, R programing and systems integration.

- **Yanira Ruiz Heredia**. PhD Molecular and Computational Biology. Master in Bioinformatics. 2018 Bioinformatics analyst at Vivia Biotech (See details below)

- **Osvaldo Graña**. PhD, Data analysis coordinator of the Bioinformatic Unit at the CNIO Spanish National Cancer Research Center. (See details below).

4.1. Most relevant CV aspects from new team members

- *Yanira Ruiz Heredia*

Educational experience

BSc in Chemistry from the Universidad Complutense of Madrid, with an Erasmus fellowship at University of Greenwich, London, UK. (2013). In 2015 awarded to an industrial PhD focused on the development and implementation of precision medicine strategies for multiple myeloma treatment based on molecular biology techniques (NGS) and pharmacodynamic drug profiling. The PhD was led by Dr. Joaquín Martinez at the Hospital Universitario 12 de Octubre and Spanish National Cancer Research Center (CNIO) and Joan Ballesteros at Vivia Biotech. In 2016 Master in Bioinformatic and Computational Biology in CNIO and headed by professor Alfonso Valencia. In 2017 awarded to an international fellowship in the Biostatistic and Compulatonal Biology group led by Dr. Nikhil Munshi and Giovanni Parmigiani at Dana- Farber-Harvard Medical School in Boston, Massachuttets.

Expertise

- Experimental designs and set up of NGS assays for the 2 main sequencing system, Illumina and Ion Torrent, including DNA/RNA extraction, quantification, library preparation and sequencing.
- Development of an in-house experimental and analytical method for minimal residual disease assessment in multiple myeloma by NGS.
- Implementing of a novel ultra-deep targeted sequencing strategy to explore the mutational landscape of a very homogenous cohort of MM newly diagnosed patients and identification of new prognostic factors and predictive biomarkers of the response to treatment. NGS data processing, analysis, visualization and interpretation, as well as extensive statistical analysis and clinical data management.
- Study the role of mitochondrial DNA alterations in the metabolism and pathogenesis of multiple myeloma from Whole Genome Sequencing and Whole Exome Sequencing data.
- Participation in the development of a novel method for minimal residual disease evaluation in acute myeloid leukemia patients by NGS.

- Determining the optimal experimental and analytical conditions to describe the pharmacodynamic behaviour of the most commonly used drugs in the treatment of multiple myeloma from fresh bone marrow aspirates by multiparametric flow cytometry.
- Constructing the pharmacodynamics profiles of anti-myeloma drugs by applying population model, as a first step toward the creation of a personalized medicine test which enable the prediction of the clinical response to treatment from ex vivo experiments.

Publications

- Abnormalities in mitochondrial DNA copy number has pathogenetic and prognostic implications in Multiple Myeloma. Ruiz-Heredia Y, Samur KM, Munshi N, Martinez-Lopez J et al. Ongoing.
- Novel deep targeted sequencing method for minimal residual disease monitoring in acute myeloid leukemia. Onecha E, Linares M, Rapado I, Ruiz-Heredia Y et al. Accepted as original article in Haematologica journal.
- Mutational screening of newly diagnosed multiple myeloma patients by deep targeted sequencing. Ruiz-Heredia Y, Sànchez-Vega B, Onecha E, Barrio S et al. Haematologica. 2018 Jun 28. pii: haematol.2018.188839.
- Concurrent progressive multifocal leukoencephalopathy and central nervous system infiltration by multiple myeloma: A case report. Ruiz-Heredia Y, Sanchez-Vega B, Barrio S, Linares M et al. J Oncol Pharm Pract. 2018 Jan 1:1078155218769367.
- Analytical and clinical validation of a novel in-house deep-sequencing method for minimal residual disease monitoring in a phase II trial for multiple myeloma. J Martinez-Lopez , B Sanchez-Vega, S Barrio, I Cuenca, Y Ruiz-Heredia et al.
  Leukemia volume31, pages1446–1449 (2017).
- Differentiation stage of myeloma plasma cells: biological and clinical significance. Paiva B, Puig N, Cedena MT, de Jong BG, Ruiz Y et al. Leukemia. 2017 Feb;31(2):382-392.
- How useful is molecular modelling in combination with ion mobility mass spectrometry for 'small molecule' ion mobility collision cross-sections? Cris Lapthorn, Frank S. Pullen, Babur Z. Chowdhry, Patricia Wright,George L. Perkins and Yanira Heredia. Analyst, Issue 20, 2015.
- Characterization of nanoparticulated phases in the manganese oxo/hydroxide system obtained in supercritical water: Optimized conditions for selected compositions. Ruiz Y et al. The Journal of Supercritical Fluids Mar 7, 2013.

- ***Osvaldo Graña***

Professional experience

- Spanish National Cancer Research Center (CNIO) Madrid, Since 2006

In charge of the coordination of the bioinformatics service provided to the CNIO research groups, and also to some research groups outside the CNIO. Next Generation Sequencing (NGS) data analysis in its different branches (RNA-seq, Small-RNA-seq, ChIP-seq, BS-seq, RIP-seq, DNA-seq, Methyl-CAP). Developed different bioinformatic tools and specific pipelines for the analysis of RNA-seq (nextpresso) and BS-seq (bicycle) data, that are publicly available. In-house training in NGS pipeline execution and NGS data analysis. NGS experimental design assistance, optimal data visualization, integration and distribution, manuscript preparation.

- Center of Molecular Biology (CBM) Madrid 2000-2005

Under the supervision of professor Alfonso Valencia, developed computational algorithms based on neural networks to predict the three-dimensional structure of proteins. Wrote different programs to evaluate the quality of 3D protein models predicted by research groups in the field worldwide, in the context of EVA and the CASP and CAFASP international competitions.

Teaching activities

Teaching activities in several bioinformatic aspects: RNA-seq, ChIP-seq and Genome Browsing in the Master in Bioinformatics at the Universidad Autónoma de Madrid. RNA-seq in the Master's degree in Bioinformatics applied to Personalized Medicine and Health (Escuela Nacional de Sanidad - Instituto de Salud Carlos III). Coordinates the subject 'Advanced Bioinformatics' of the Universidad Francisco de Vitoria, and teaches De-Novo Genome Assembly, ChIPseq, RNA-seq and Genome Browsing. Previously was a teacher in the Master of Bioinformatics at the Universidad Complutense de Madrid, in the Master of Bioinformatics at the Instituto de Salud Carlos III, in the Master and Doctorate Program TADSI at the Universidad de Vigo, and in the Master in Molecular Biology, Cellular Biology and Molecular Biomedicine at the Universidad Autónoma de Madrid.

Publications

PubMed link:

https://www.ncbi.nlm.nih.gov/myncbi/browse/collection/50212415/?sort=date&direction=descending