CRDC Success Story:

Bridging the gap between data scientists and biologists and clinicians

Dr. Daoud Meerzaman
Computational Genomics and Bioinformatics Branch (CGBB)
NCI Center for Biomedical Informatics and IT

October 17, 2024

NIH NATIONAL CANCER INSTITUTE

Computational Genomics & Bioinformatics Branch (GCBB)

Computational Genomics & Bioinformatics Branch

- Provide state-of-art data analysis on a variety of multifaceted cancer-related research projects including proteogenomic, AI, and clinical high-throughput data.

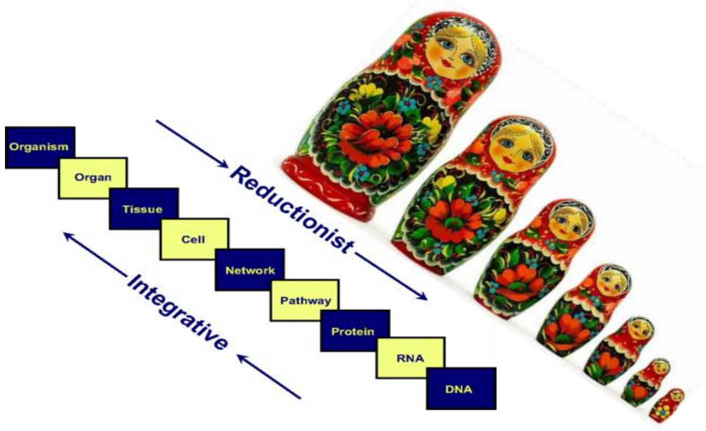
Computational Genomics & AI

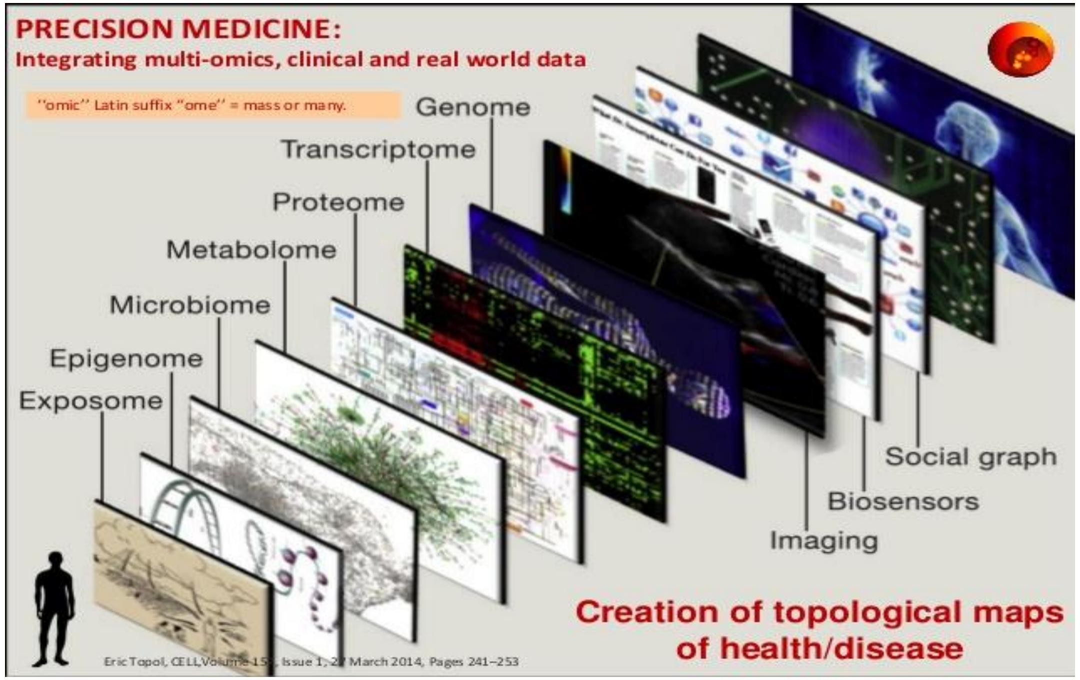
- NGS and Integrated analysis of proteogenomic data for Intramural NCI and moonshot projects (APOLLO, CIMAC-CIDC and SeQC2).
- Multimodal AI and ML approaches for patient outcome prediction. Deep learning algorithms, CNNs.

Scientific & Tools Consulting and training

- Generate and implement a web-based workflow in CRDC (CGC) and OmicCircos and (MOGSA)
- Widely used tools for NGS analyses and Visualization tools
- ITCR Scientific software
- CBIIT Summer Internship Program

From Bench to Bedside: Integrative Data Analysis and Precision Medicine





PRECISION MEDICINE:
Integrating multi-omics, clinical and real world data

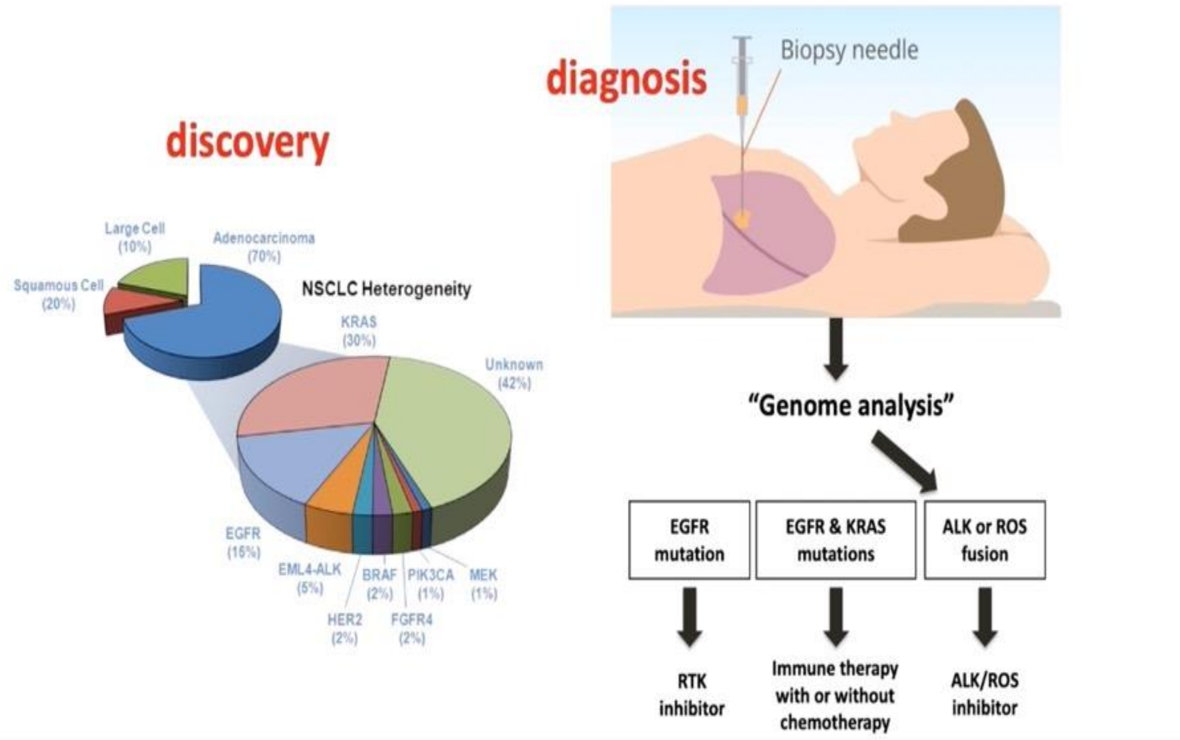
"omic" Latin suffix "ome" = mass or many.

Creation of topological maps of health/disease

Eric Topol, CELL Volume 15, Issue 1, 27 March 2014, Pages 241–253

https://www.slideshare.net/TWilckens/symbiotic-innovation-082015

Adult cancers: genomics provides a treatment playbook



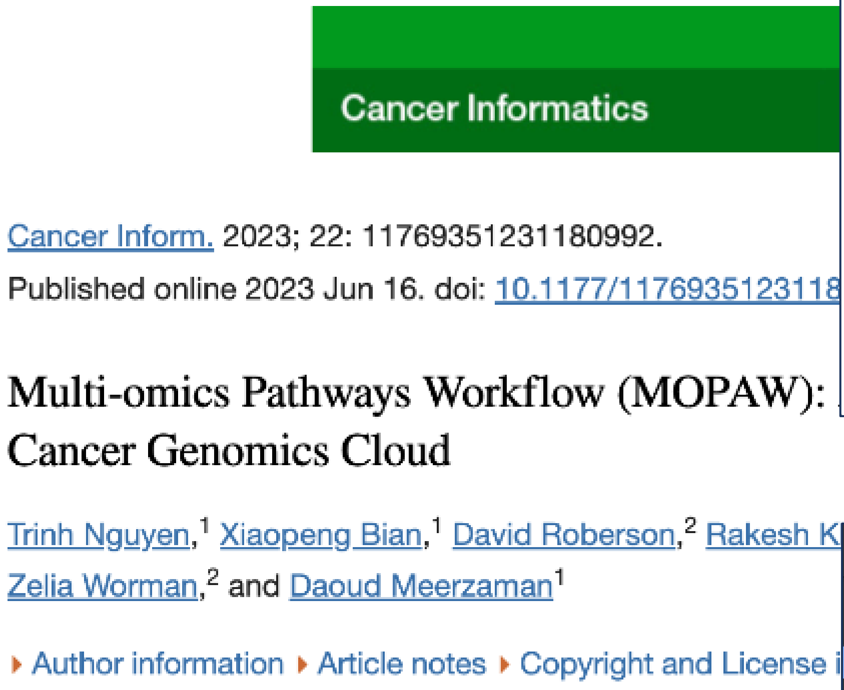
https://www.slideshare.net/JTADrexel/bioinformatics-2512758?next_slideshow=2

bridging the gap between data scientists and biologists and clinicians

From Complex Code to Simple Click: Empowering Users with Accessible Tools using CRDC

Velsera SB-CGC

- Access to many different data sets no need to download data
- Offers one on one demo and training to use the system.
- more than 600 analytical and bioinformatics tools and workflows readily available
- It requires minimal or no coding experience to use these tools and workflows.

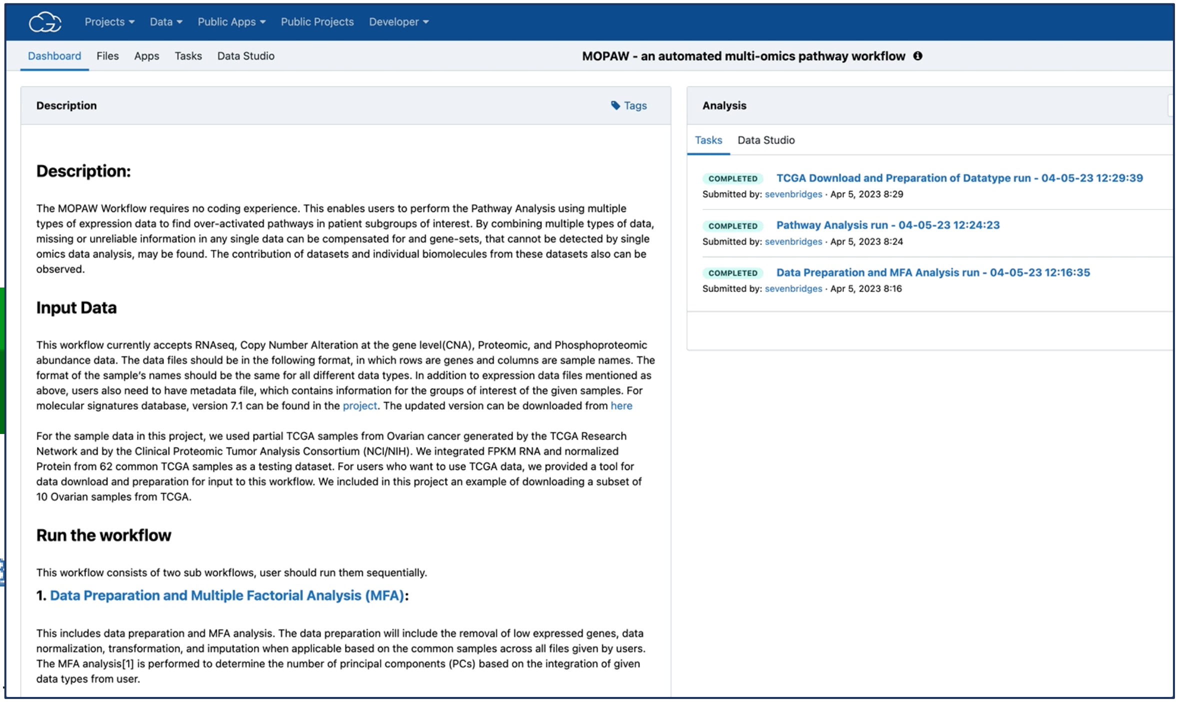


[Cancer Inform.](#) 2023; 22: 11769351231180992.

Published online 2023 Jun 16. doi: 10.1177/1176935123118

Multi-omics Pathways Workflow (MOPAW):
Cancer Genomics Cloud

Trinh Nguyen,[1] Xiaopeng Bian,[1] David Roberson,[2] Rakesh K
Zelia Worman,[2] and Daoud Meerzaman[1]



MOPAW - an automated multi-omics pathway workflow

Description:

The MOPAW Workflow requires no coding experience. This enables users to perform the Pathway Analysis using multiple types of expression data to find over-activated pathways in patient subgroups of interest. By combining multiple types of data, missing or unreliable information in any single data can be compensated for and gene-sets, that cannot be detected by single omics data analysis, may be found. The contribution of datasets and individual biomolecules from these datasets also can be observed.

Input Data

This workflow currently accepts RNAseq, Copy Number Alteration at the gene level(CNA), Proteomic, and Phosphoproteomic abundance data. The data files should be in the following format, in which rows are genes and columns are sample names. The format of the sample's names should be the same for all different data types. In addition to expression data files mentioned as above, users also need to have metadata file, which contains information for the groups of interest of the given samples. For molecular signatures database, version 7.1 can be found in the project. The updated version can be downloaded from here

For the sample data in this project, we used partial TCGA samples from Ovarian cancer generated by the TCGA Research Network and by the Clinical Proteomic Tumor Analysis Consortium (NCI/NIH). We integrated FPKM RNA and normalized Protein from 62 common TCGA samples as a testing dataset. For users who want to use TCGA data, we provided a tool for data download and preparation for input to this workflow. We included in this project an example of downloading a subset of 10 Ovarian samples from TCGA.

Run the workflow

This workflow consists of two sub workflows, user should run them sequentially.

1. Data Preparation and Multiple Factorial Analysis (MFA):

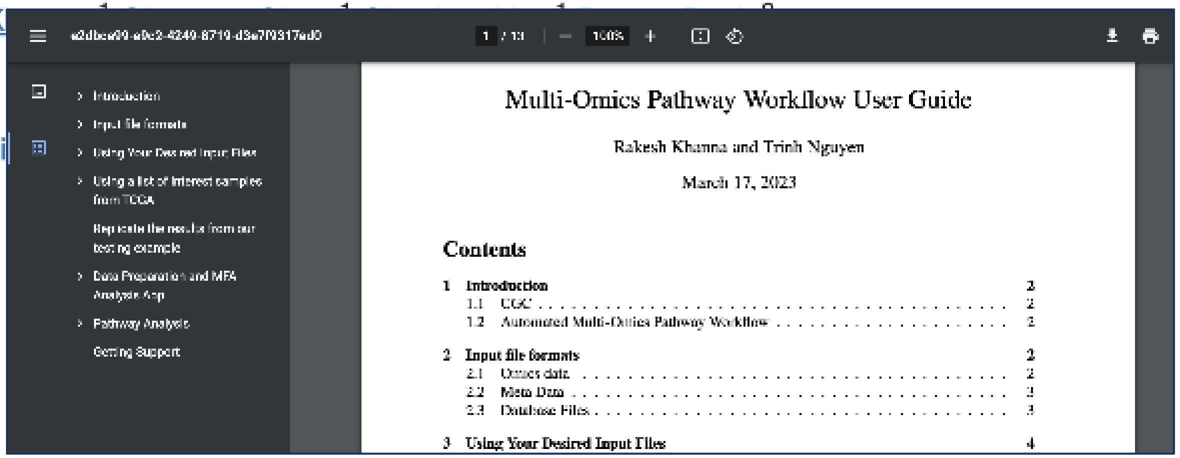
This includes data preparation and MFA analysis. The data preparation will include the removal of low expressed genes, data normalization, transformation, and imputation when applicable based on the common samples across all files given by users. The MFA analysis[1] is performed to determine the number of principal components (PCs) based on the integration of given data types from user.

COMPLETED **TCGA Download and Preparation of Datatype run - 04-05-23 12:29:39**
Submitted by: sevenbridges · Apr 5, 2023 8:29

COMPLETED **Pathway Analysis run - 04-05-23 12:24:23**
Submitted by: sevenbridges · Apr 5, 2023 8:24

COMPLETED **Data Preparation and MFA Analysis run - 04-05-23 12:16:35**
Submitted by: sevenbridges · Apr 5, 2023 8:16

https://cgc.sbgenomics.com/u/sevenbridges/mopaw-1



Multi-Omics Pathway Workflow User Guide

Rakesh Khanna and Trinh Nguyen

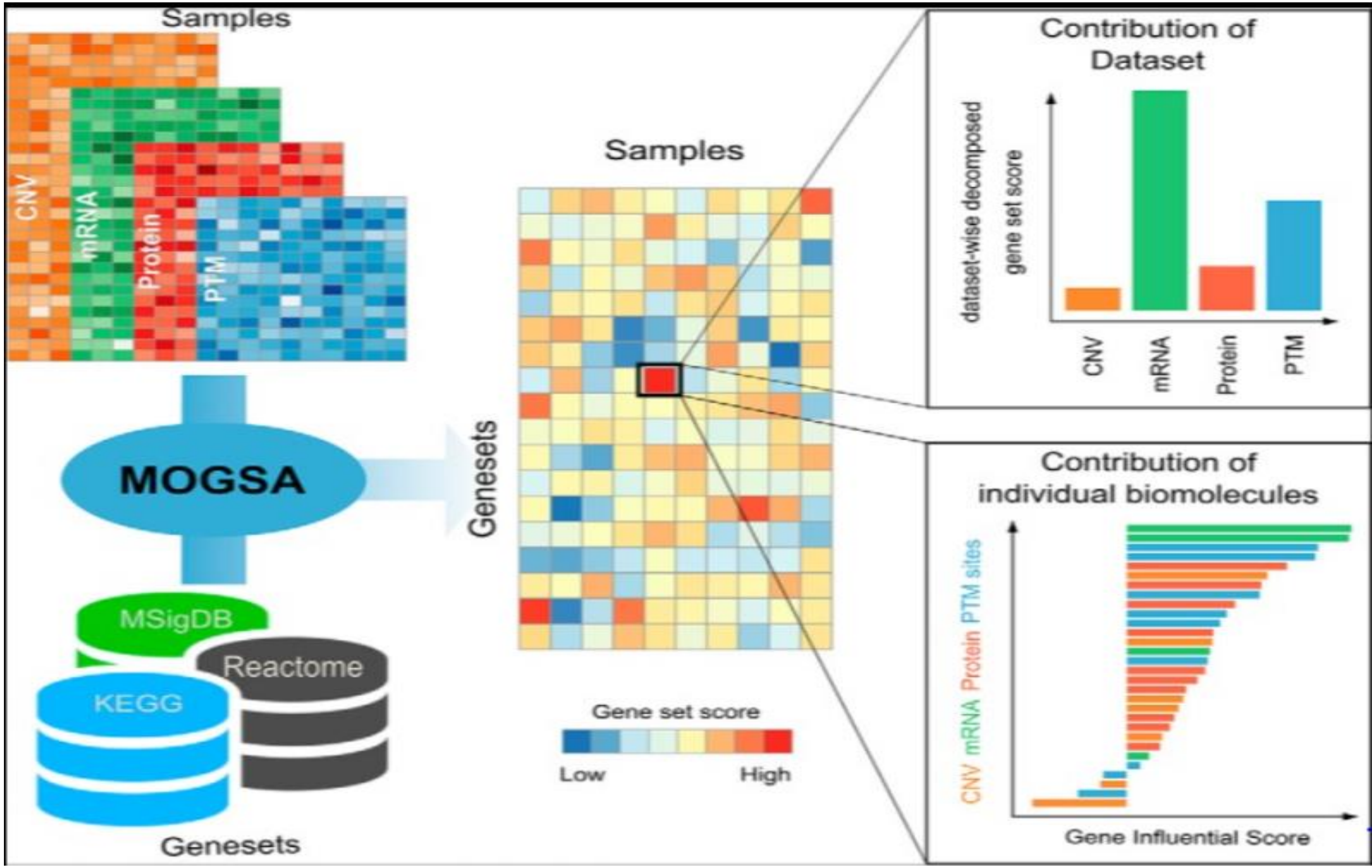
March 17, 2023

Contents

1 Introduction 2
1.1 CGC 2
1.2 Automated Multi-Omics Pathway Workflow 2
2 Input file formats 2
2.1 Omics data 2
2.2 Meta Data 3
2.3 Database Files 3
3 Using Your Desired Input Files 4

Multi-omics gene-set analysis (MOGSA)

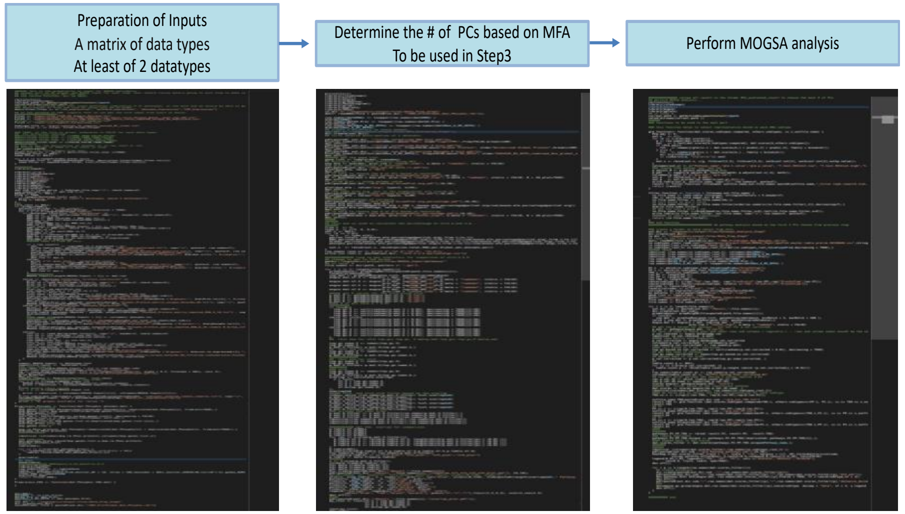
A multivariate single sample gene-set analysis method that integrates multiple experimental and molecular data types measured over the same set of samples.



https://pubmed.ncbi.nlm.nih.gov/31243065/

From Complex Code to Simple Click: Empowering Users with Accessible Tools using CRDC

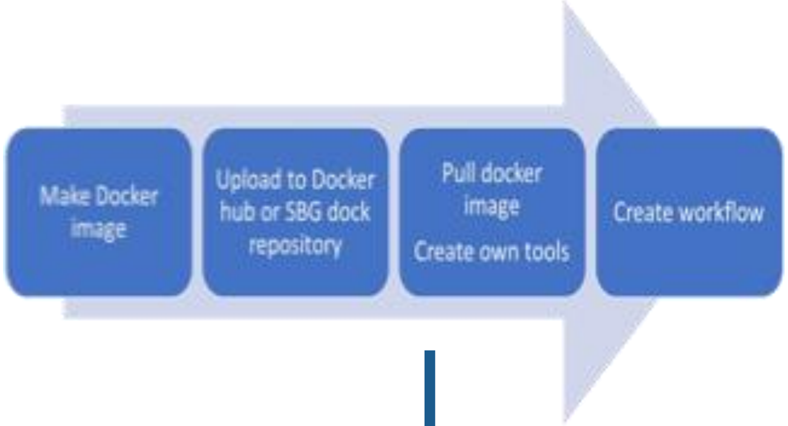
Perform MOGSA Analysis in 3 Steps



Preparation of Inputs
A matrix of data types
At least of 2 datatypes

Determine the # of PCs based on MFA
To be used in Step3

Perform MOGSA analysis

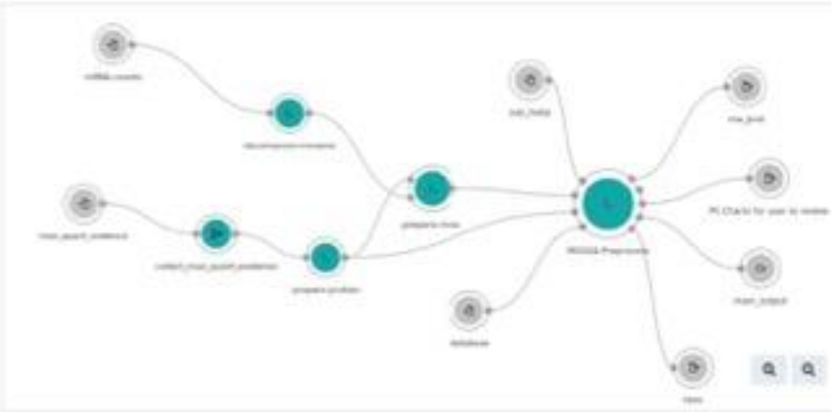


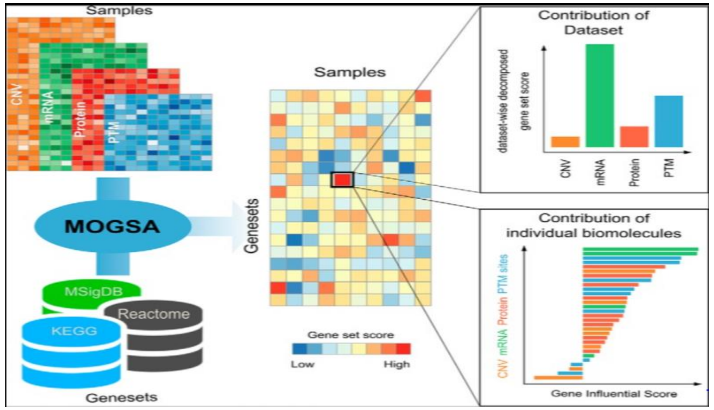
Make Docker image

Upload to Docker hub or SBG dock repository

Pull docker image
Create own tools

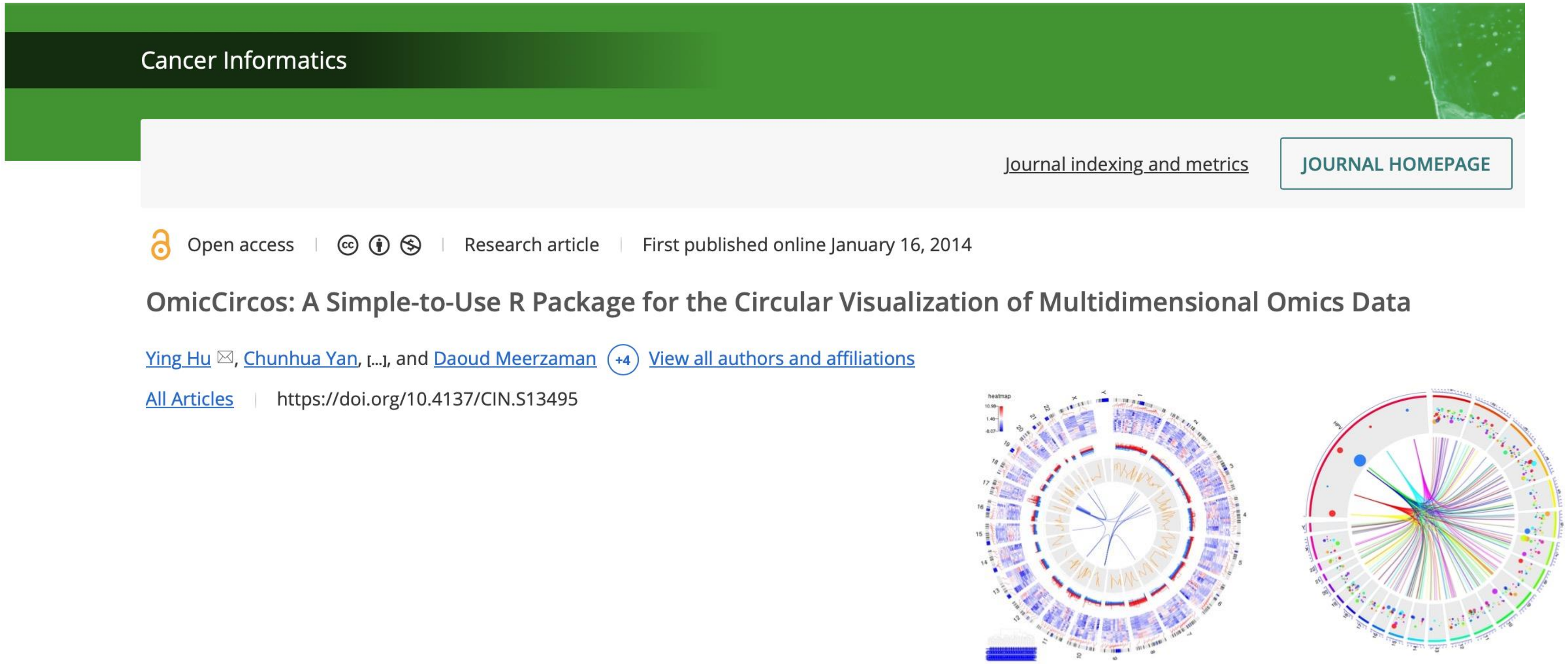
Create workflow





OmicCircos

- *Since December 2013*
- *Sited over by 160 articles*
- Total views and downloads of this article: 4323
- Total downloads of the tool more than 43000



Cancer Informatics

OmicCircos: A Simple-to-Use R Package for the Circular Visualization of Multidimensional Omics Data

Ying Hu, Chunhua Yan, [...], and Daoud Meerzaman +4 View all authors and affiliations

All Articles | https://doi.org/10.4137/CIN.S13495

Implementation of OmicCircos on CRDC-CGC

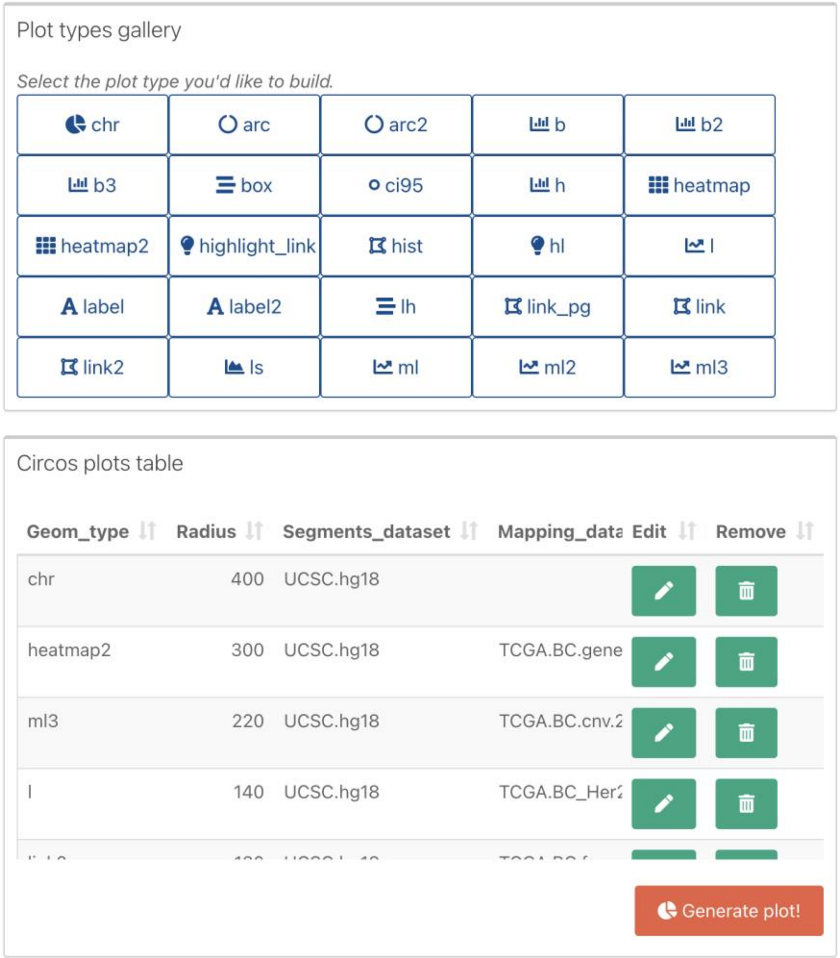
- Bioconductor R package into an R shiny application named OmicCircos App. This App requires no coding experience.

OmicCircos App

OmicCircos App is R Shiny application created around OmicCircos R package for more effective generation of high-quality circular plots for visualizing

Tool Integration and Data Usability

Graphical User Interface- OmicCircos App

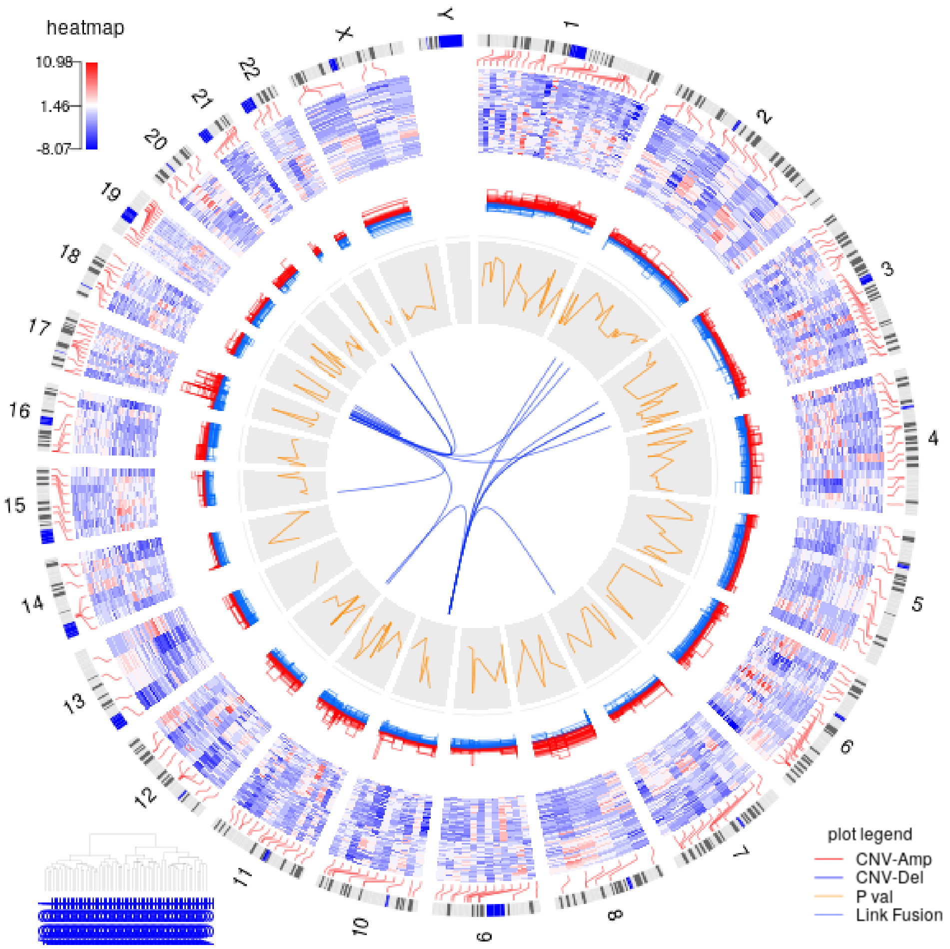


Plot types gallery

Select the plot type you'd like to build.

Circos plots table

Geom_type	Radius	Segments_dataset	Mapping_data	Edit	Remove
chr	400	UCSC.hg18		✎	🗑
heatmap2	300	UCSC.hg18	TCGA.BC.gene	✎	🗑
ml3	220	UCSC.hg18	TCGA.BC.cnv.	✎	🗑
l	140	UCSC.hg18	TCGA.BC_Her	✎	🗑

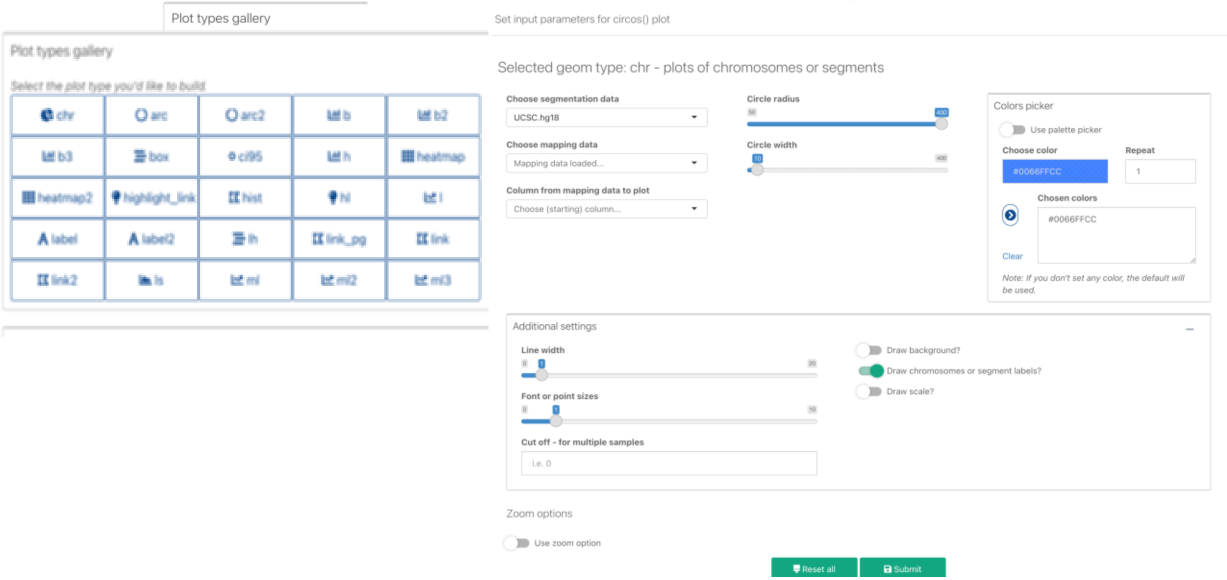


Genomic variations for Basal subtype of BCR

Generate plot!

Track 1, Chromosome

Step 1: Plot of Chromosome or segment



Set input parameters for circos() plot

Selected geom type: chr - plots of chromosomes or segments

Choose segmentation data: UCSC.hg18

Choose mapping data: Mapping data loaded...

Column from mapping data to plot: Choose (starting) column...

Circle radius: 400 (50–400)

Circle width: 10 (10–400)

Colors picker

○ Use palette picker

Choose color: #0066FFCC

Repeat: 1

Chosen colors: #0066FFCC

Clear

Note: If you don't set any color, the default will be used.

Additional settings

Line width: 1 (0–20)

Font or point sizes: 1 (0–10)

Cut off - for multiple samples: i.e. 0

○ Draw background?

● Draw chromosomes or segment labels?

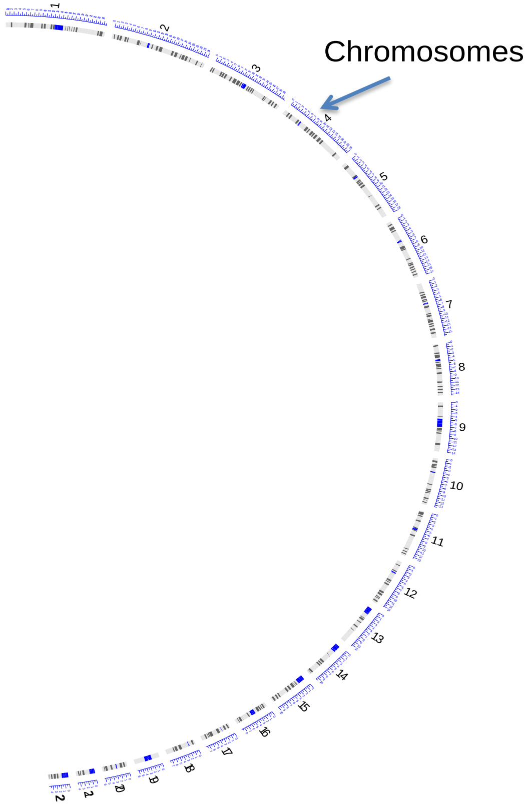
○ Draw scale?

Zoom options

○ Use zoom option

Reset all

Submit



Gene Expression for Basal subtype of BCR

Generate plot!

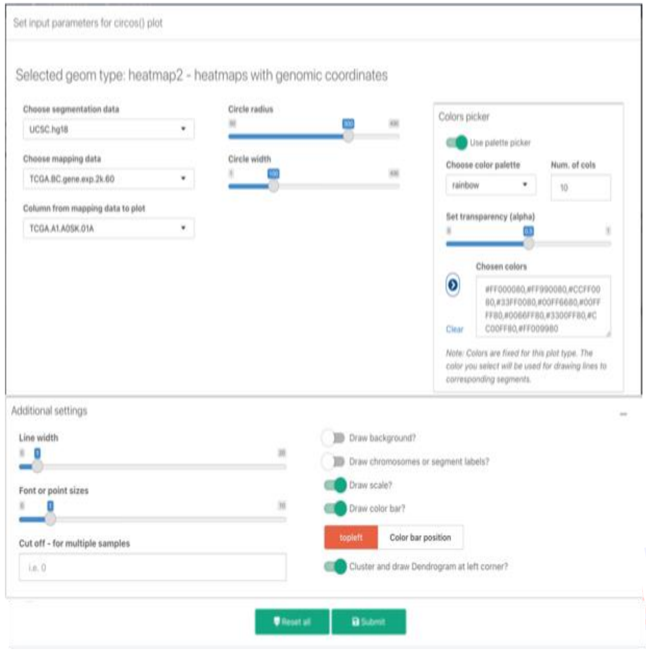
Track 2, Gene expression

Plot types gallery

Select the plot type you'd like to build.

Plot types gallery

Select the plot type you'd like to build.



Set input parameters for circos() plot

Selected geom type: heatmap2 - heatmaps with genomic coordinates

Choose segmentation data: UCSC.hg18

Choose mapping data: TCGA.BC.gene.exp.2k.60

Column from mapping data to plot: TCGA.A1.A0SK.01A

Circle radius

Circle width

Colors picker

Use palette picker

Choose color palette: rainbow

Num. of cols: 10

Set transparency (alpha)

Chosen colors: #FF000080,#FF990080,#CCFF0080,#33FF0080,#00FF6680,#00FFFF80,#0066FF80,#3300FF80,#CC00FF80,#FF009980

Note: Colors are fixed for this plot type. The color you select will be used for drawing lines to corresponding segments.

Additional settings

Line width

Font or point sizes

Cut off - for multiple samples: i.e. 0

Draw background?

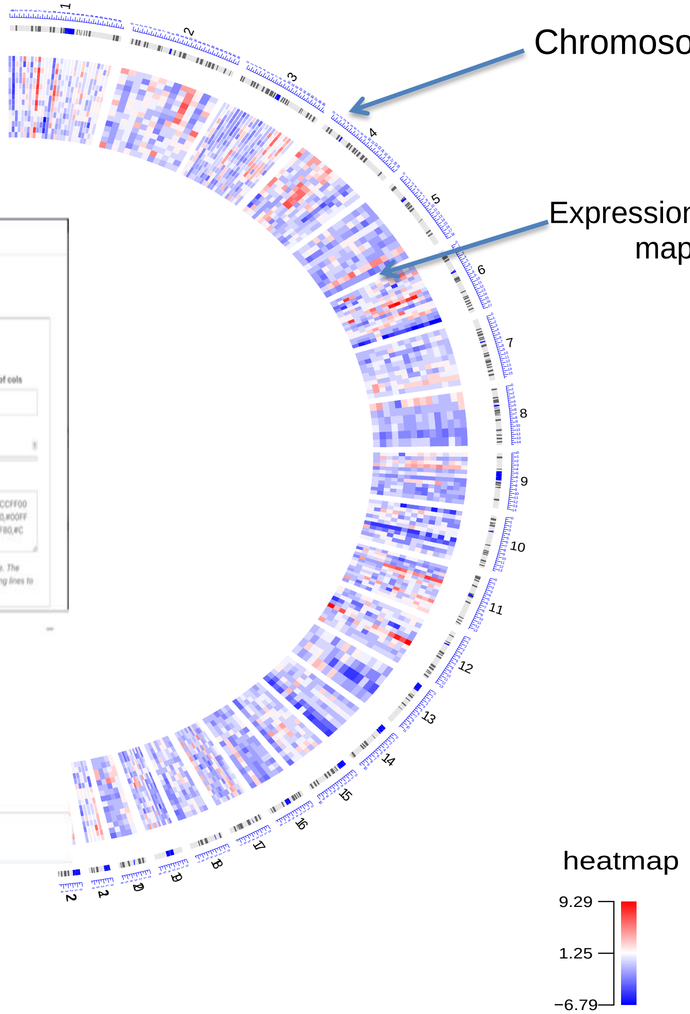
Draw chromosomes or segment labels?

Draw scale?

Draw color bar?

Color bar position: topleft

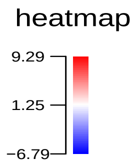
Cluster and draw Dendrogram at left corner?

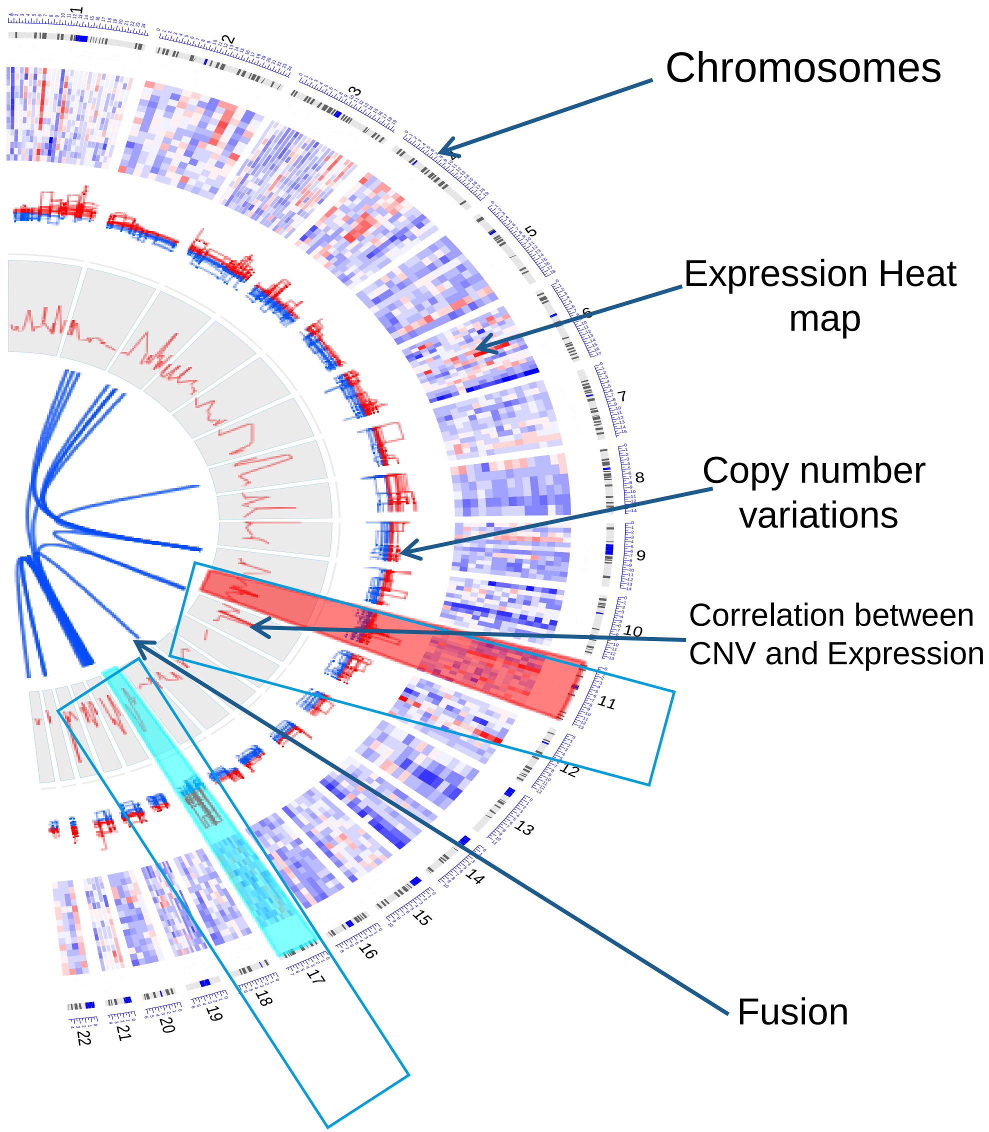


Chromosomes

Expression Heat map

Zooming on chromosome 11 and 17 for Basal subtype of BCR





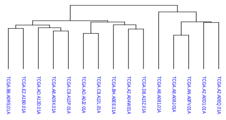
Chromosomes

Expression Heat map

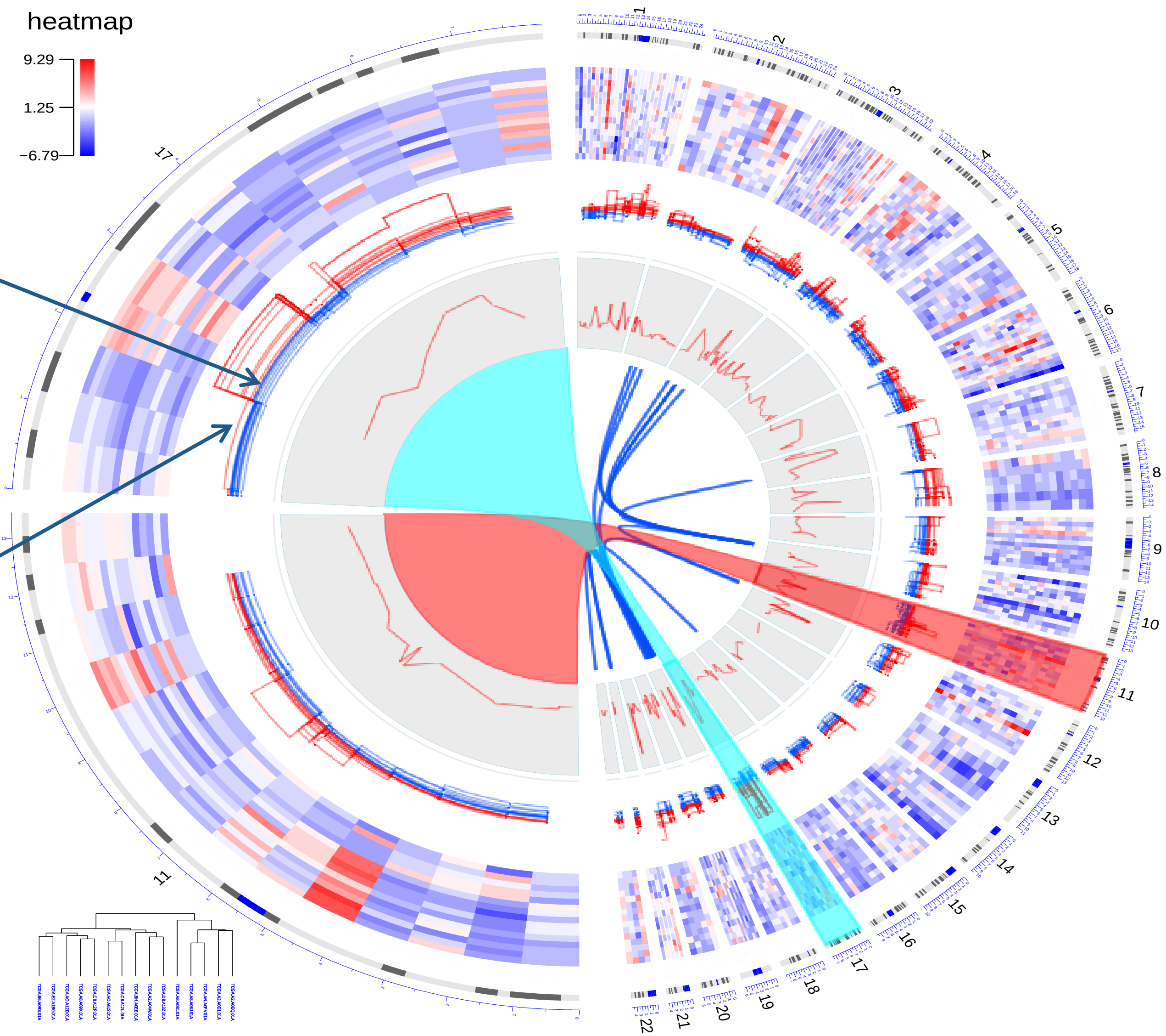
Copy number variations

Correlation between CNV and Expression

Fusion



Zooming on chromosome 11 and 17

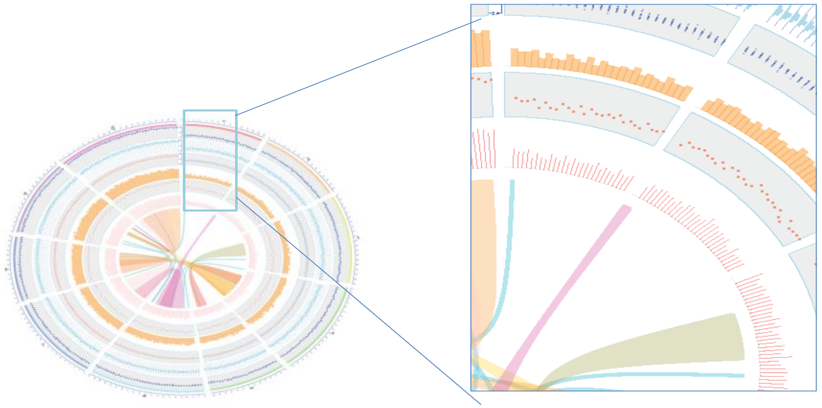


ERBB2 (Her2)

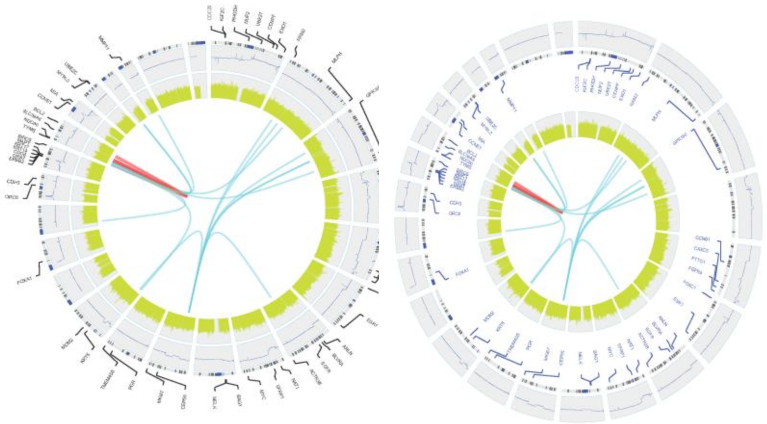
P53

OmicCircos capabilities

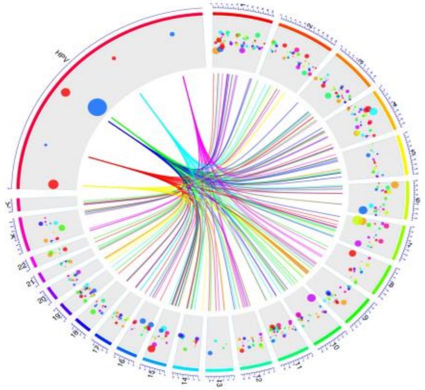
Display the Results in Boxplot, Histogram and Scatterplot



OmicCircos: text annotation

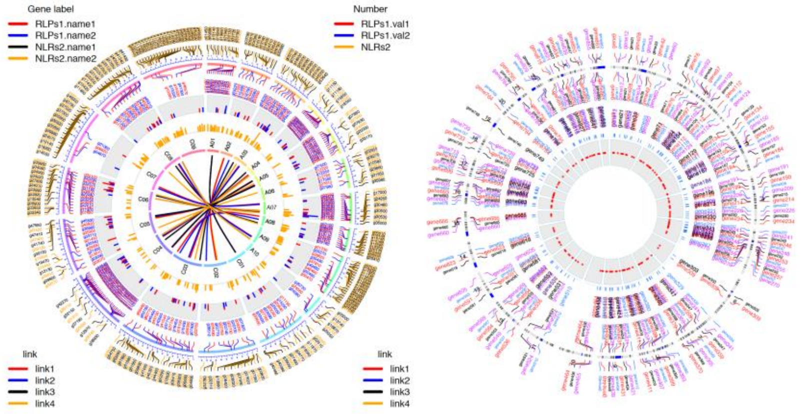


Interaction of genomes of HPV and human RNA seq association HPV and human



Dot means variation,
increase variation bigger dots

Complex genome annotation



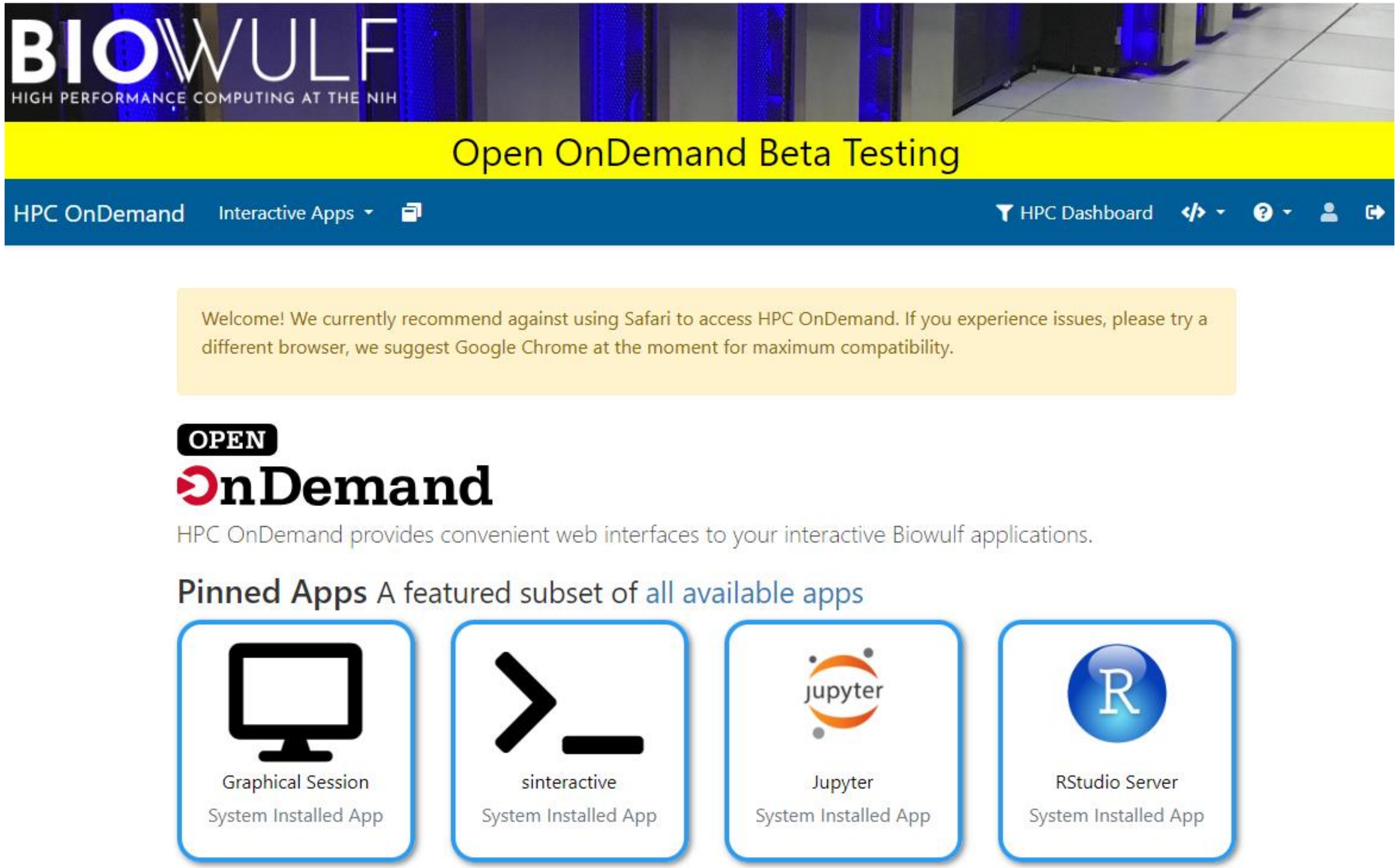
OmicCircos Demo

OmicCircos

Powerful, Informative, Publishable
Circos Plots, Made Easy

- https://cgc.sbgenomics.com/webapps

Biowulf on Demand



Open OnDemand Beta Testing

Welcome! We currently recommend against using Safari to access HPC OnDemand. If you experience issues, please try a different browser, we suggest Google Chrome at the moment for maximum compatibility.

HPC OnDemand provides convenient web interfaces to your interactive Biowulf applications.

Pinned Apps A featured subset of all available apps

3DVizSNP

BMC Part of Springer Nature

Search

BMC Bioinformatics

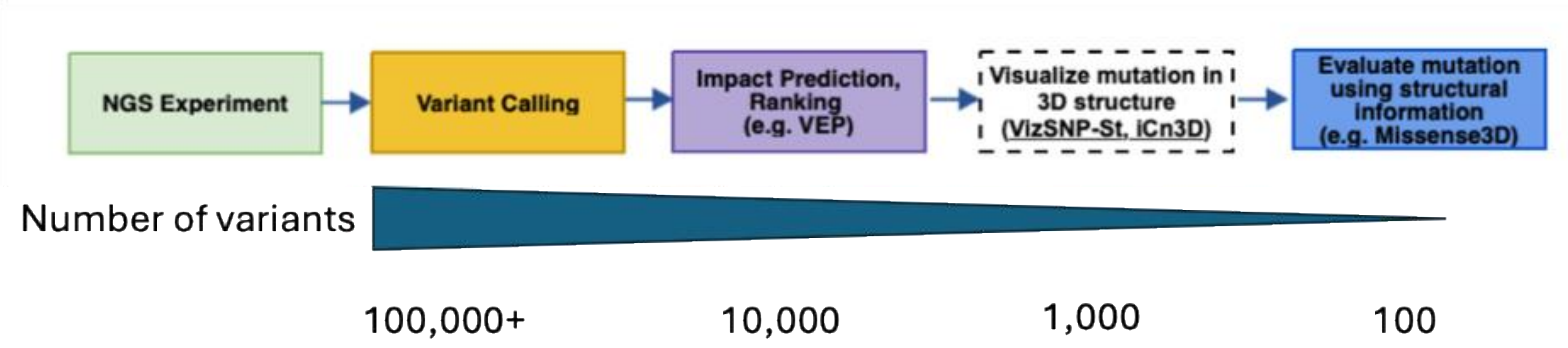
Software | Open access | Published: 09 June 2023

3DVizSNP: a tool for rapidly visualizing missense mutations identified in high throughput experiments in iCn3D

Michael Sierk, Shashikala Ratnayake, Manoj M. Wagle, Ben Chen, Brian Park, Jiyao Wang, Philippe Youkharibache & Daoud Meerzaman

BMC Bioinformatics **24**, Article number: 244 (2023) | Cite this article

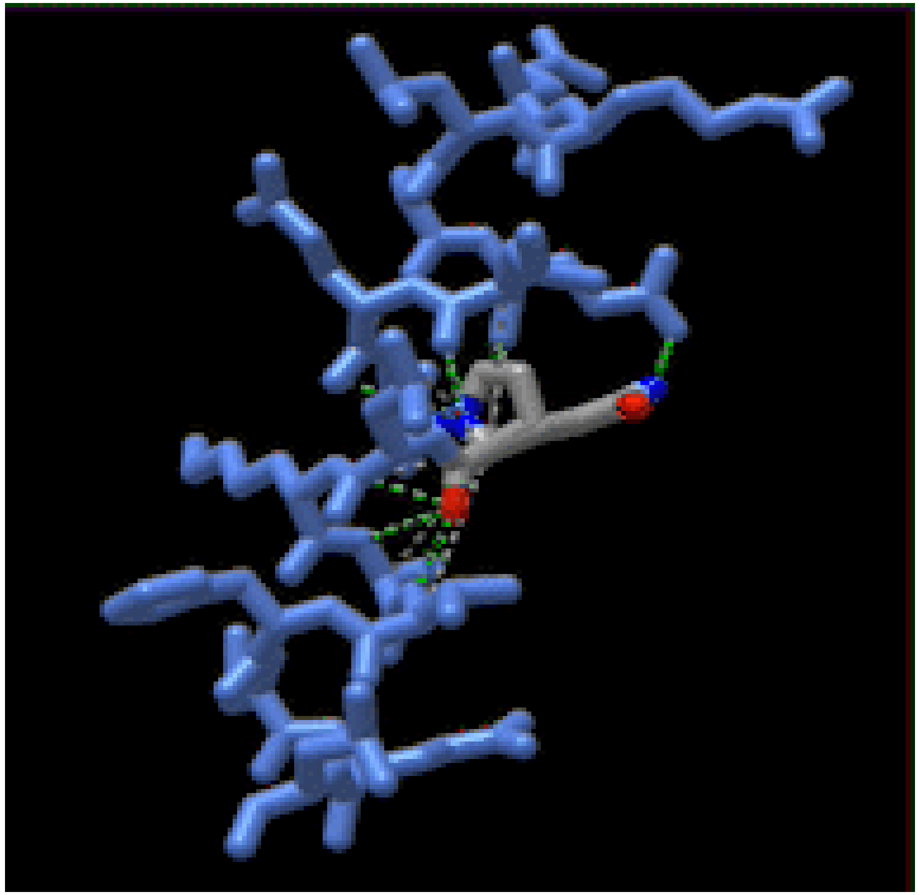
What role does 3DVizSNP play?

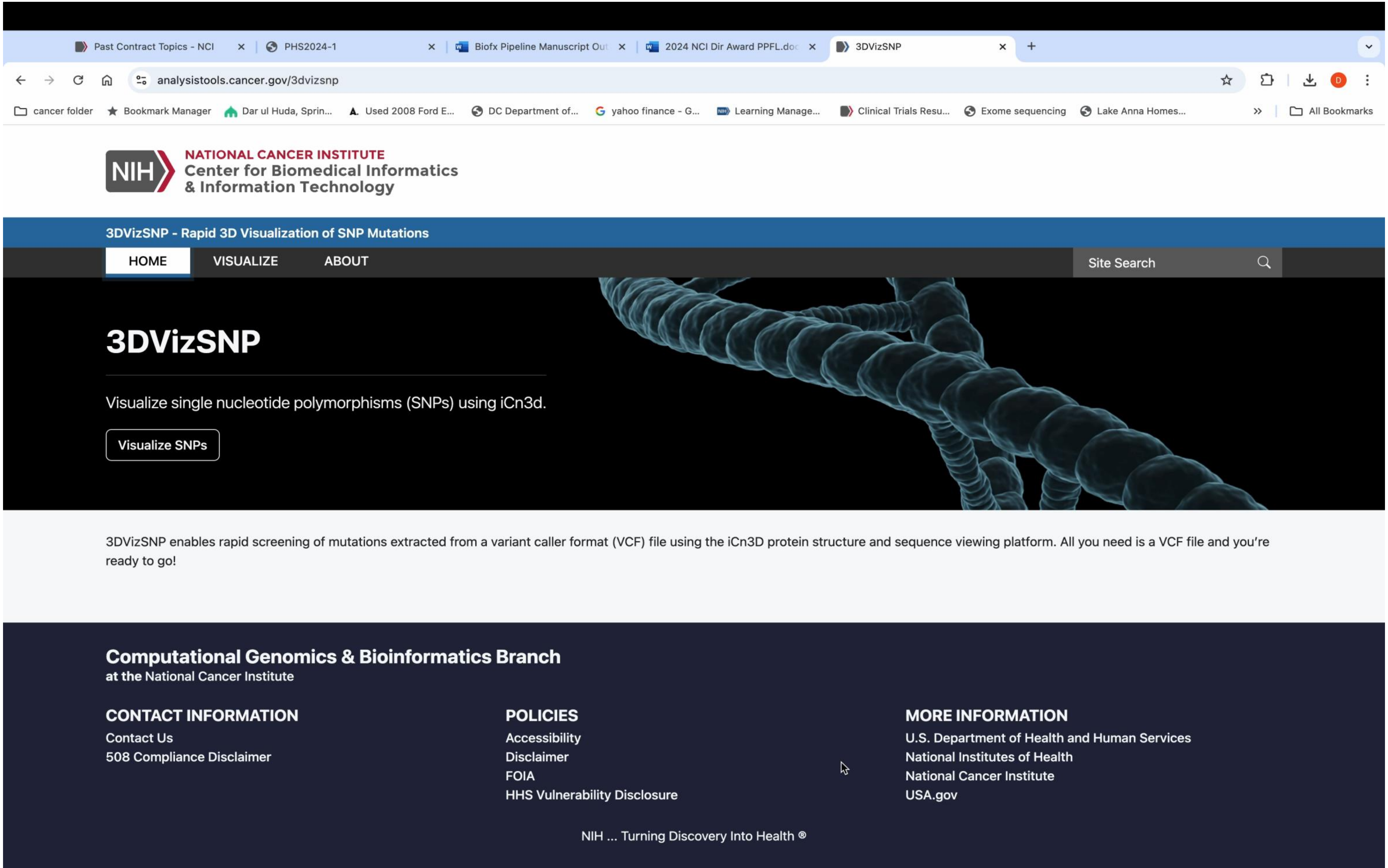


Number of variants

100,000+ 10,000 1,000 100

Results





3DVizSNP - Rapid 3D Visualization of SNP Mutations

3DVizSNP

Visualize single nucleotide polymorphisms (SNPs) using iCn3d.

3DVizSNP enables rapid screening of mutations extracted from a variant caller format (VCF) file using the iCn3D protein structure and sequence viewing platform. All you need is a VCF file and you're ready to go!

Computational Genomics & Bioinformatics Branch
at the National Cancer Institute

CONTACT INFORMATION
Contact Us
508 Compliance Disclaimer

POLICIES
Accessibility
Disclaimer
FOIA
HHS Vulnerability Disclosure

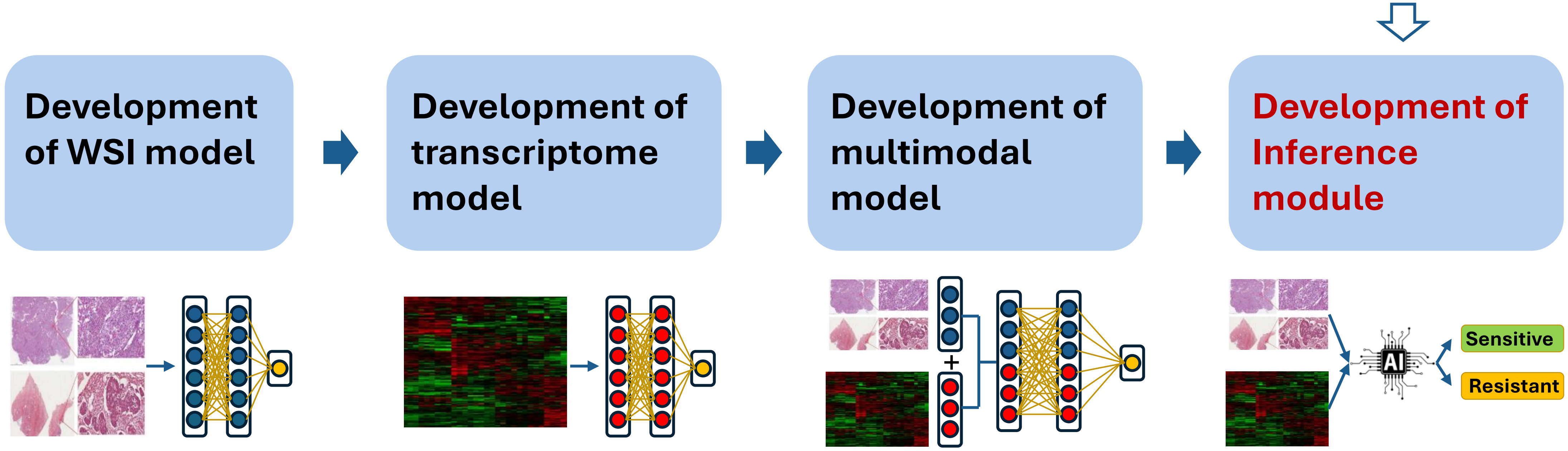
MORE INFORMATION
U.S. Department of Health and Human Services
National Institutes of Health
National Cancer Institute
USA.gov

NIH ... Turning Discovery Into Health ®

Overall Plan

Goal: Development of a Multimodal AI Model to Predict Chemotherapy Response of Ovarian Cancer

We are in this step



Development of WSI model

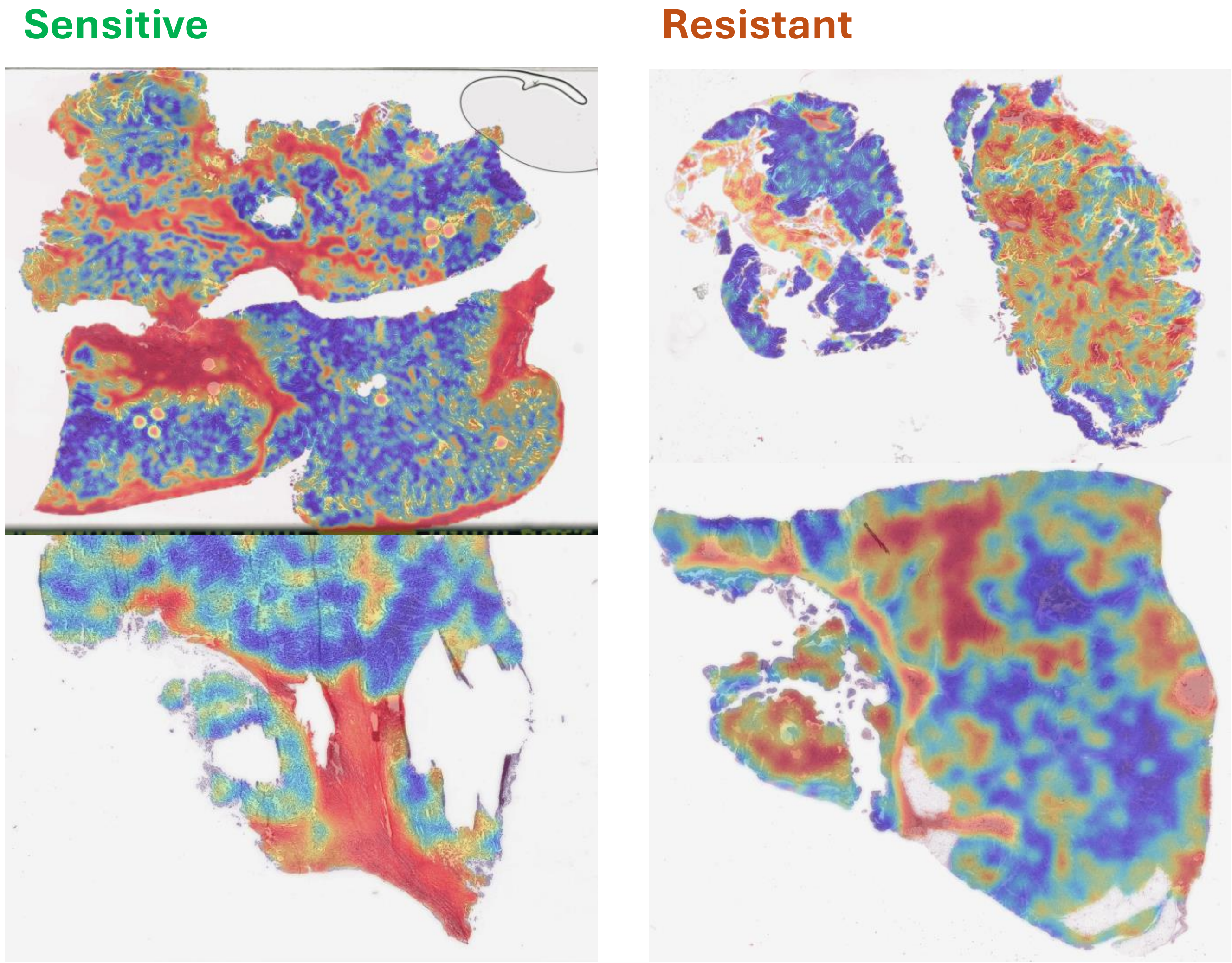
Development of transcriptome model

Development of multimodal model

Development of Inference module

Example of Attention Heatmap of Ovarian Cancer WSIs

- Tissue Segmentation
- Patch Creation
- Feature Extraction
- **Attention Scoring**
- Model Development



Sensitive

Resistant

=> Higher attention area is used more importantly for training

Acknowledgments

NCI

- Michael Sierk
- Brian Park
- Phyllip Cho
- Ziao Ye
- Kailing Chen
- Madhu Kanigicherla
- Kevin Mata Rodriguez
- Satya Gugulothu
- Daoud Meerzaman

CANCER GENOMICS CLOUD
SEVEN BRIDGES

The Velsera Miners team with special thanks to:

- Marija Gačić
- Nikola Novakovic
- Marko Trifunović
- Marko Zecevic
- Aleksander Gazibara
- Rowan Beck