

# TiP – A Genomics Platform Facilitated by OmicSoft Curated Land Data

Omicsoft User Group Meeting, September 18, 2019  
Yonghong Xiao, substituting Lihua Yu  
H3 Biomedicine, Cambridge MA



## H3

Human. Health. Hope.

An Eisai oncology company



# Outline



- H3 - A data driven cancer therapeutics company to create precision medicine; data driven drug discovery and development is our goal
- H3 - Data Science:
  - Focus on data democratization – data access and usability designed with users (biologists) and scientific questions in mind as our guiding principle.
  - Focus on translate data into actions – therefore our philosophy of collaborating with strategic partners in most other areas.
- TiP – Translational Informatics Platform
  - Its position in H3 genomics eco-system
    - Data integration
    - Data reporting
    - Iterative Data query
  - How do we work with strategic partners – Omicsoft Oncoland data
  - Benefits of using Omicsoft land data vs. internal processed of data

# H3 Biomedicine: Human • Health • Hope



## A CLINICAL STAGE ONCOLOGY BIOTECHNOLOGY COMPANY

FOUNDED | 2011

FOCUS | Novel and highly targeted oncology treatments

HEADQUARTERED | Cambridge, Massachusetts

ORGANIZATION | Private

BUSINESS STRUCTURE | Wholly owned subsidiary of Eisai Co. Ltd.

### Human.

H3 scientists analyze cancer patients' data to uncover new disease insights and deliver impactful medicines to patients.



### Health.

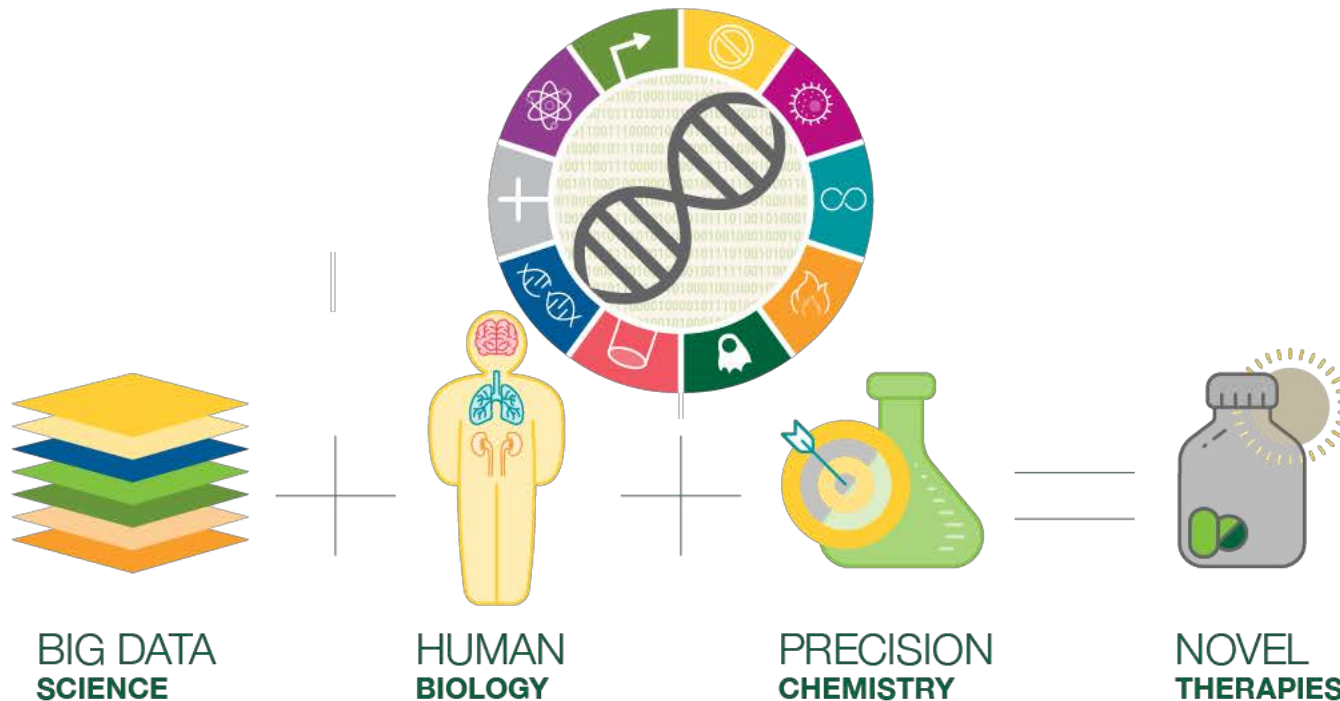
We believe in the use of big data science and precision chemistry which aim to deliver the right drug to the right patient.



### Hope.

We aspire to provide hope. Hope for patients and their families. Hope for health care professionals involved in cancer care.

# H3 – To unlock cancer genomic clues, create precision medicine, and transform hope into health



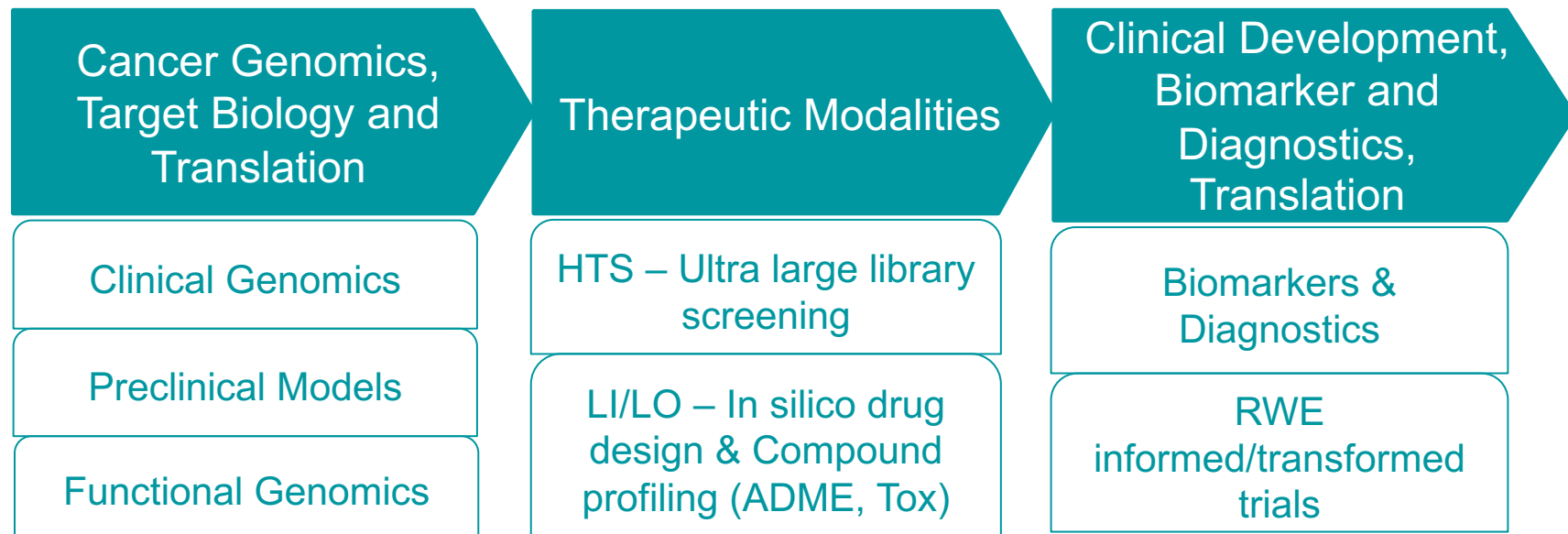
H3 is responsible *for advancing innovative new medicines from discovery to clinical proof-of-concept guided by hhc\* concept.*

hhc: human health care

# H3 Data Science: Aspire to be the most impact focused organization



|             |  |
|-------------|--|
| Lead        | Data Culture                                   |
| Democratize | Data access to all                             |
| Transform   | Data into knowledge, intelligence and decision |



**Data Practice/Culture**  
Data Tracking – Management – Analysis, Visualization, Exploration

# Big Data, Deeping Mining – Human Genetics Driven Targets



|  |                              |  |
|--|------------------------------|--|
|  | <b>Cancer Genomics</b>       | <b>TCGA</b>  |
|  | <b>Diagnostic Sequencing</b> | <b>ICGC</b>  |
|  | <b>Clinical Genomics</b>     | <b>AACR</b> <small>American Association for Cancer Research</small><br><b>PROJECTGENIE</b> <small>Genomics Evidence Neoplasia Information Exchange</small> |

|  |                                 |   |
|--|---------------------------------|---|
|  | <b>Model Genomics Profiling</b> | <b>CCLC</b> <small>Cancer Cell Line Encyclopedia</small><br><b>Cell Lines Project</b> <small>Mutation profiles of</small>       |
|  | <b>Genetic Screening</b>        | <b>Project Achilles</b><br><b>CTD<sup>2</sup></b> <small>Cancer Target Discovery and Development</small>                        |
|  | <b>Chemical Screening</b>       | <b>GDSC</b> <small>The Genomic Drug Sensitivity in Cancer</small><br><b>ChemPartner</b> <small>Dedicated to LifeScience</small> |

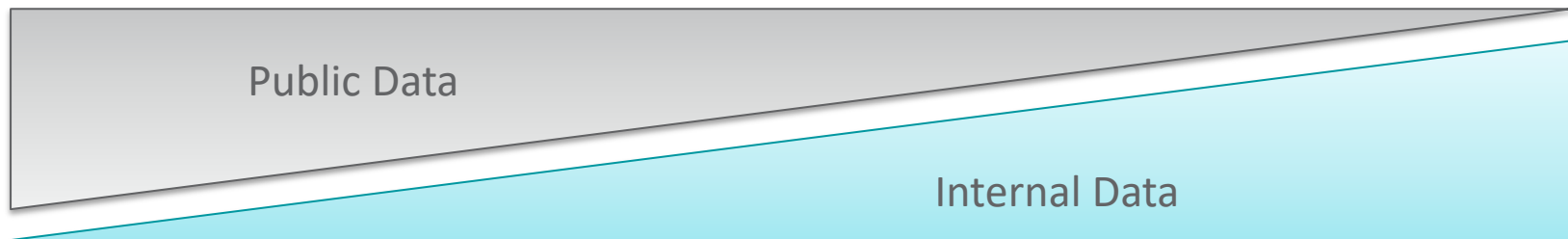
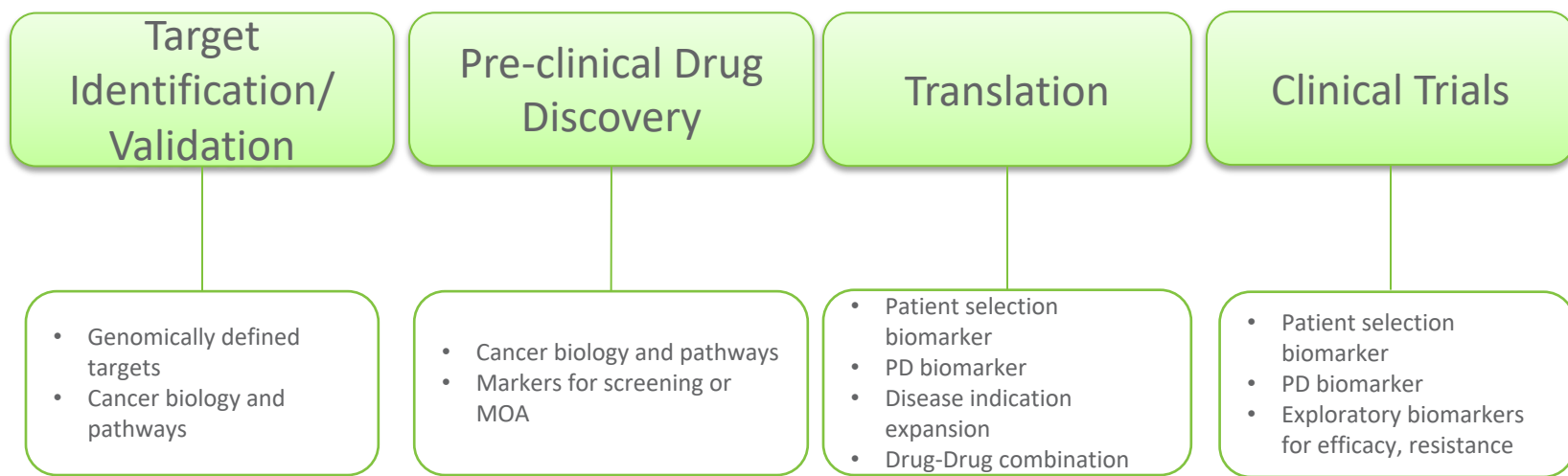
## Deep Mining Platforms

|  |  |
|--|--|
|  | <b>H3 Biomedicine TiP</b><br><small>Translational Informatics Platform</small> |
|  | <b>Pharmacogenomics Correlation Analysis Portal</b>                            |
|  | <b>cBioPortal</b><br><small>FOR CANCER GENOMICS</small>                        |
|  | <b>depmap</b>  |
|  | <b>SevenBridges</b>  |



# H3's Genomics Ecosystem

Goals: Leverage both internal and external, current and historical data across portfolio and throughout our discovery/development cycles for idea testing/generation and data driven decision making.



# The Data Ecosystem



## Genomics Data

- External
- Internal

## H3 Data Science+Partners

- Data Collection
- Data cleaning, standardization, re-annotation
- Data processing, analysis

## TiP: A Path to Leveraging Genomics Data

- Data reduction and integration
- Query, visualization and communication
- Pharmacogenomics internal, external data integration

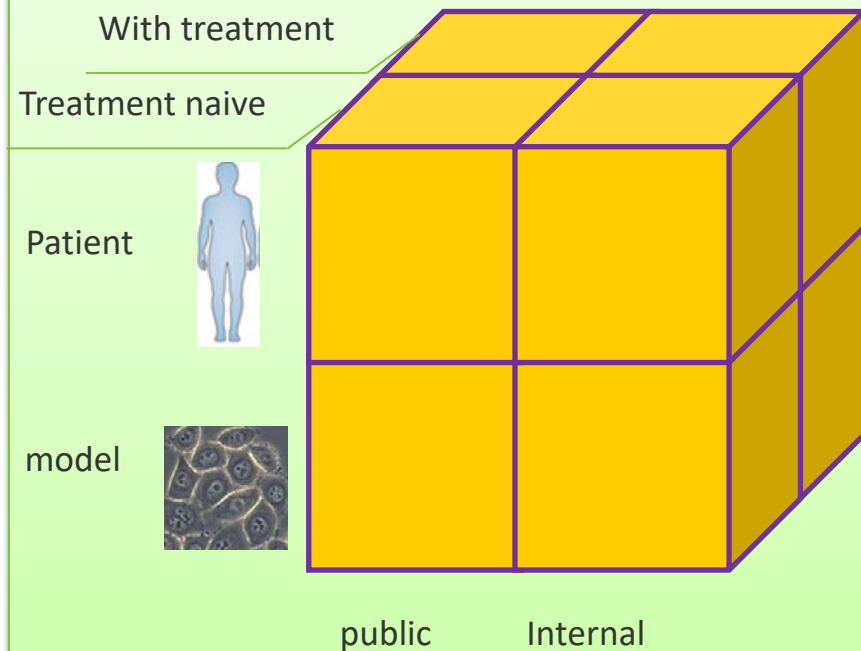




# TiP – Data source



- Variety of categories



- All data together for one stop shop

Focus on

- Cancer genomics
- Functional genomics
- Compound profiling

Agnostic

- Internal vs. external
- Patient samples vs. pre-clinical models
- w/ or w/o treatment

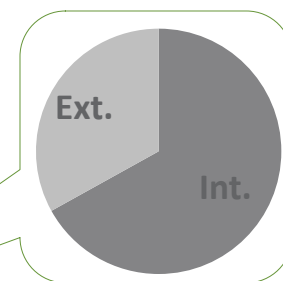
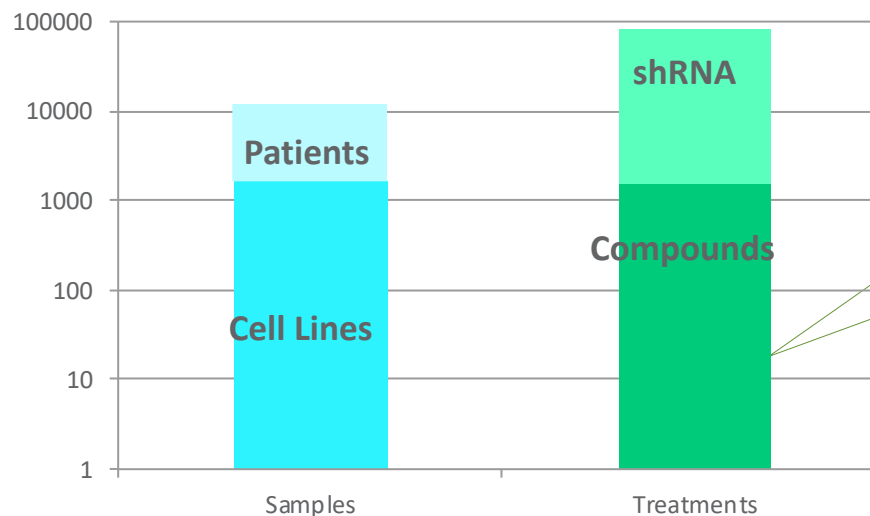


# TiP – Data Overview - circa 2014

## Datasets

CCLE  
Sanger  
CTD2  
GDSC  
NCI-60

Achilles  
COLT OICR  
H3 Internal  
TCGA



9  
Datasets

18 Protocols  
+  
123K Entities

10 Data Types  
+  
74 Annotation  
Types

881M Data  
Points  
+  
368K  
Annotations

All external data were cleaned and processed at H3



# TiP – Data Overview - circa 2019

~80% of the data are from in Omicsoft Oncoland data subscription

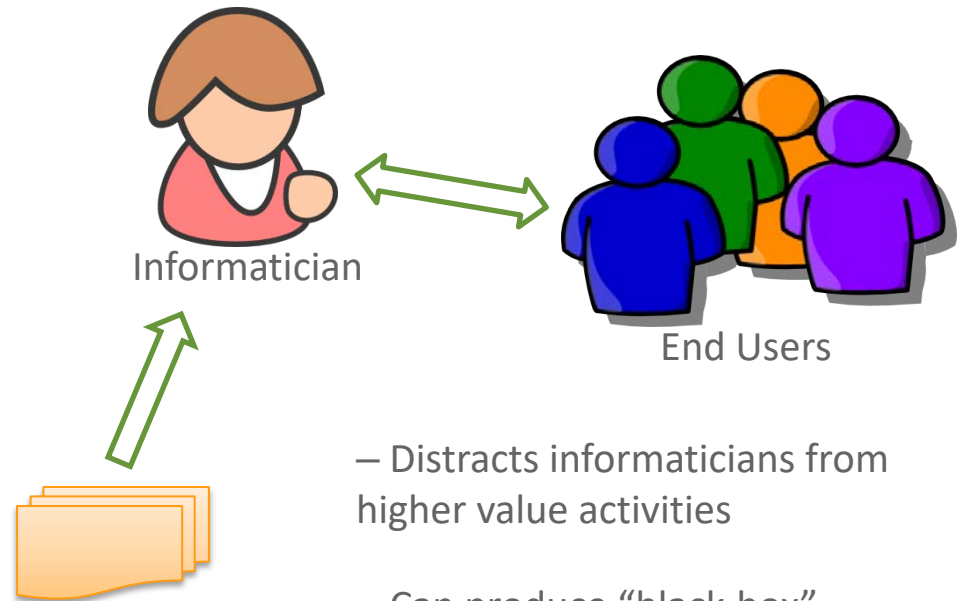
- Patient tumor: The Cancer Genome Atlas (**TCGA**), International Cancer Genome Consortium (**ICGC**), Multiple Myeloma (**CoMMpass**)
- Cancer cell lines: Cancer Cell Line Encyclopedia (**CCLE**), Sanger Cancer Cell Line
- Drug response and cancer genomics: Genomics of Drug Sensitivity (**GDSC**), Cancer Target Discovery And Development (**CTD<sup>2</sup>**)
- Genetic perturbation: **Achilles** project (shRNA, sgRNA)
- **H3 internal** cell line compound treatment data

# User Strategy for Delivering Data



## Traditional Strategy:

- Informatician as Data Broker
- Uses technical skills to extract, integrate and analyze data
- Interfaces with business to answer specific questions



- Distracts informaticians from higher value activities
- Can produce “black-box” experience for end users
- Even best informatics group has a bandwidth limit
- Especially difficult for exploratory analysis

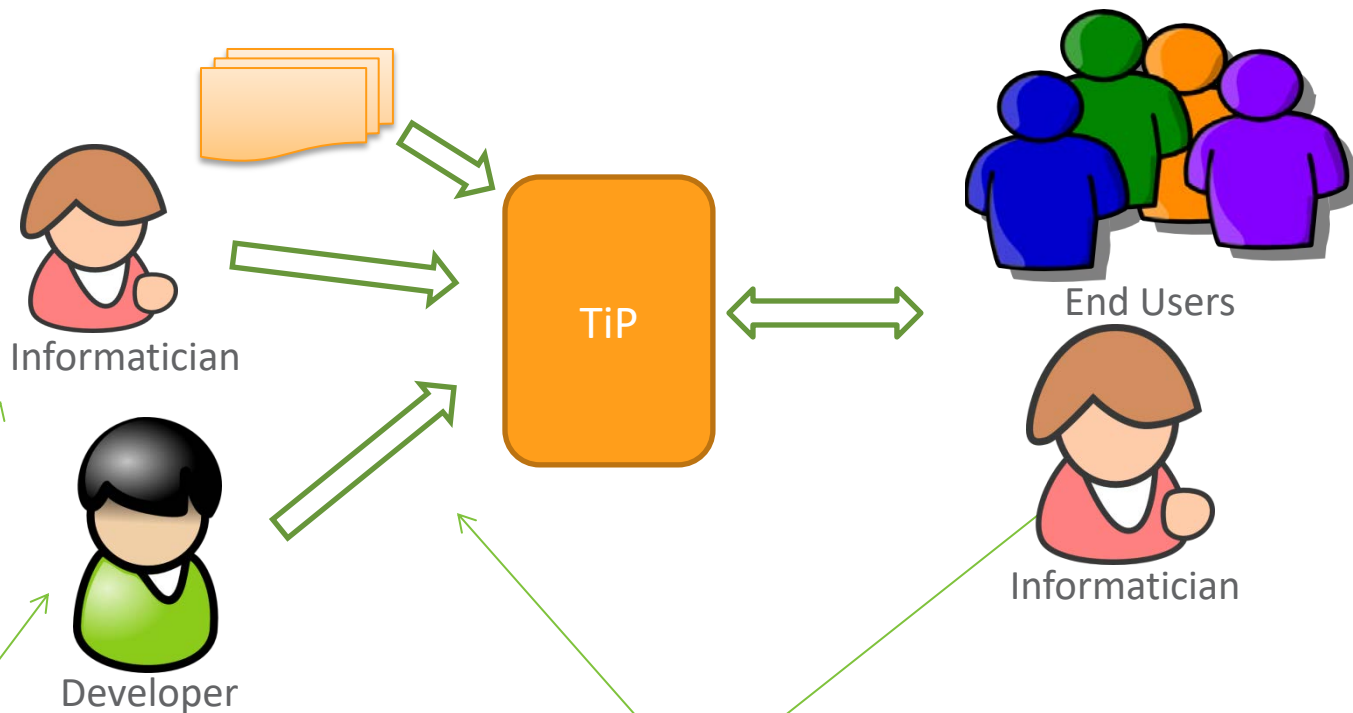


# User Strategy for Delivering Data

Informatician as Data Steward

Ensure data is:

- Clean
- Consistent
- Clear
- Unambiguous



Gaps

S/W Technical vs. Domain  
– High complexity of data  
– Developers need to understand meaning and purpose

Natural tension between how informatician and biologist will use system  
  
Where possible – hide complexity, use sensible defaults, remove the necessity to know about the technical detail of data  
  
Partner to deliver highest value



# Iterative User Query Interface

Heavy design focus on usability

- User Centered Design
- Story Boarding / Prototyping
- Use of familiar concepts (Amazon facets, shopping cart)
- Immediate user feedback, interactivity, and overall speed



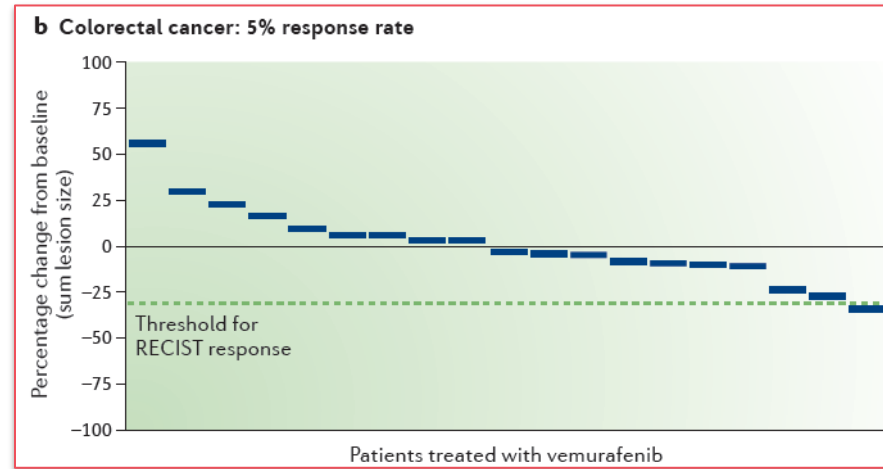
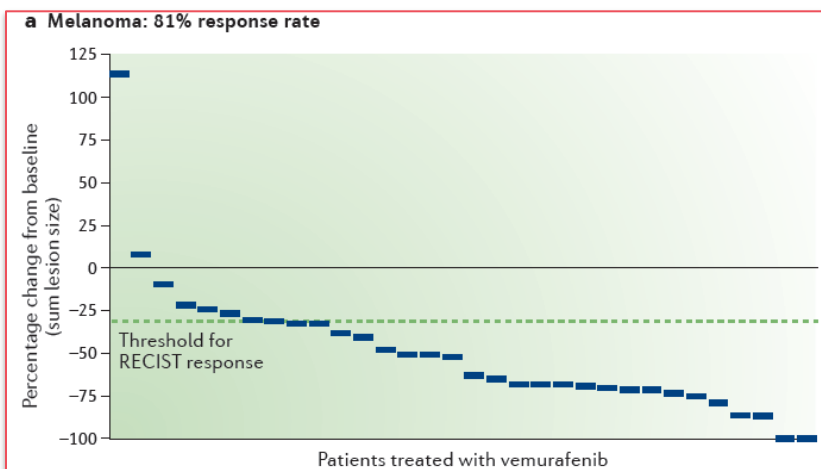
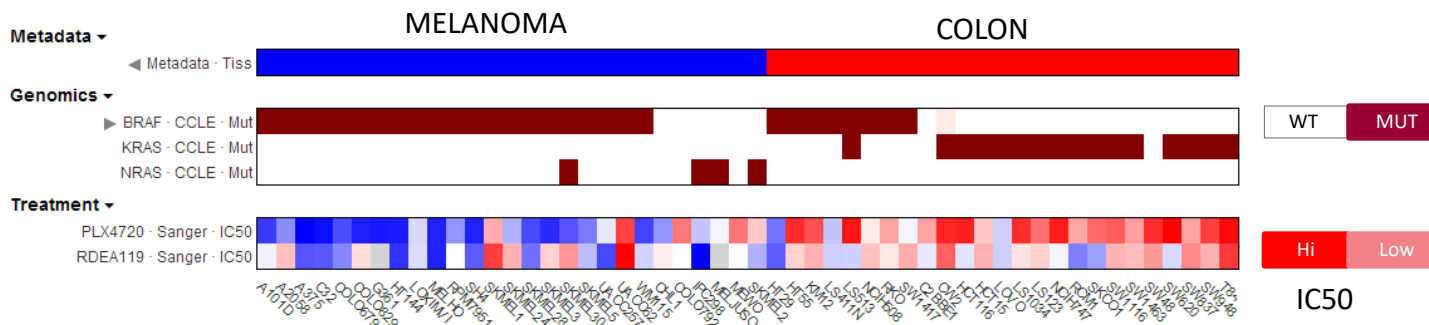
**Filtering**

Choose filters...

**Annotation**

- > Cohort
- ▼ Patient Sample Type All None
  - Additional - New Primary (1)
  - Additional Metastatic (1)
  - Metastatic (281)
  - Primary Tumor (7846)
  - Recurrent Tumor (41)
  - Solid Tissue Normal (1399)
- > Primary Site Of Disease
- ▼ Sample Type All None
  - Cell Line (1642)
  - Patient (10015)

# Correlating Drug Response to Genetic Biomarkers



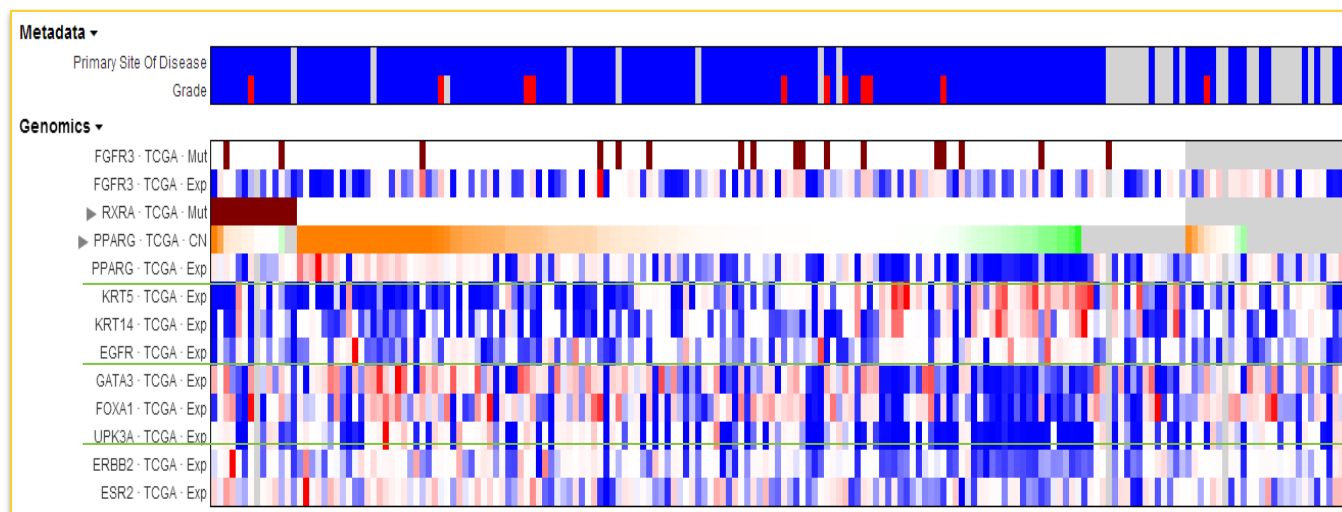
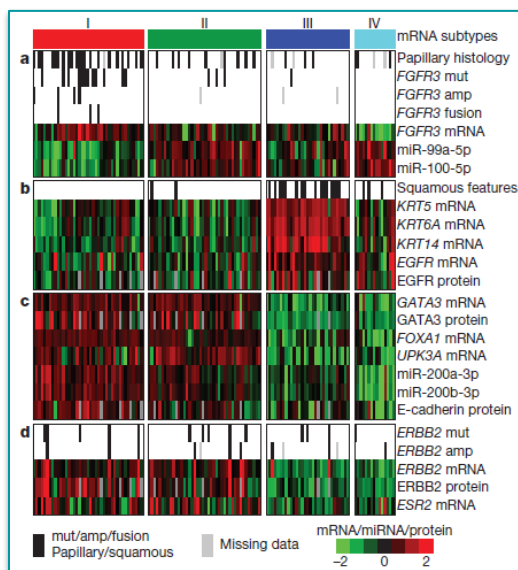
Bollag G et al. Nat Rev Drug Discov. 2012 Nov;11(11):873-86.

# Assessing targets under pathway or disease context



TCGA-BLCA WG paper  
-classify BLCA into 4 molecular classes

TCGA-TiP  
align RXRA mut/PPARG amplified samples into BLCA classes



Class I & II

Class III&IV

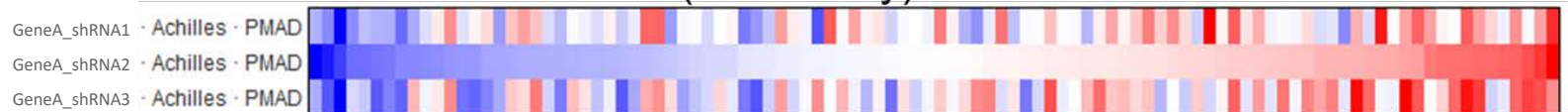


# Assessing synthetic lethal targets using functional genomics

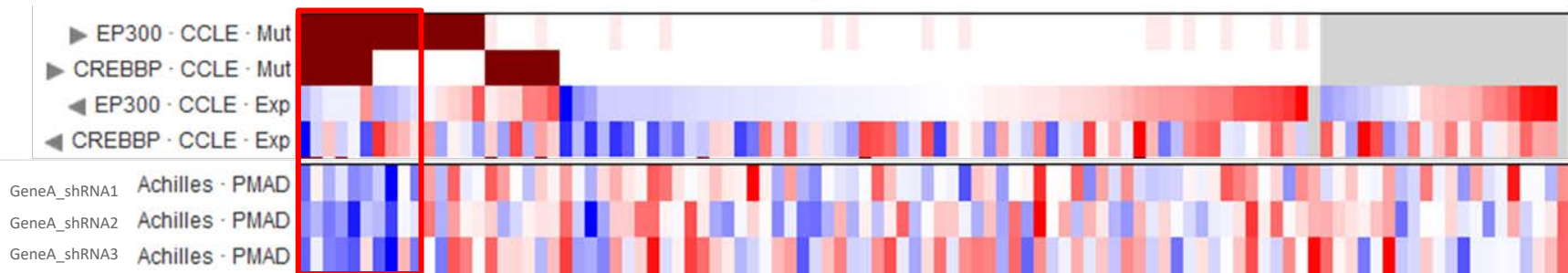
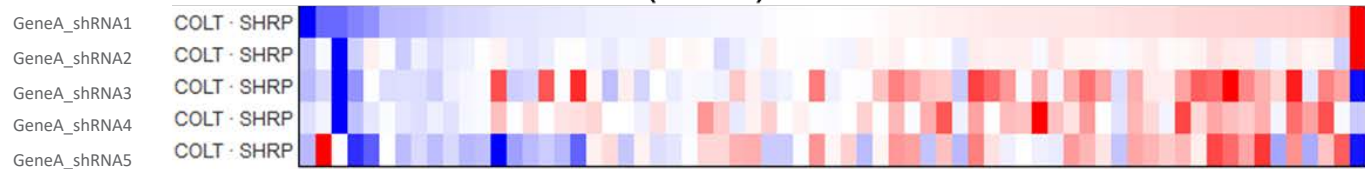


## EP300/CREBBP-Mutated Cancer Cells Are Selectively Sensitive to Gene X shRNAs

### Achilles (microarray)



### COLT (NGS)



On-target analysis confirmed by ATARiS (Shao DD, et al. Genome research, 23, 665–78)

# Gene Report – Deliver the Data as A Biologist see it



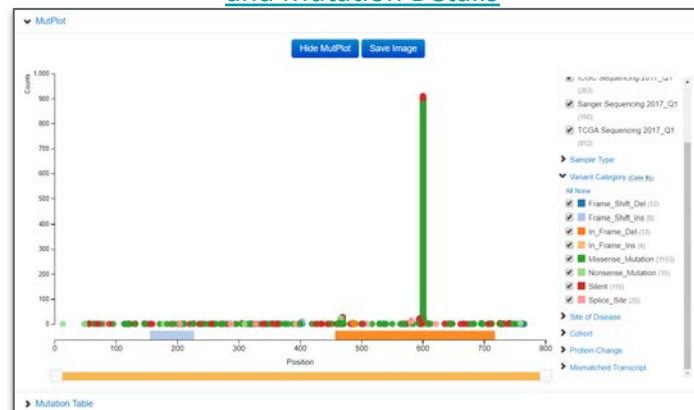
## Mutation – Visualization of types of mutations, MutSig Significance, and Mutation Details

H3 TIP Help Hello, Guest Login

### BRAF Gene Report

Showing best match for "Braf", other possible matches: BRAFP1, HMG20B

**Entrez Id** 673  
**Cosmic** BRAF  
**Full Name** B-Raf proto-oncogene, serine/threonine kinase  
**Aliases** RAFB1, NS7, BRAF1, B-RAF1, B-RAF  
**Type** Protein-coding  
**Map Location** 7q34  
**Summary** This gene encodes a protein belonging to the raf/mil family of serine/threonine protein kinases. This protein plays a role in regulating the MAP kinase/ERKs signaling pathway, which affects cell division, differentiation, and secretion. Mutations in this gene are associated with cardiofaciocutaneous syndrome, a disease characterized by heart defects, mental retardation and a distinctive facial appearance. Mutations in this gene have also been associated with various cancers, including non-Hodgkin lymphoma, colorectal cancer, malignant melanoma, thyroid carcinoma, non-small cell lung carcinoma, and adenocarcinoma of lung. A pseudogene, which is located on chromosome X, has been identified for this gene.  
**Notes** B-Raf Proto-oncogene, Serine/threonine Kinase



## Summary of Genomics Aberrations



MutSig Table

| Gene | Analysis               | Cohort          | q-Value     | q-Value Rank | # Mutations | # Patients |
|------|------------------------|-----------------|-------------|--------------|-------------|------------|
| BRAF | TCGA MutSig 2016_01_28 | THCA (TCGA)     | 3.6505e-13  | 1            | 291         | 291        |
| BRAF | TCGA MutSig 2016_01_28 | LUAD (TCGA)     | 2.365122e-8 | 11           | 43          | 39         |
| BRAF | TCGA MutSig 2016_01_28 | GBMLGG (TCGA)   | 0.00155595  | 34           | 12          | 11         |
| BRAF | TCGA MutSig 2016_01_28 | GBM (TCGA)      | 0.002172733 | 15           | 6           | 6          |
| BRAF | TCGA MutSig 2016_01_28 | COADREAD (TCGA) | 0.003816366 | 353          | 125         | 109        |
| BRAF | TCGA MutSig 2016_01_28 | COAD (TCGA)     | 0.004369344 | 331          | 111         | 96         |
| BRAF | TCGA MutSig 2016_01_28 | SKCM (TCGA)     | 0.04487881  | 34           | 169         | 145        |
| BRAF | TCGA MutSig 2016_01_28 | KIRP (TCGA)     | 0.06382661  | 16           | 4           | 4          |
| BRAF | TCGA MutSig 2016_01_28 | READ (TCGA)     | 0.08037777  | 27           | 14          | 13         |
| BRAF | TCGA MutSig 2016_01_28 | STAD (TCGA)     | 0.1133041   | 928          | 18          | 18         |

| Gene | Sample Type | Experiment              | Sample Name | Disease  | Cohort | Variant Category  | AA Position | Protein Change | Genome Change        | Alt. Allele Frequency | Total Depth |
|------|-------------|-------------------------|-------------|----------|--------|-------------------|-------------|----------------|----------------------|-----------------------|-------------|
| BRAF | Cell Line   | CCLE Sequencing 2017_G1 | 22RV1       | prostate |        | 3'UTR             | NA          | NA             | g chr7:140433295delA | 0.344                 | 32          |
| BRAF | Cell Line   | CCLE Sequencing 2017_G1 | 22RV1       | prostate |        | Missense_Mutation | 597         | p.L597R        | g chr7:140453145A>C  | 0.327                 | 787         |
| BRAF | Cell Line   | CCLE Sequencing 2017_G1 | 8305C       | thyroid  |        | Missense_Mutation | 600         | p.V600E        | g chr7:140453136A>T  | 0.429                 | 156         |
| BRAF | Cell Line   | CCLE Sequencing 2017_G1 | 8505C       | thyroid  |        | Missense_Mutation | 600         | p.V600E        | g chr7:140453136A>T  | 0.929                 | 42          |
| BRAF | Cell Line   | CCLE Sequencing 2017_G1 | A101D       | skin     |        | Missense_Mutation | 600         | p.V600E        | g chr7:140453136A>T  | 0.551                 | 256         |
| BRAF | Cell Line   | CCLE Sequencing 2017_G1 | A2058       | skin     |        | Missense_Mutation | 600         | p.V600E        | g chr7:140453136A>T  | 0.281                 | 32          |



# Business Impacts of TiP

- Lower the barrier of accessing the right data for all scientists
- Free up computational biologists to work on data analysis rather than data munging/wrangling
- Scientific discussion based on **live** interaction/iteration with data
- All of the above lead to significantly shorter cycle of initial hypothesis generation and testing



# Partnership is essential in our business

1. Internal resource investment should focus on core business
  - Data scientists enable data-driven drug discovery and development
  - Be the conductor!!
2. Strategic partners for infrastructure, operation, supporting scientific functions
  - Scale, resource to maintain stable operation
  - Core business investment to stay state-of-art
  - Network of customers, collaborators to extend scientific network and bring new science or best practice back to customers
3. Reward for internal data scientists
  - Focus on science not operation
  - Networks for collaborators
  - Opportunity to lead/manage external resources





# Benefits of Partnership with OmicSoft for TiP

- Omicsoft serves as our external genomics data processing, standardization partner primarily via Oncoland

The image shows two screenshots of genomics data portals. On the left is 'The Cancer Genome Atlas Data Portal' with a search interface for TCGA data. On the right is the 'Cancer Genomics Hub' website, which provides information about its secure repository for cancer genome data.

Broad Firehose  
UCSC  
Wash U  
MSKCC  
....  
(All GDACs)

Example:  
TCGA

No

Omicsoft Oncoland data subscription

TiP

- Fraction of the internal investment, regular updates, continued growth of new datasets
- H3 data scientists focus on
  - Internal data processing/standardization
  - Science part of the analysis – “transform”

# Summary and Lessons Learned



- Goal (Data Driven Drug Discovery and Development) determines means:
  - Focusing on knowledge discovery - enable all scientists
  - Focusing on end goal with fit-for-purpose technologies
- TiP – a Translational Informatics Platform enabling non-informatic scientists to examine omics and functional screening data, and generate hypothesis
- Engage partnership (such as Omicsoft) to help us:
  - Keeping pace with new data and data types while allowing informaticians focus more on scientific questions
  - Scalable, flexible resources, stable operation

# Acknowledgement



- Alex Ramos
- Jacob Feala
- Chia-Ling Huang
- Stephen Kottman



- Luis LeBolo



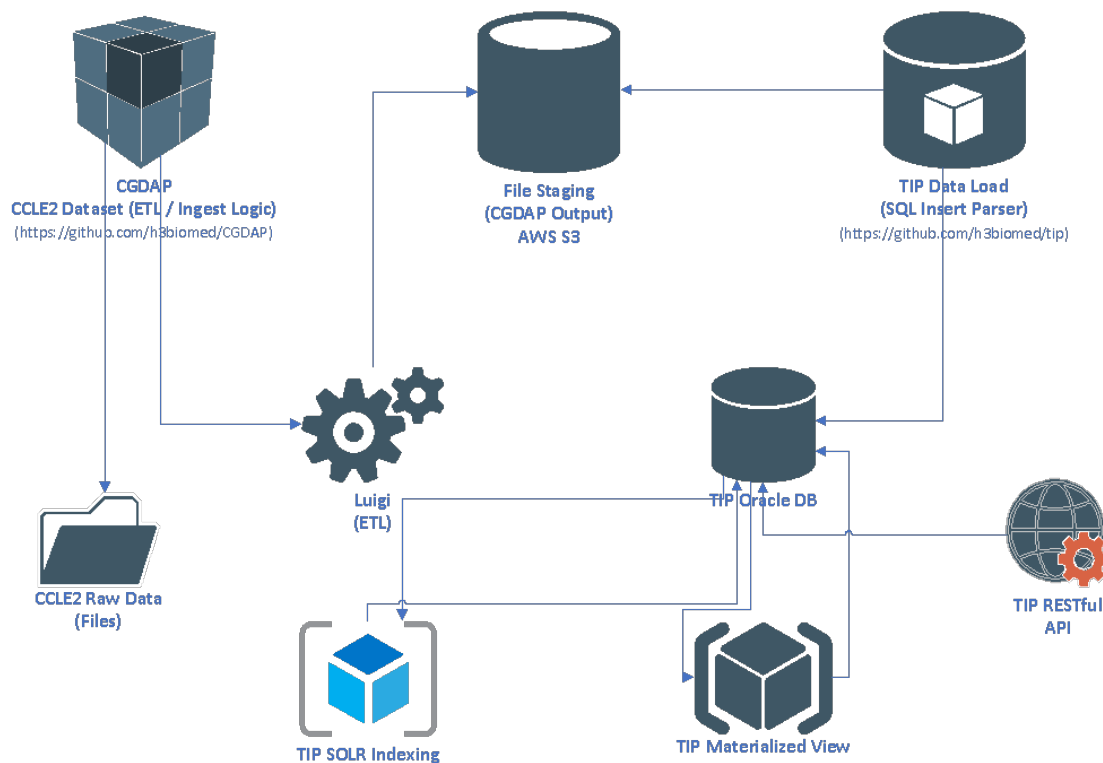
- Matt Newman
- Scott Magin
- Support team

# Backups





# TiP software engineering architecture



Andrew Brown, Arrayo

# Story Line



- H3
- H3 Data Science:
  - emphasize our focus on data democratization – data access and usability designed with users (biologists) and scientific questions in mind as our guiding principle.
  - Emphasize our focus on translate data into actions – therefore Our Philosophy of collaborating with strategic partners in most other areas.
- TiP
  - Its position in H3 genomics eco-system
    - Data integration
    - Data reporting
    - Iterative Data query
  - How do we work with strategic partners?
    - Omicsoft land data – data munging
    - Spotfire to sophisticated data visualization and exploration
  - Benefit of using Omicsoft land data vs. internal procesed of data



# Partnership

1. Internal resource investment should focus on core business
  - Data scientists enable data-driven drug discovery and development
  - Be the conductor!!
2. Strategic partners for infrastructure, operation, supporting scientific functions
  - Scale, resource to maintain stable operation
  - Core business investment to stay state-of-art
  - Network of customers, collaborators to extend scientific network and bring new science or best practice back to customers
3. Reward for internal data scientists
  - Focus on science not operation
  - Networks for collaborators
  - Opportunity to lead/manage external teams

