

# Enhancing Medical Discovery in Cancer Center using Omicsoft Array Suite, IPA

Jun Yin, Ph.D.  
Director, Bioinformatics Shared Resource

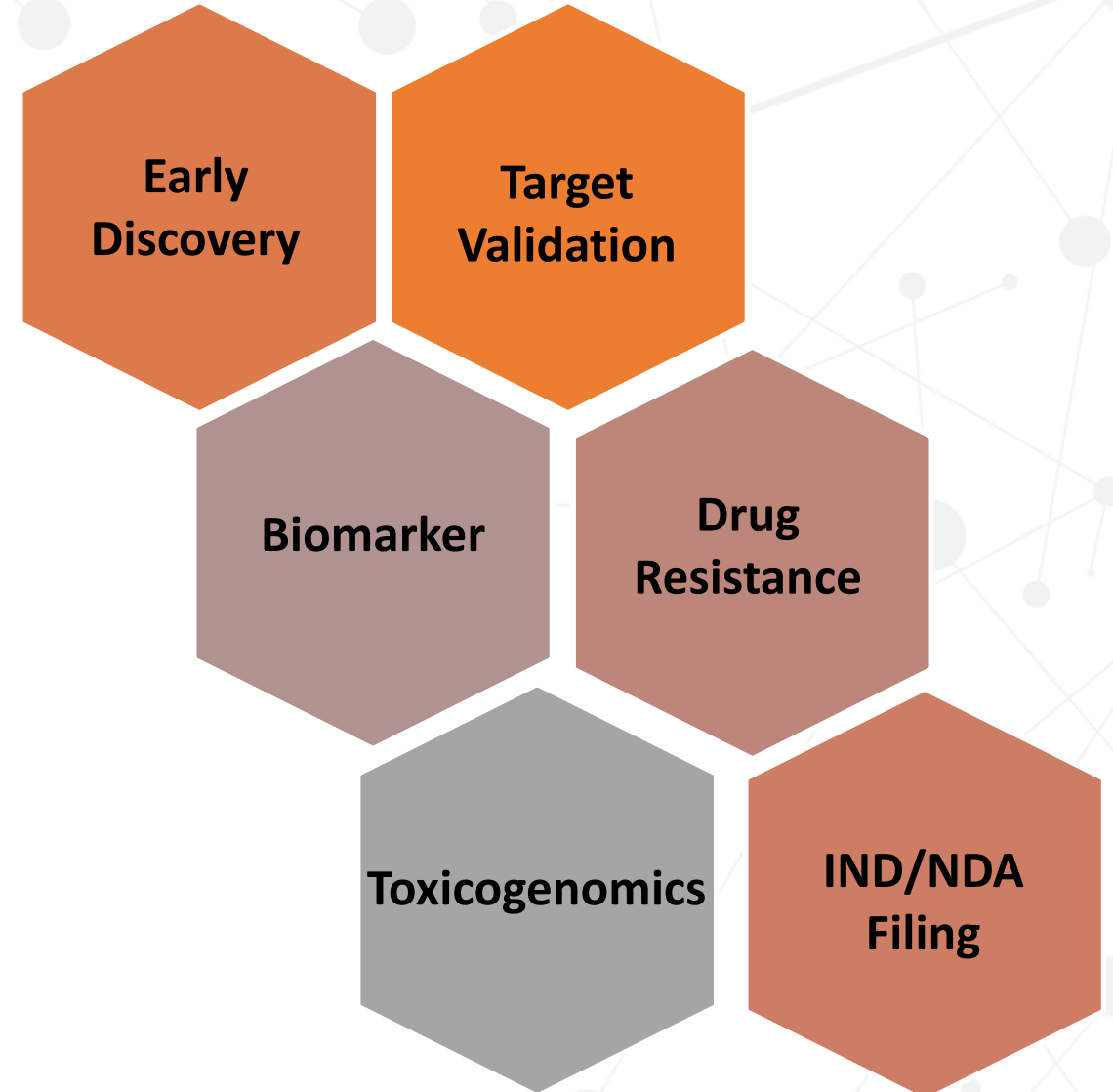
Omicsoft and IPA User Group Meeting  
9/18/2019

# Jun Yin

Director, Bioinformatics Shared Resource, SBP  
Senior Scientist, Amgen Inc.  
Postdoctoral Associate, Yale University

2019-Present  
2015-2019  
2012-2015

- Bioinformatics Lead for target discovery platform at Amgen, including CMD targets, Bi-Specific T Cell Engager (**BiTE**)
- Contributed to the FDA approval of **IMLYGIC** (oncolytic virus, melanoma) and **EVENTY** (sclerostin inhibitor for osteosclerosis)
- Supported PT, IND/NDA filings for several targets across Therapeutic Areas
- Publications in Science, Cell, PNAS etc.





## BREAKTHROUGH RESEARCH

### NCI-Designated Cancer Center

• designated in 1981, 1 of 7 in the U.S.

### Neuroscience and Aging Research Center

### Infectious and Inflammatory Disease Center

### Sanford Children's Health Research Center

### Conrad Prebys Center for Chemical Genomics

## DRUG DISCOVERY

60+

DRUG  
DISCOVERY  
PROJECTS

OVER \$100M OF  
TRANSLATIONAL  
SCIENCE GRANTS  
IN LAST 5 YEARS

ONE OF THE MOST ADVANCED  
NONPROFIT DRUG DISCOVERY  
INFRASTRUCTURES

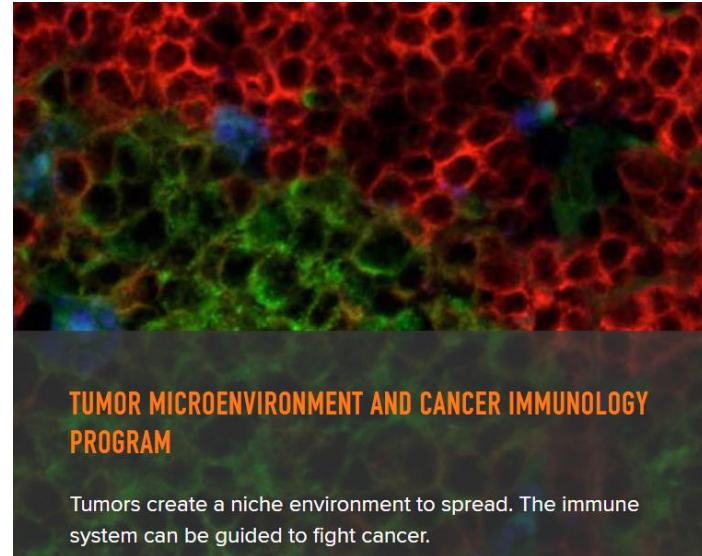
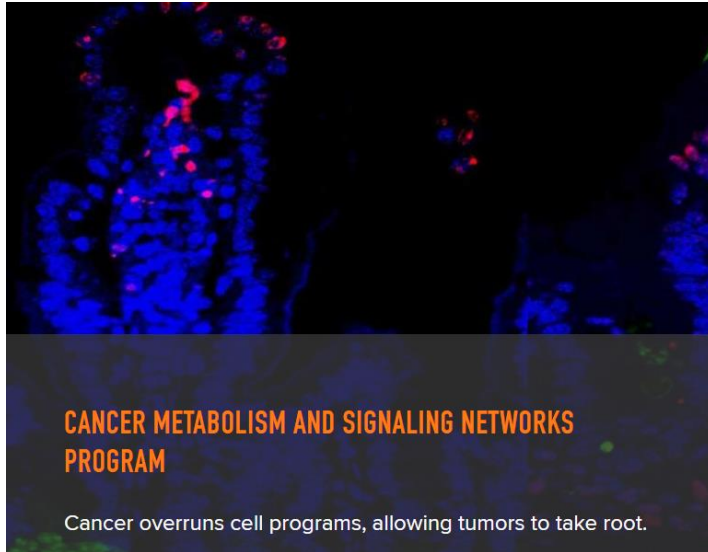
3

Clinical trials underway  
Immune therapy • Calcification  
Pancreatic cancer

4

FDA-approved treatments  
Epogen® • Targretin®  
Strensiq® • Venclexta™

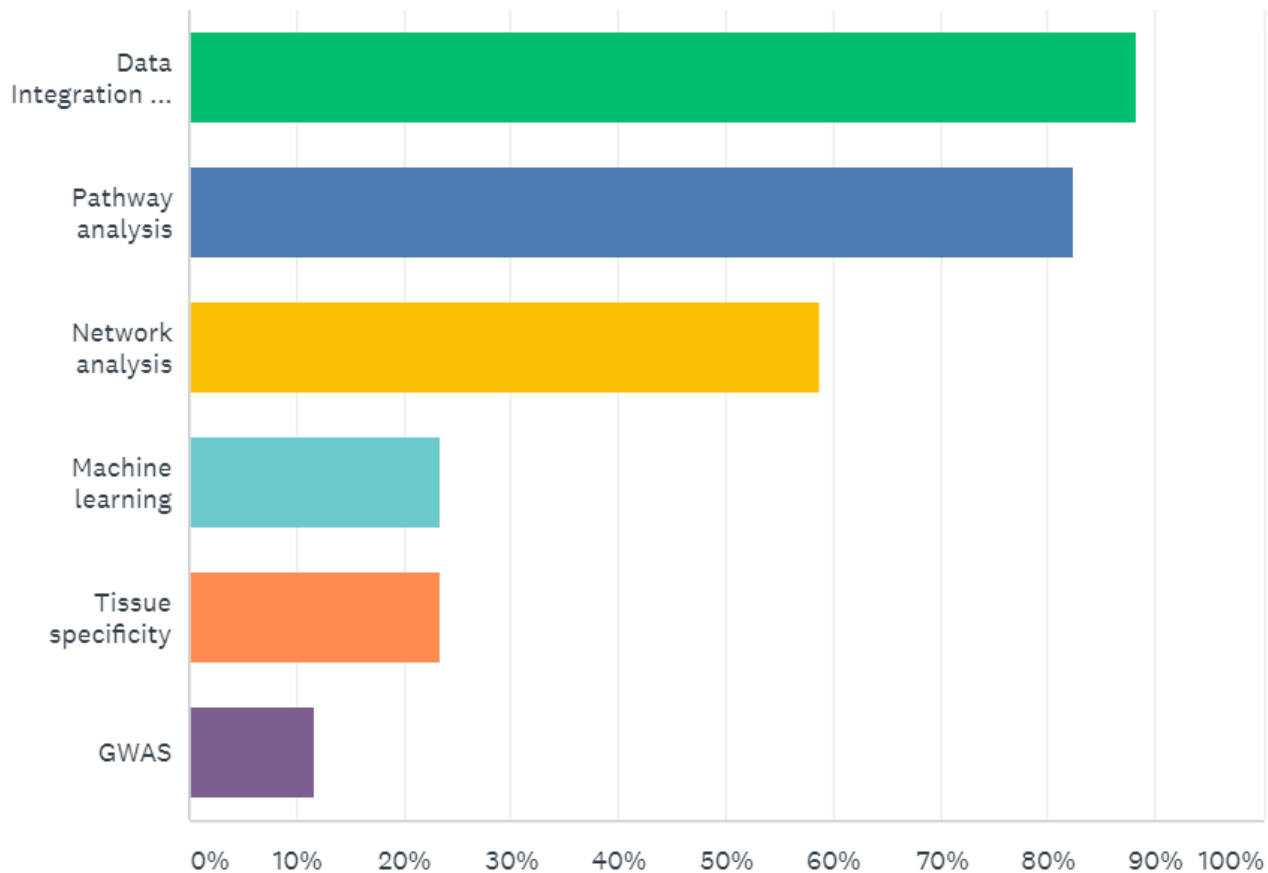
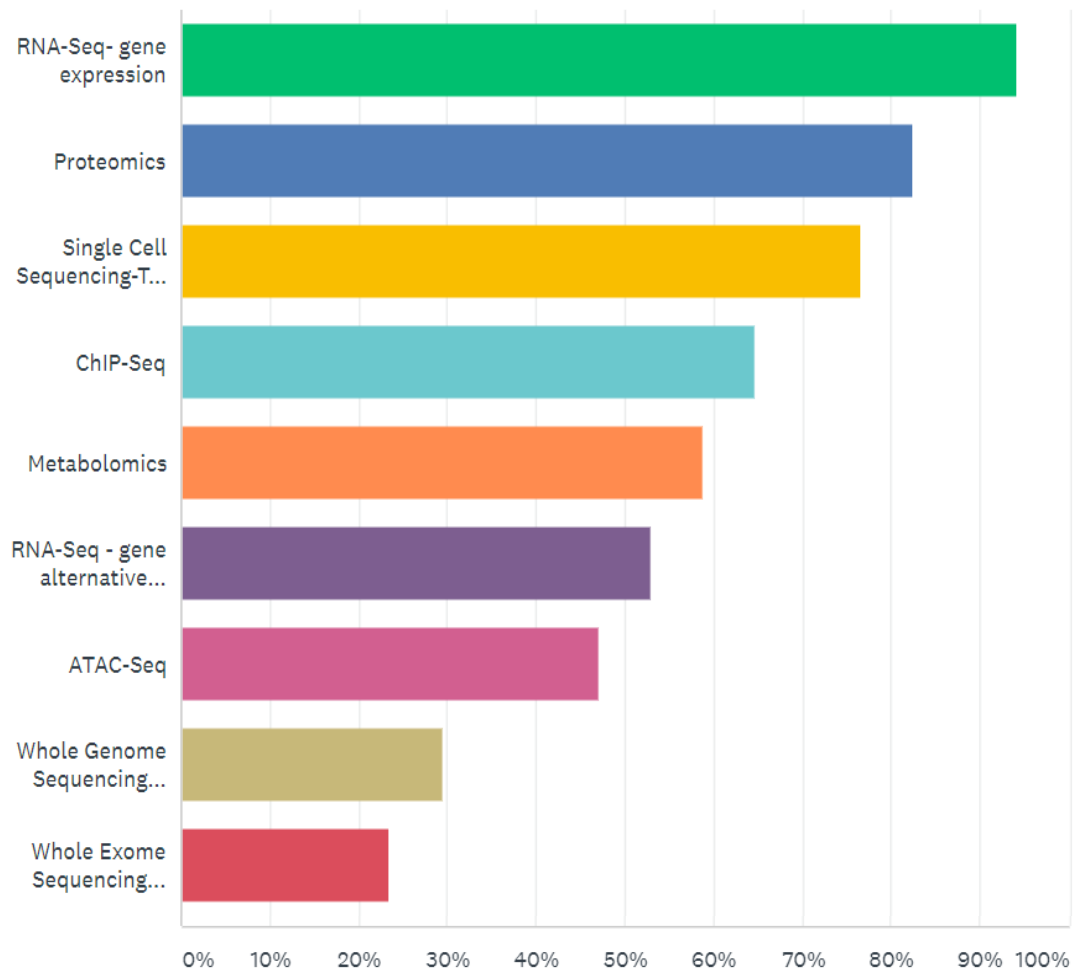
## CANCER CENTER PROGRAMS



- Preeminent status, one of the seven Basic Laboratory Cancer Centers in US
- Supported by NCI Cancer Center Support Grant (CCSG)
- About 40 PIs in three cancer programs
- Research ranges from drug discovery, screening, validation, drug resistance, cancer mechanism/metabolism, preclinical/clinical programs

**Supported >70 publications in the last few years  
>18 grant applications ...**

## INTEREST LEVEL IN INTERNAL SURVEY



## MAJOR SERVICES

### *Streamlined Biological Data Analysis*

- RNA-Seq, ChIP-Seq, ATAC-Seq
- Proteomics
- Preprocessed TCGA, CCLE, GTEx
- Ingenuity Pathway Analysis (IPA), Metacore, GSEA

### *Omics Data Mining and Integration*

- Multi-Omics data integration
- Single Cell Sequencing
- Network analysis
- Translational research for therapeutic targets and biomarker identification

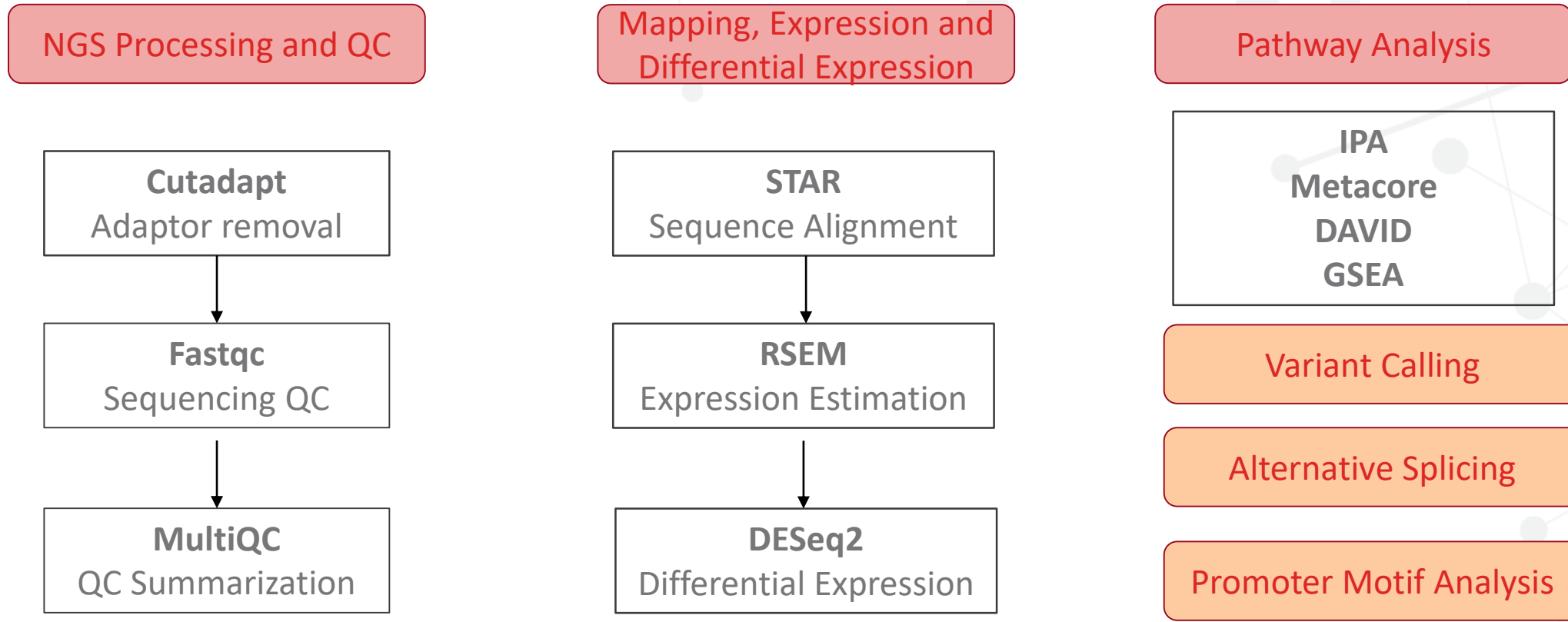
### *Biostatistics and Machine Learning*

- Machine Learning technologies, e.g. Bayesian network, Neural Network, Random Forest
- Statistical consultation e.g. survival analysis, linear regression

### *Computational Infrastructure and Training*

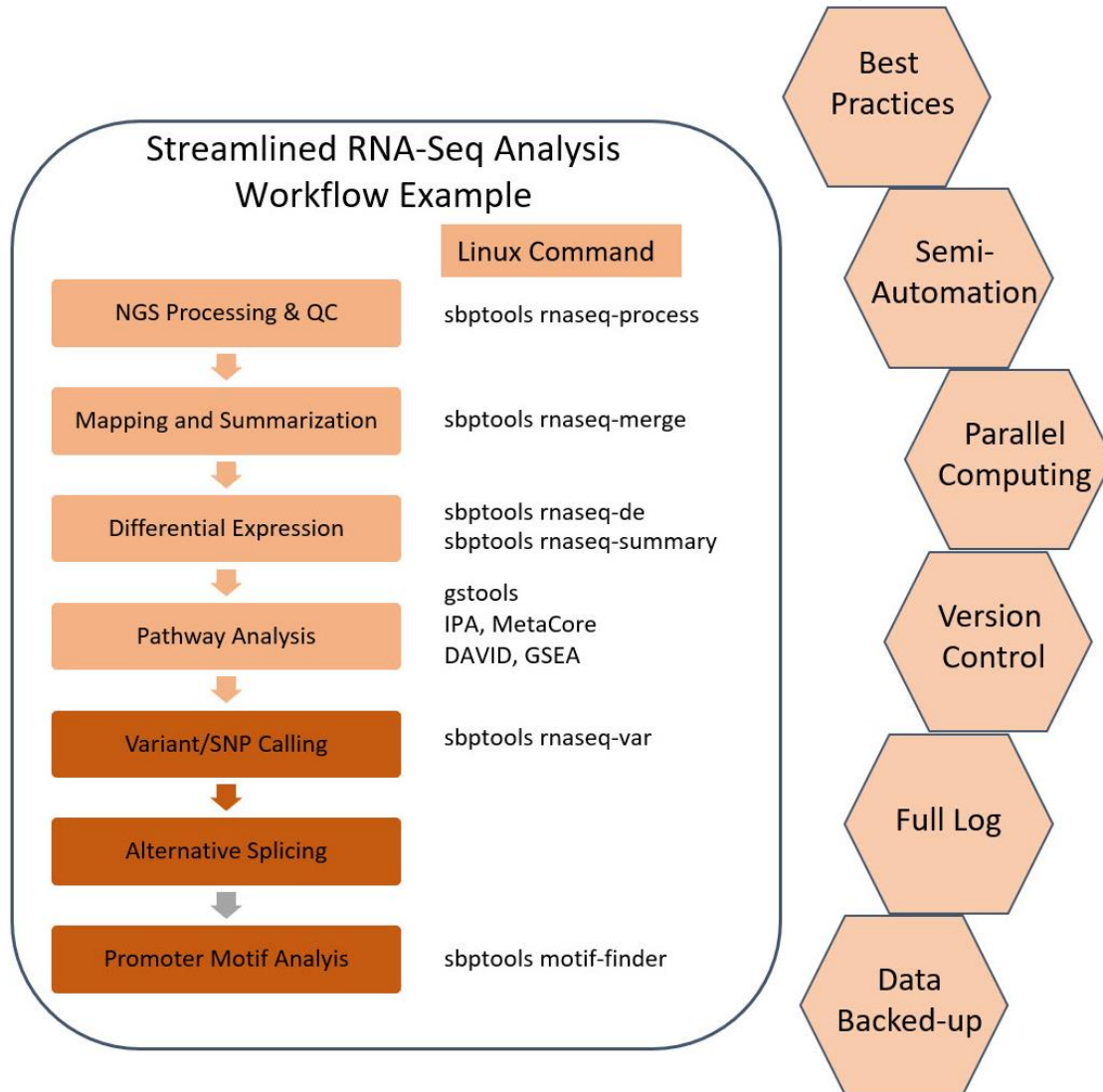
- Automated pipeline for High Throughput Screening
- Computational pipeline for lab routine analysis

## EXAMPLE: RNA-SEQ DATA ANALYSIS PIPELINE



- Similar to the standard pipeline used by ENCODE
- Alternative sequence alignment/summarization using Omicsoft oshell/oscript
- Omicsoft Array Suite for data visualization and sharing

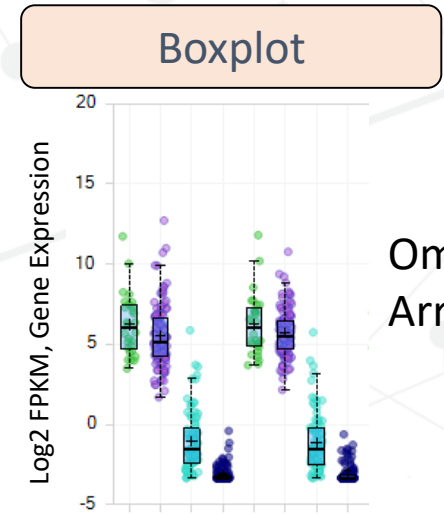
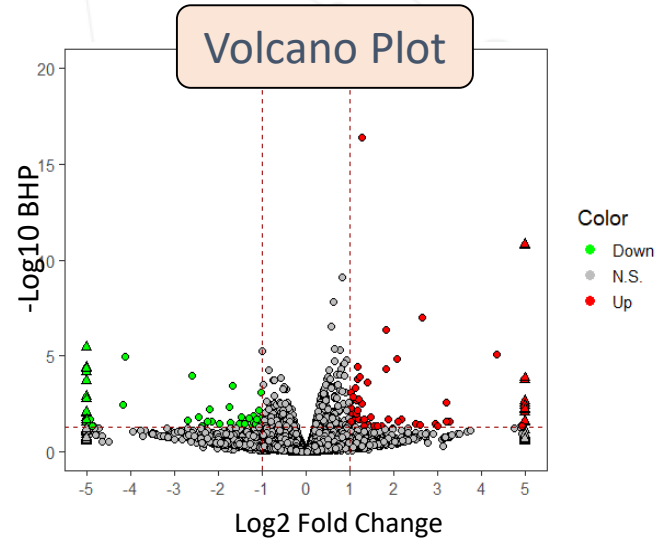
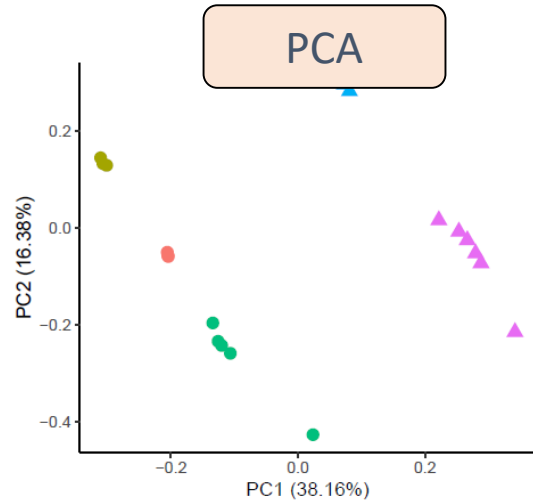
# EXAMPLE: RNA-SEQ DATA ANALYSIS PIPELINE



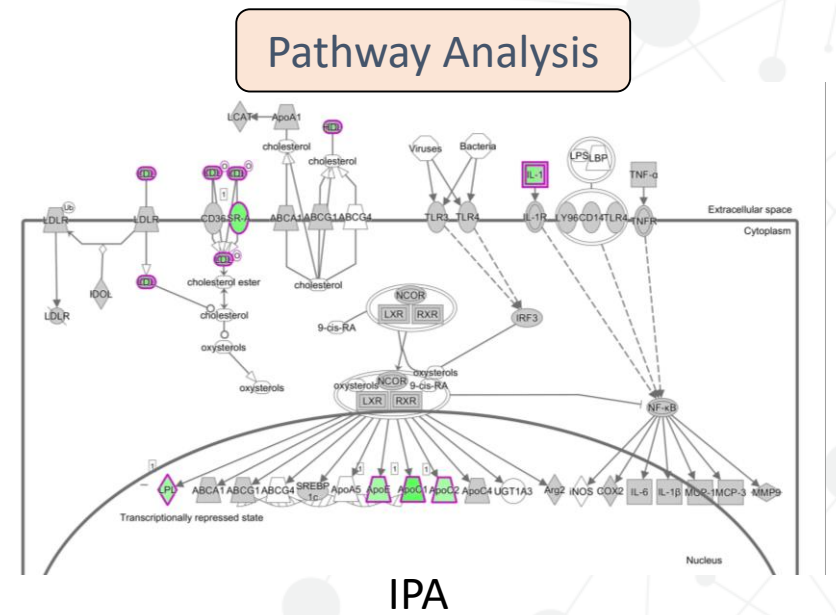
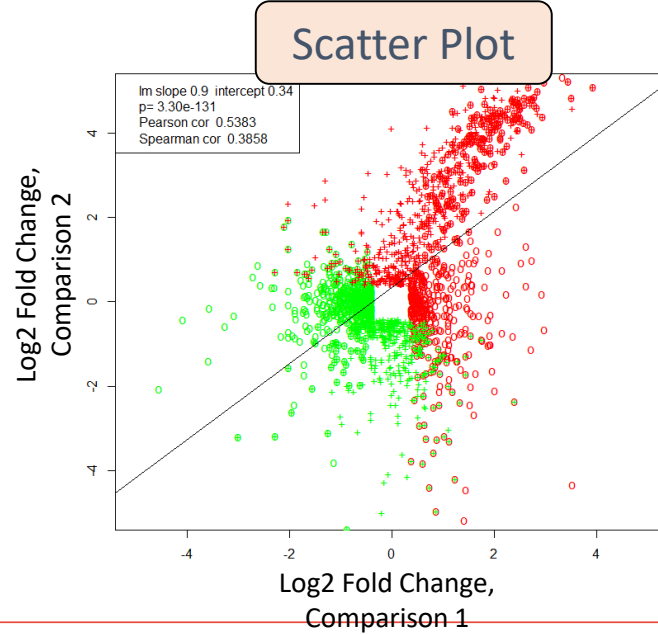
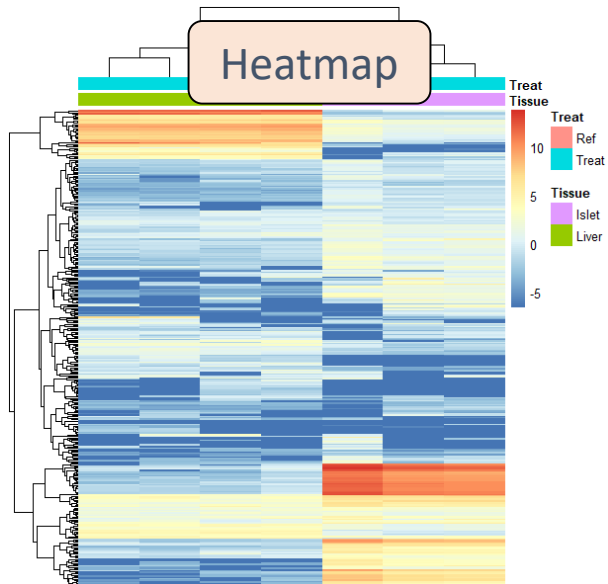
- Automated command line implementation in Linux
- Fast processing, parallel computing, standard genome and gene annotation version, software version control
- From Fastq, QC, alignment to DE gene list, and pathway analysis



# EXAMPLE: RNA-SEQ DATA ANALYSIS PIPELINE



Omicsoft  
Array Studio



# METASCAPE, PATHWAY ANALYSIS TOOL FROM SBP



**Step 1**

**Multiple Gene List**  
 Drag & drop your file (.xls,.xlsx,.csv,.txt)

Select File...

Or paste a gene list

Accept Gene ID/Symbol/RefSeq/  
 Ensembl/UniProt/UCSC

Upload File Format

Single List:  
 .xls/xlsx .csv .txt

Multiple List:  
 .xls/xlsx .csv .txt

Test Upload

single list  
 3 gene lists

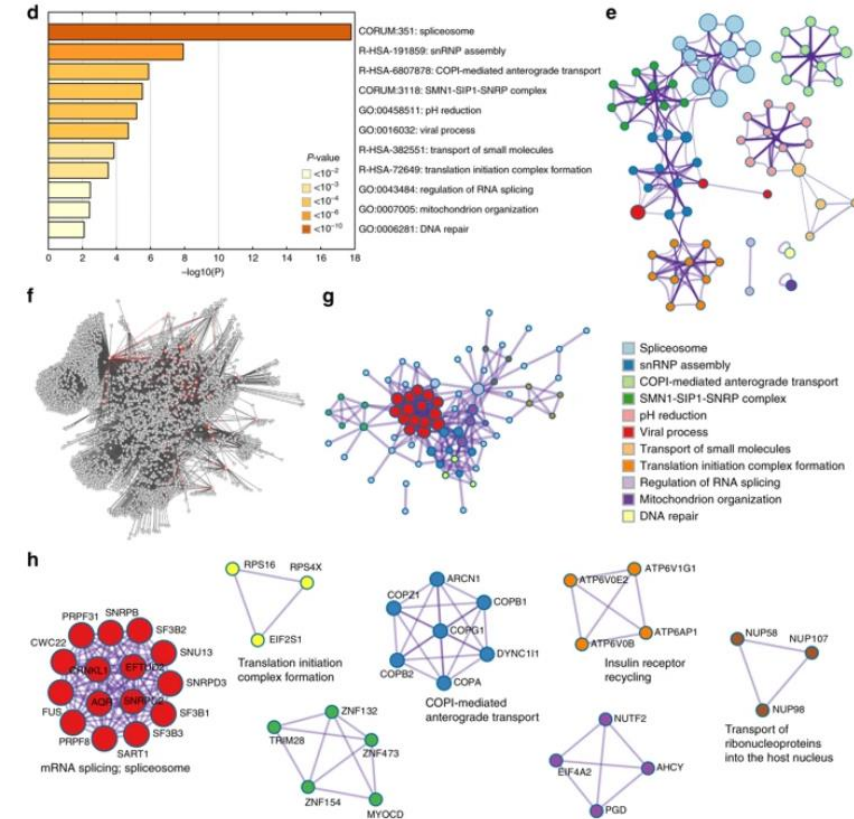
Test Identifiers

Gene Symbol [try it!](#)  
 RefSeq  
 Entrez Gene ID

**Step 2**

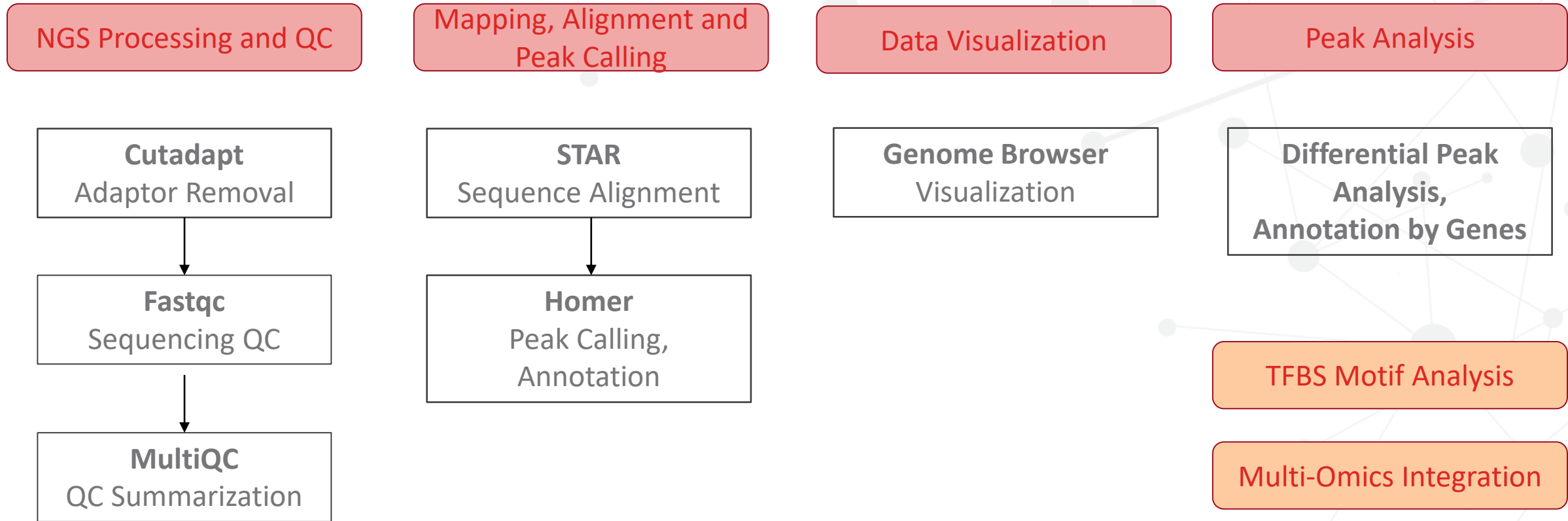
**Step 3**

Express Analysis Custom Analysis



Zhou et al. 2019. Nature Communications

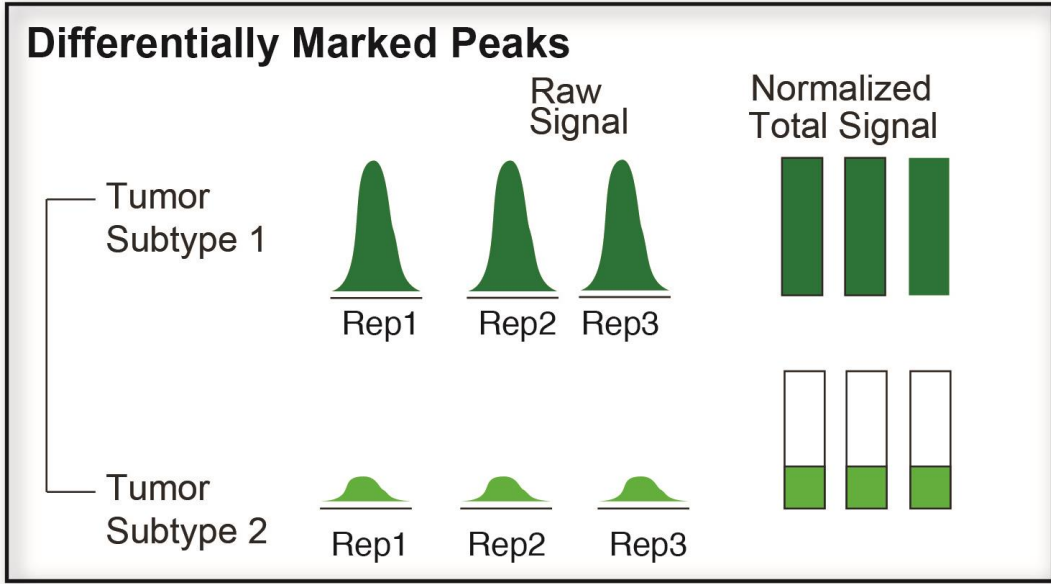
# CHIP-SEQ/ATAC-SEQ COMPUTATIONAL WORKFLOW



- Fast processing, parallel computing, standard genome and gene annotation version, software version control
- From Fastq, QC, alignment to DE gene list, and pathway analysis

# CHIP-SEQ/ATAC-SEQ COMPUTATIONAL WORKFLOW

Differential Marked Peak Calling

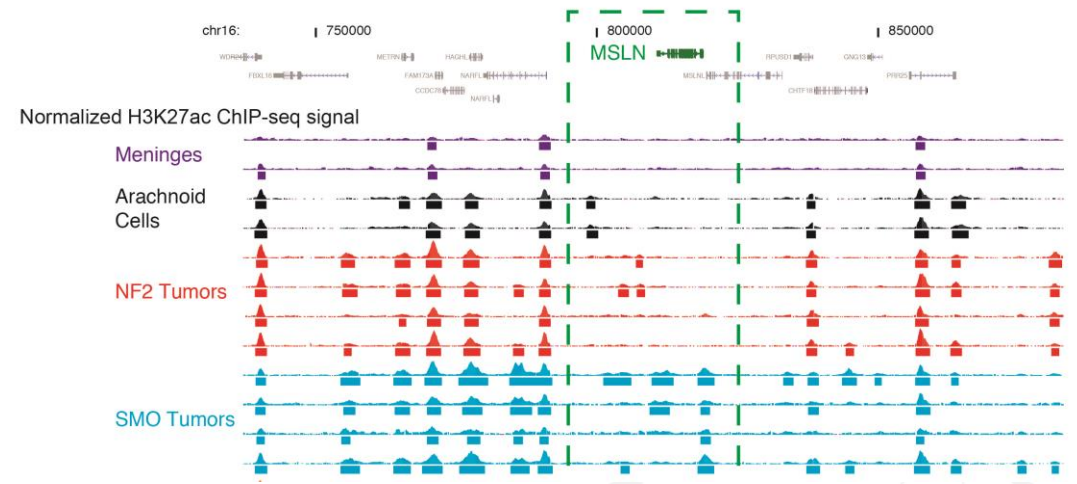


- Find reproducible peaks from replicates
- Merge reproducible peaks from different groups
- Compare ChIP-Seq signal between groups

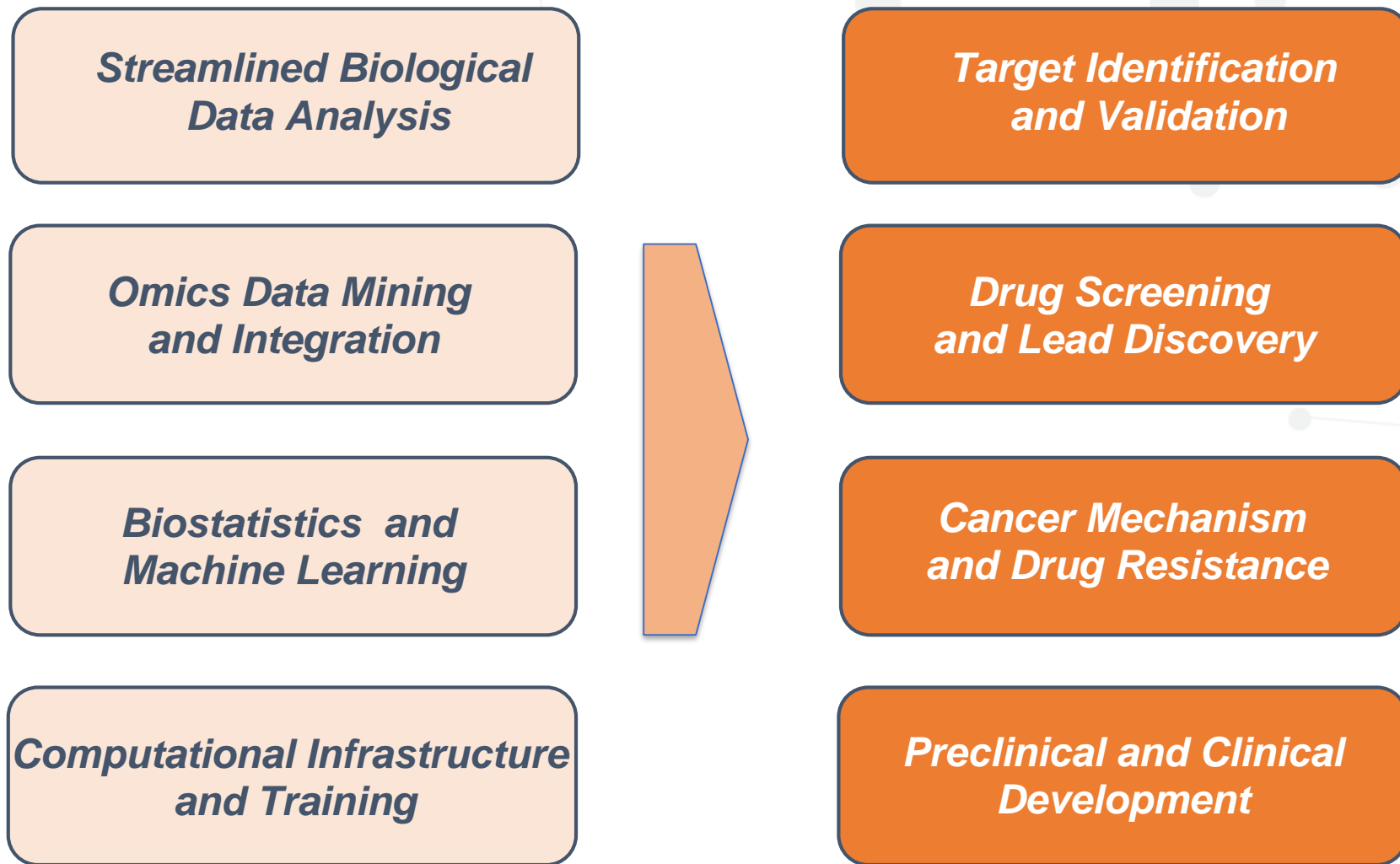
Motif Analysis



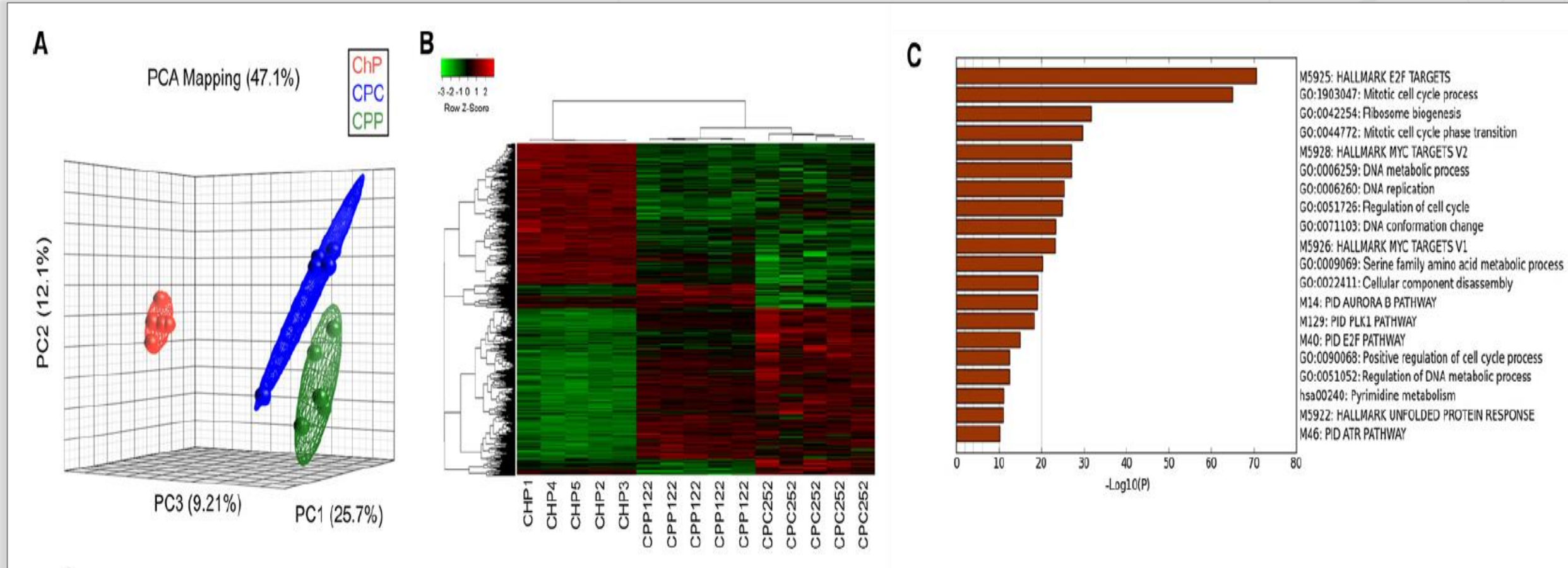
Genome Browser



## SERVICES ORIENTED TO PROMOTE CANCER RESEARCH AND DRUG DEVELOPMENT

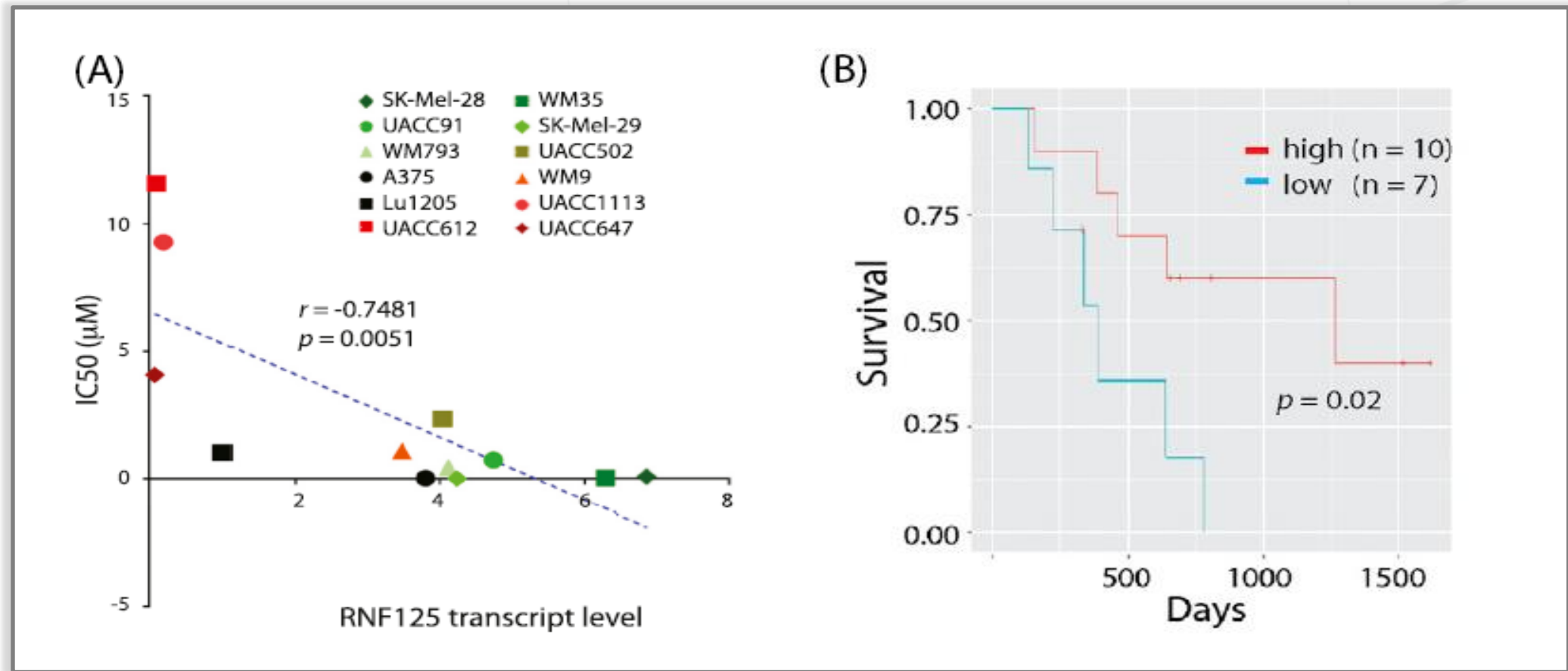


# EXAMPLE: DRUG TARGET DISCOVERY IN CHOROID PLEXUS TUMOR



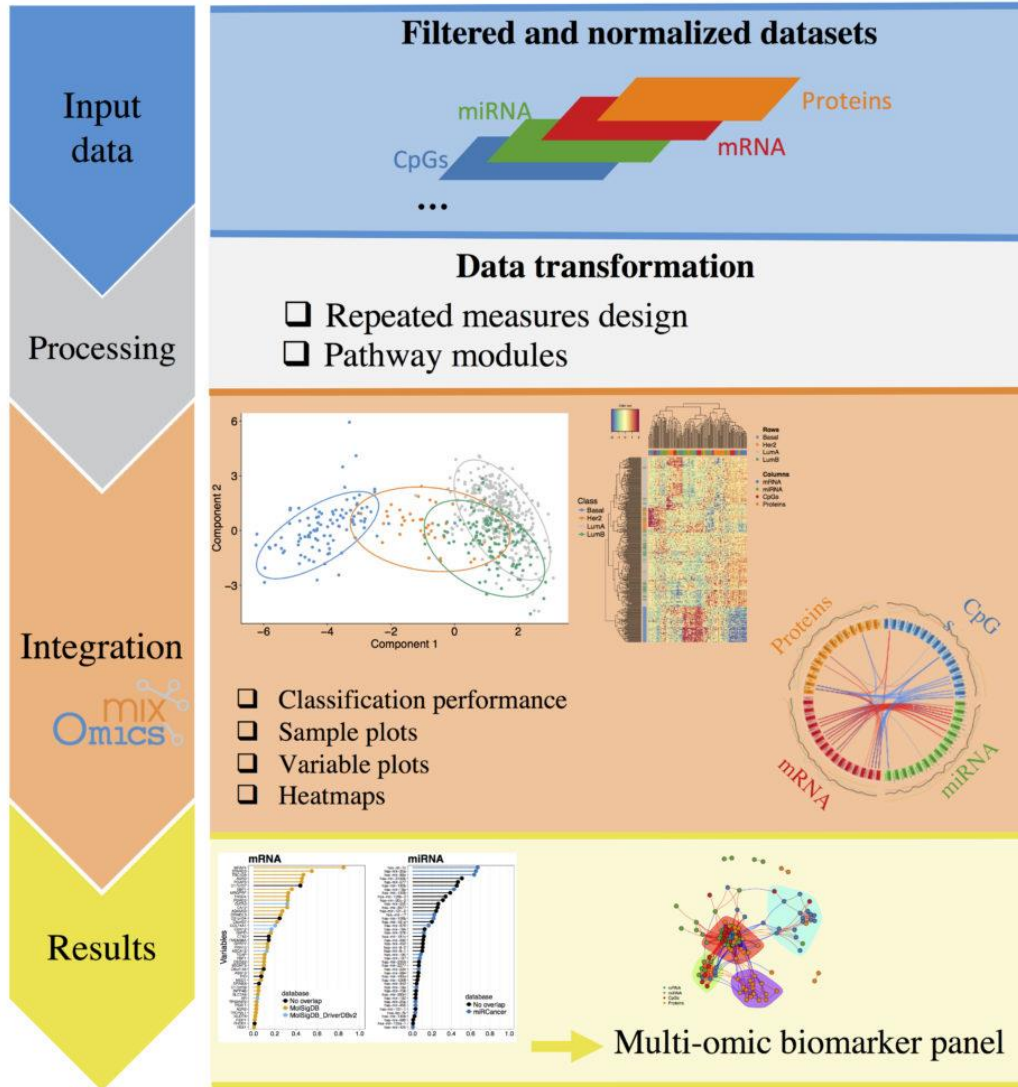
Wang J, et al. 2019. Myc and loss of p53 cooperate to drive formation of choroid plexus carcinoma. *Cancer Research*.

## EXAMPLE: BRAF INHIBITOR DRUG RESISTANCE IN MELANOMA



Kim H, et al. Downregulation of the Ubiquitin Ligase RNF125 Underlies Resistance of Melanoma Cells to BRAF Inhibitors via JAK1 Deregulation. *Cell Rep*.

# MULTI-OMICS DATA INTEGRATION



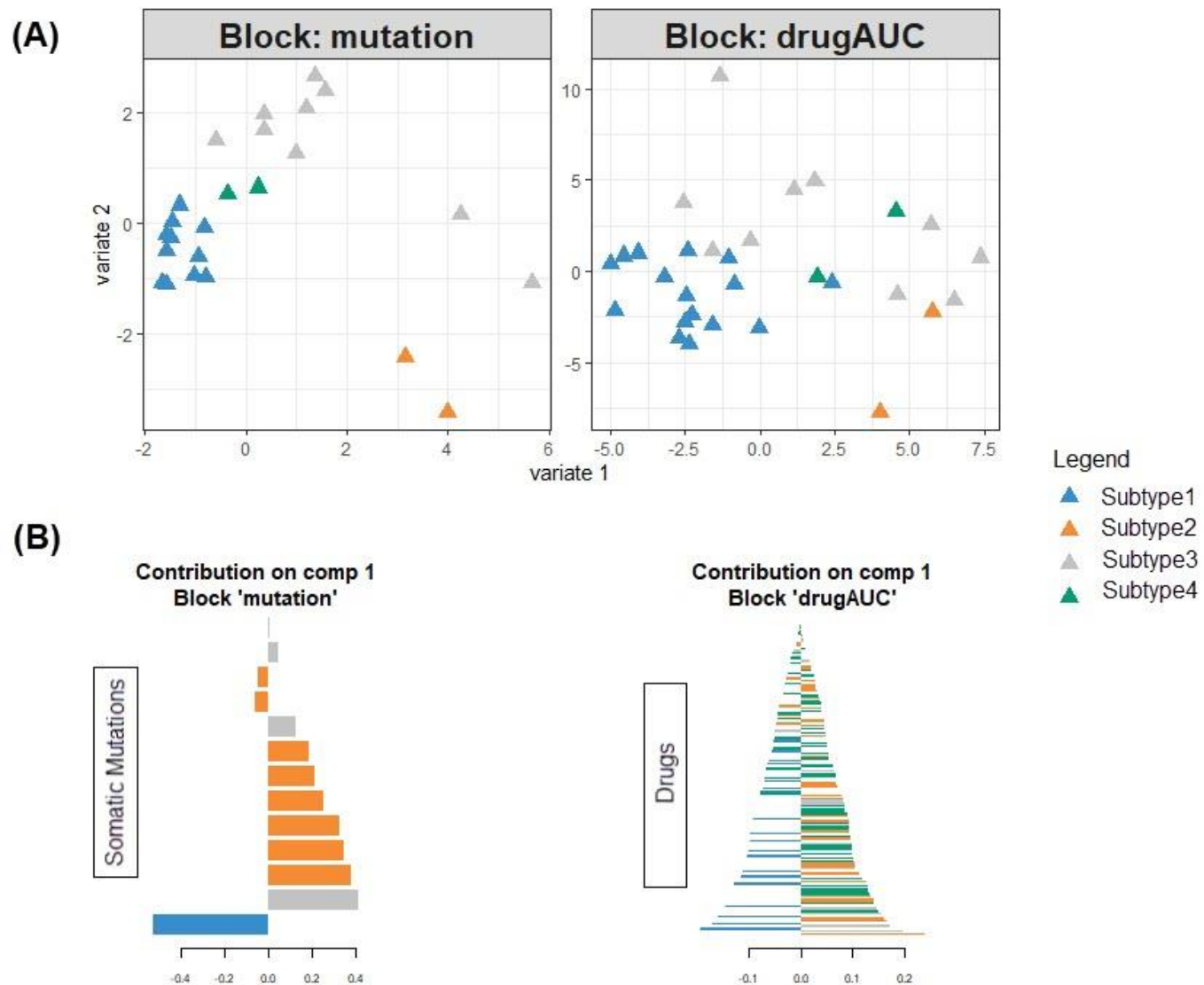
D  
I  
A  
B  
L  
O

- High requirement for the data
  - Samples need to be matched for different technologies
  - Large number of replicates (5-6 samples per group)

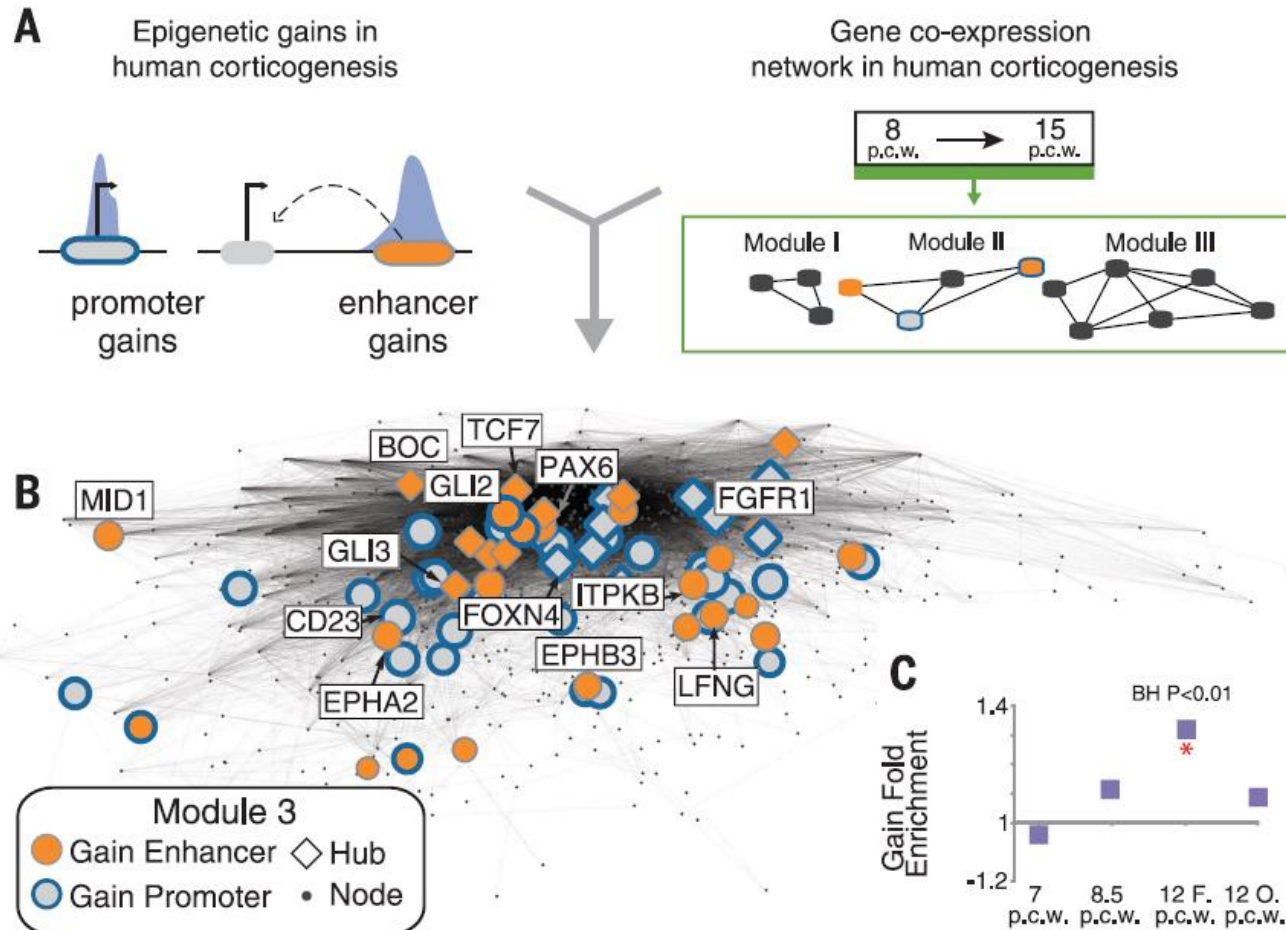
Andrew Hodges



# EXAMPLE: DATA INTEGRATION OF SOMATIC MUTATION AND DRUG SCREENING



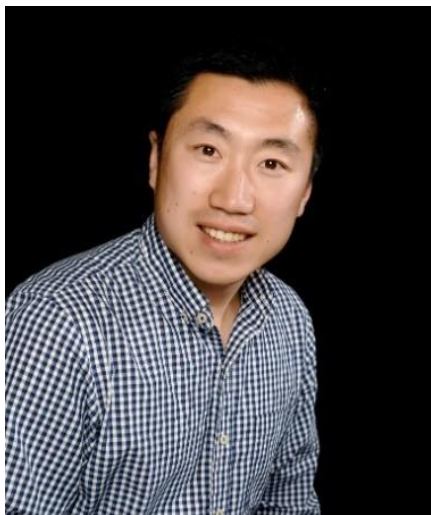
# EXAMPLE: MULTI-OMICS DATA INTEGRATION FOR RNASEQ AND CHIPSEQ



- Overlay Epigenomics features on gene expression network

Reily, S.\*, Yin, J.\* etc. 2015. Science.

## BIOINFORMATICS CORE STAFF



Jun Yin, Ph.D.  
Director, Bioinformatics Core

Genomics

Data mining

Drug Discovery



Andrew Hodges, Ph.D.  
Bioinformatics Project Manager

Machine Learning

Drug Screening

Data Integration



Rabi Murad  
Bioinformatics Scientist

Genomics

Multi-Omics

Pipeline Dev



Zoe Li,  
Bioinformatics Technician

Data Management

Software Engineering