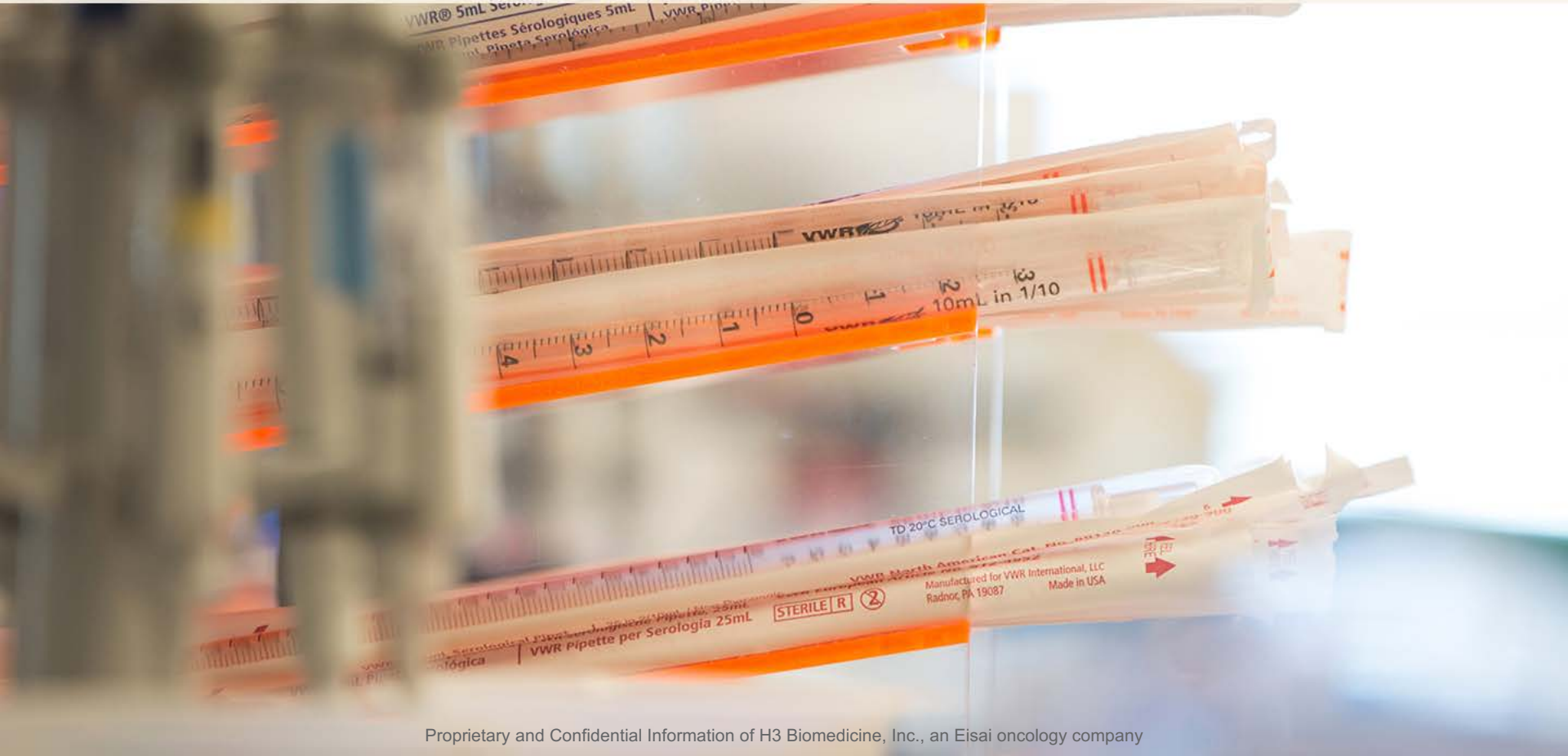
TiP – A Genomics Platform Facilitated by OmicSoft Curated Land Data

Omicsofr 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

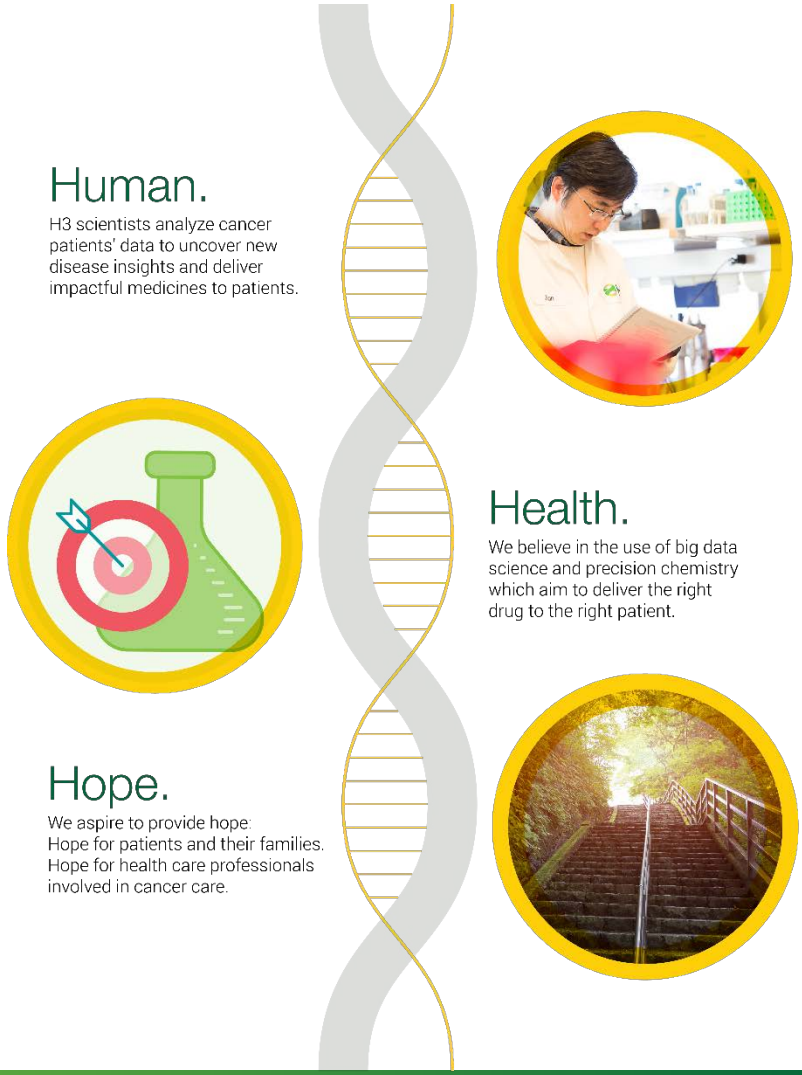
FOUNDED | 2011

FOCUS | Novel and highly targeted oncology treatments

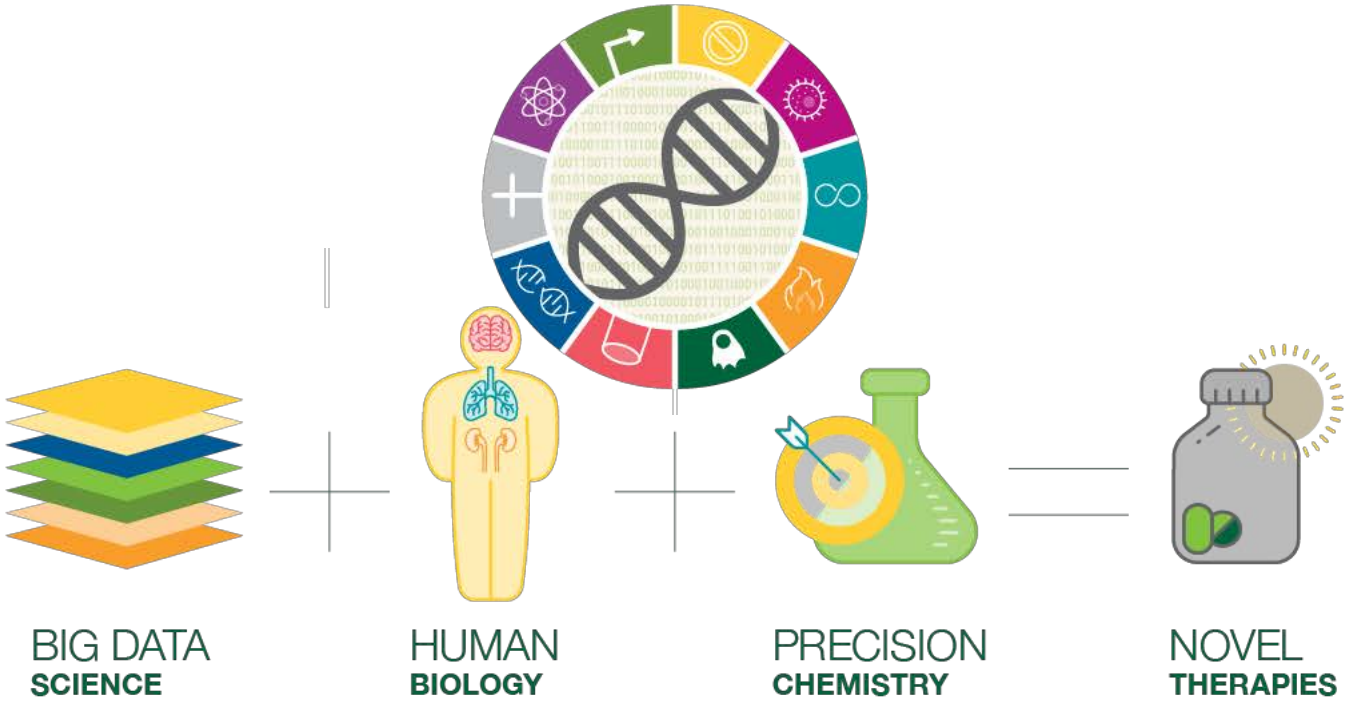
HEADQUARTERED | Cambridge, Massachusetts

ORGANIZATION | Private

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



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

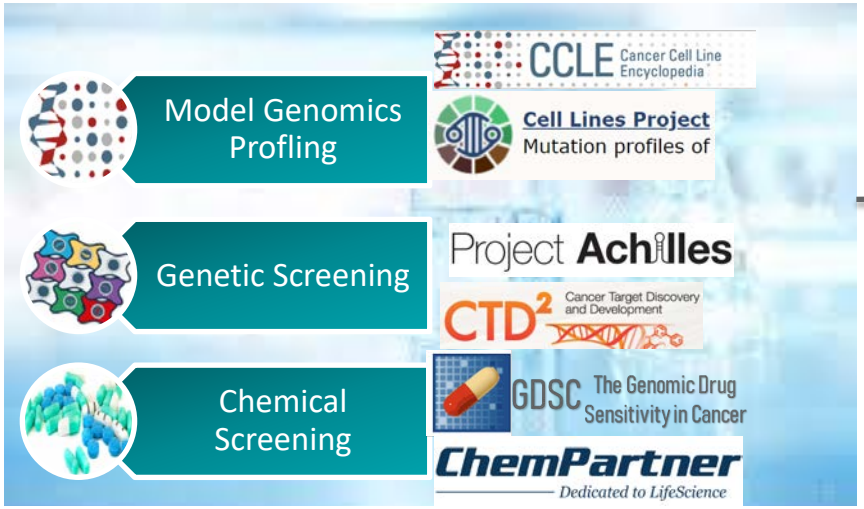
H3 Data Science: Aspire to be the most impact focused organization

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

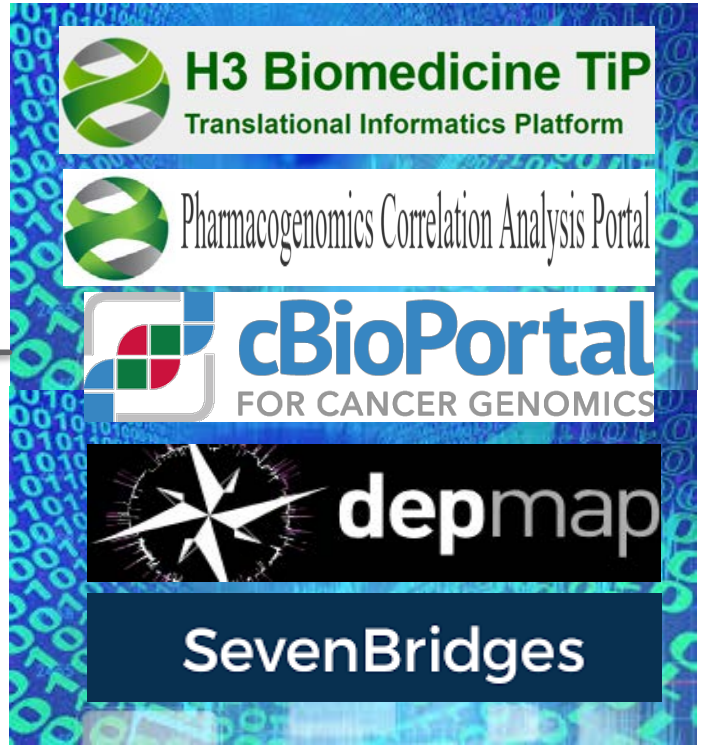
Cancer Genomics, Target Biology and Translation	Therapeutic Modalities	Clinical Development, Biomarker and Diagnostics, Translation
Clinical Genomics	HTS – Ultra large library screening	Biomarkers & Diagnostics
Preclinical Models	LI/LO – In silico drug design & Compound profiling (ADME, Tox)	RWE informed/transformed trials
Functional Genomics		

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

Big Data, Deeping Mining – Human Genetics Driven Targets



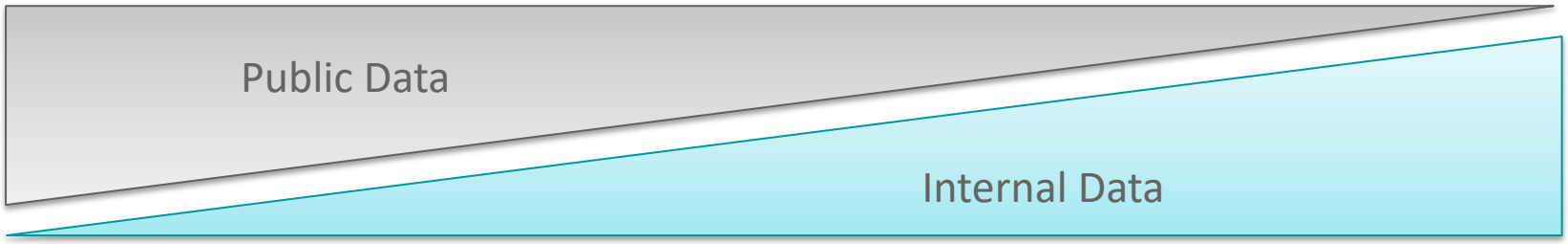
Deep Mining Platforms



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.

Target Identification/ Validation	Pre-clinical Drug Discovery	Translation	Clinical Trials
• Genomically defined targets • Cancer biology and pathways	• Cancer biology and pathways • Markers for screening or MOA	• Patient selection biomarker • PD biomarker • Disease indication expansion • Drug-Drug combination	• Patient selection biomarker • PD biomarker • Exploratory biomarkers for efficacy, resistance



The Data Ecosystem

Generation → Processing → Management → Integration → Analysis Mining

Genomics Data

- External
- Internal

H3 Data Science+Partners

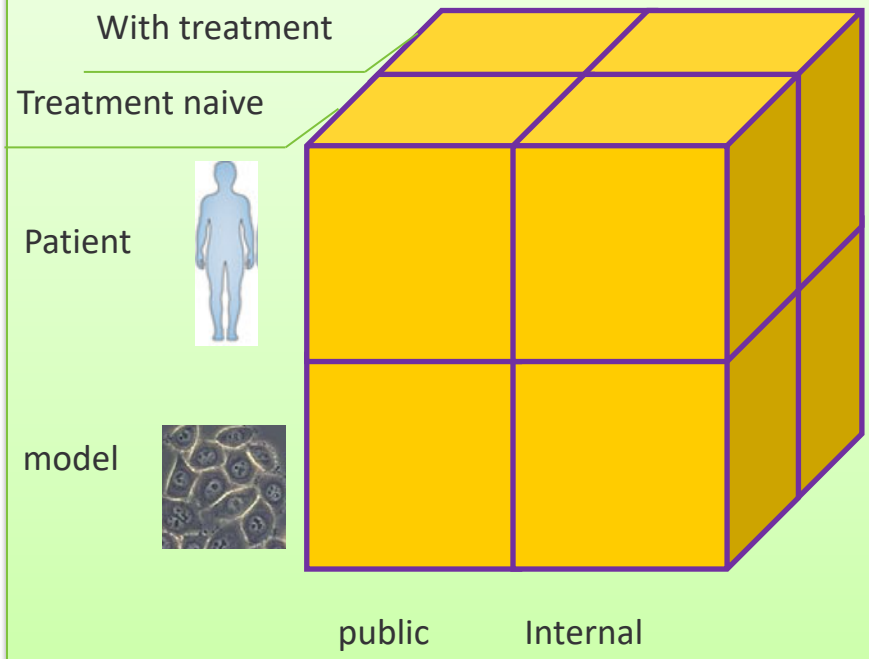
- Data Collection
- Data cleaning, standardization, re-annotation
- Data processing, analysis

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

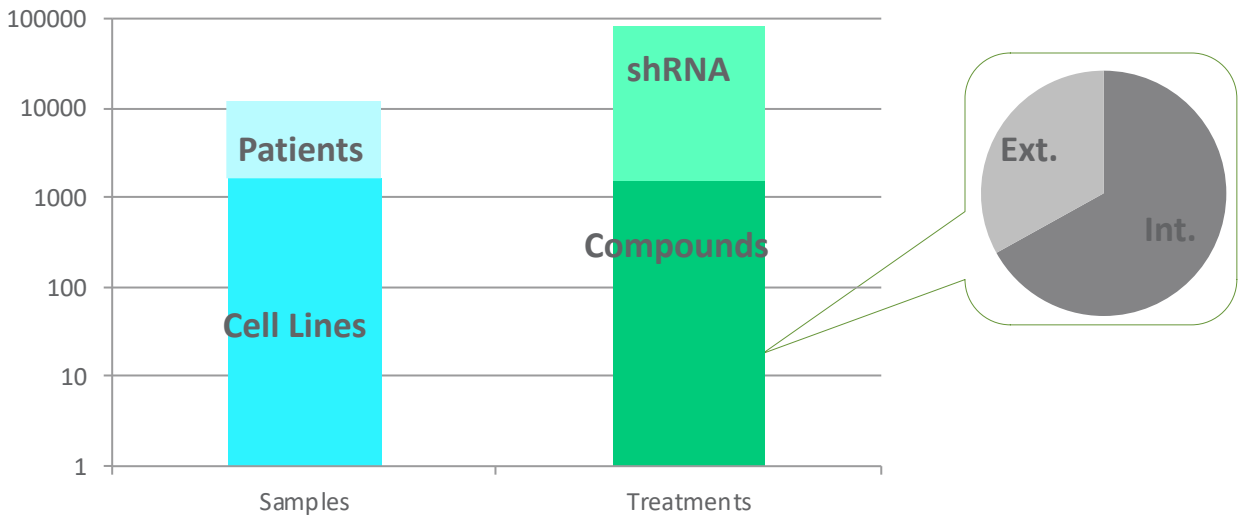
Focus on
- Cancer genomics
- Functional genomics
- Compound profiling

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

TiP – Data Overview - circa 2014

Datasets

CCLE	Achilles
Sanger	COLT OICR
CTD2	H3 Internal
GDSC	TCGA
NCI-60	



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

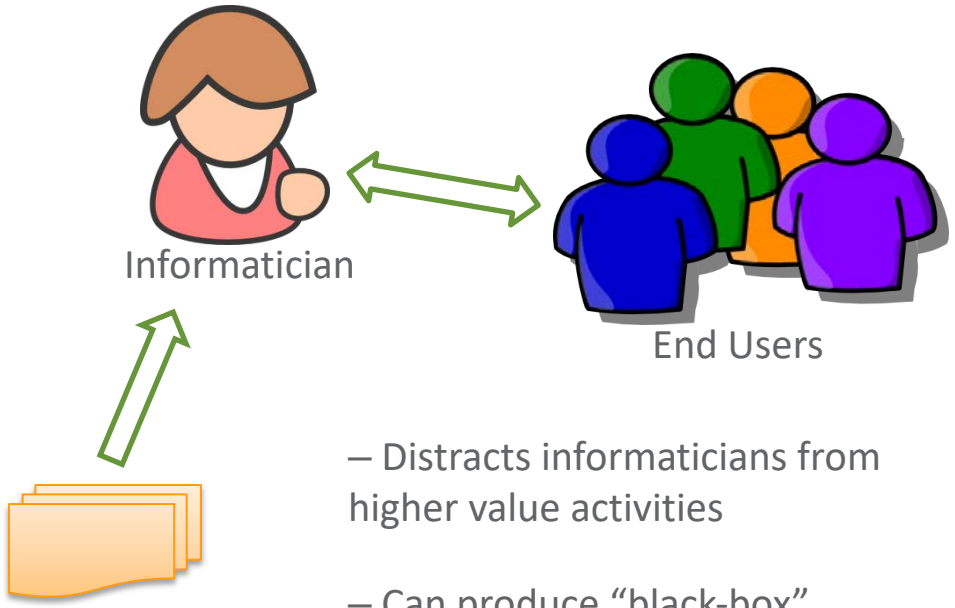
~80% of the data are from in Omicsoft Oncoland data subscription

- 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^2**)
- 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

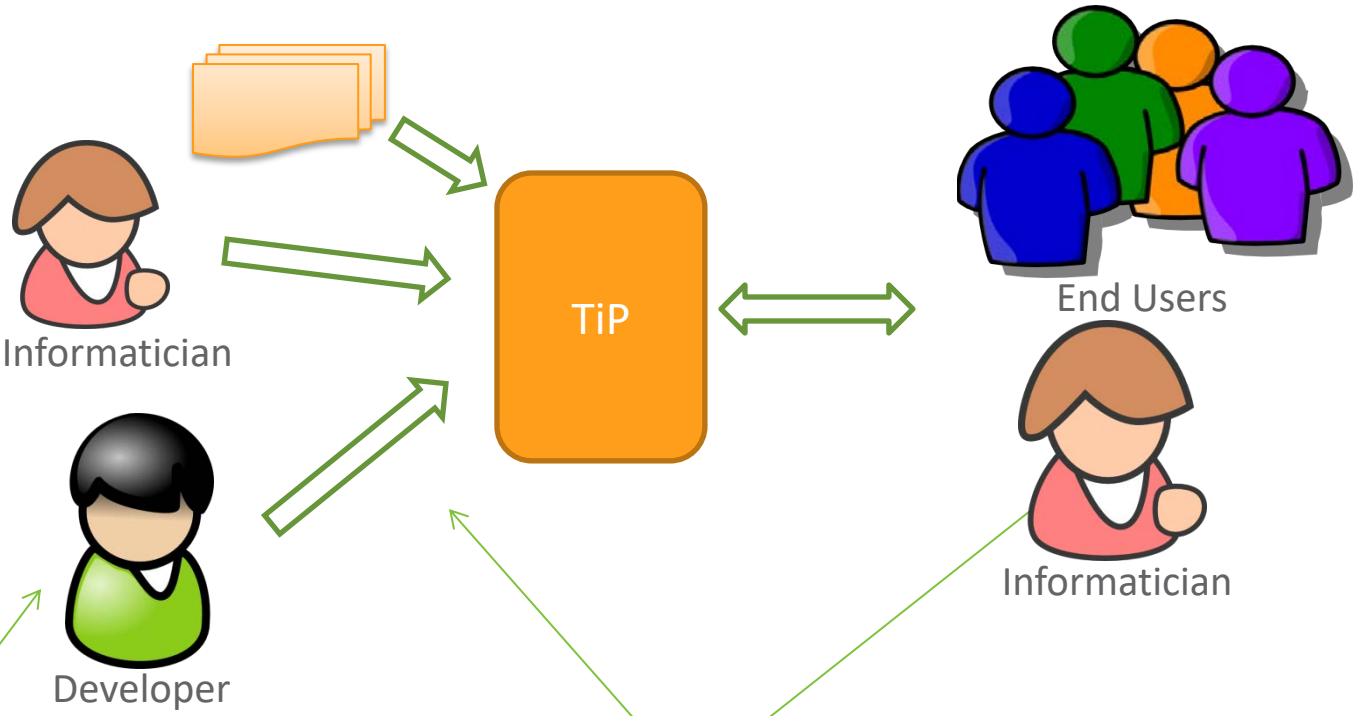
– Even best informatics group has a bandwidth limit

– 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

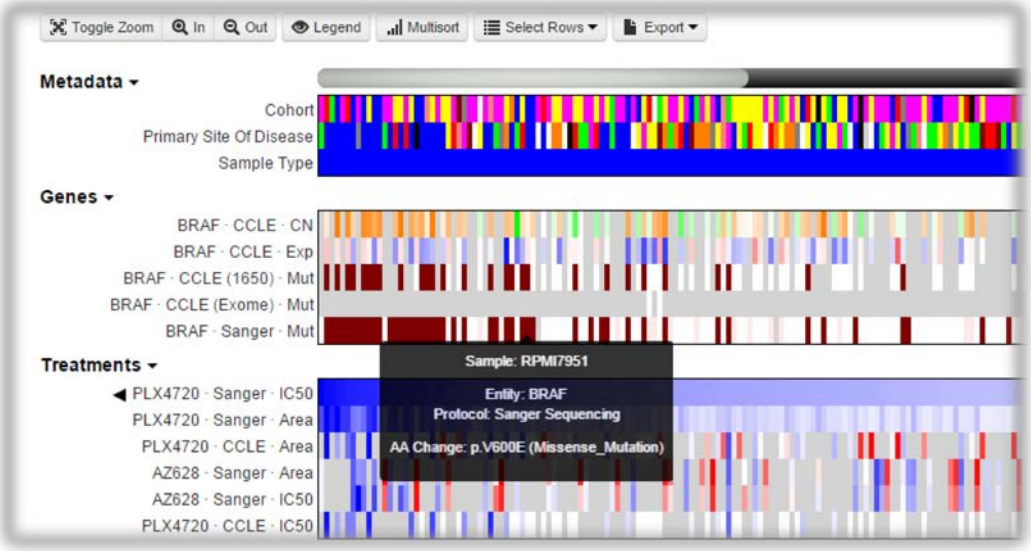
Iterative User Query Interface

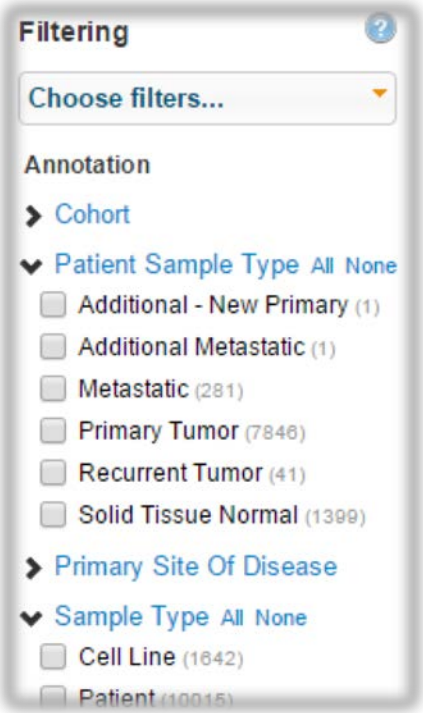
Heavy design focus on usability

- User Centered Design
- Story Boarding / Prototyping

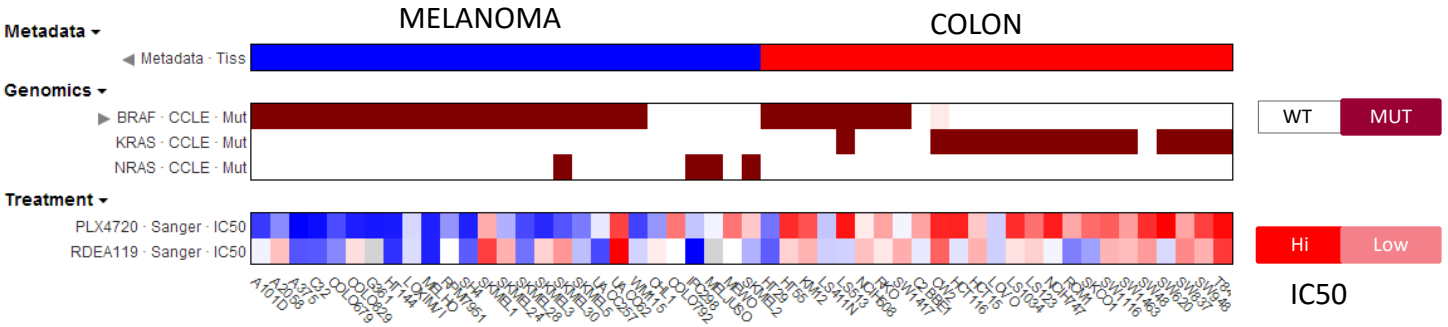
- Use of familiar concepts (Amazon facets, shopping cart)

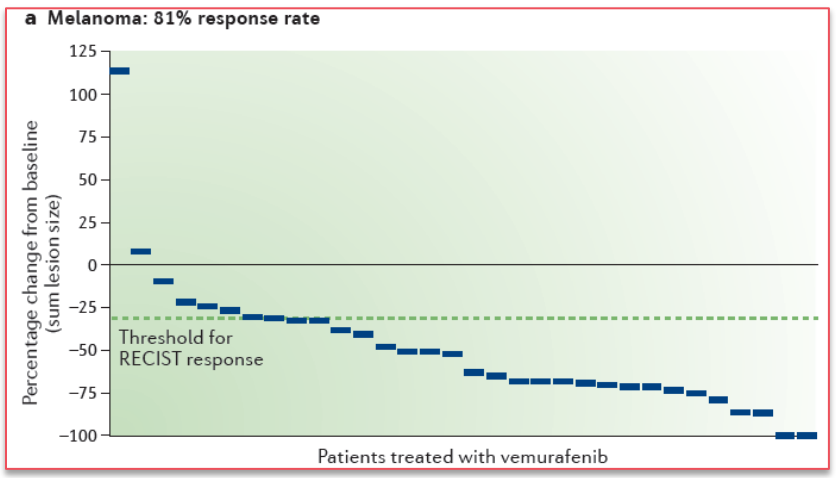
- Immediate user feedback, interactivity, and overall speed

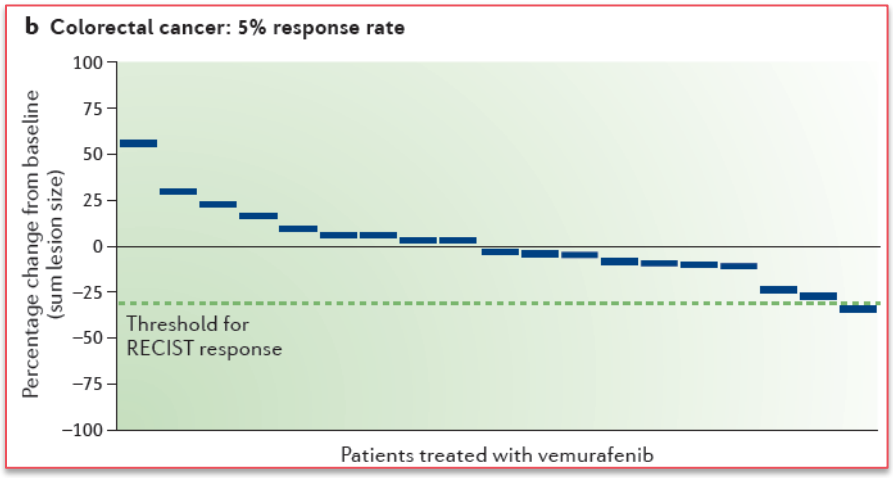




Correlating Drug Response to Genetic Biomarkers



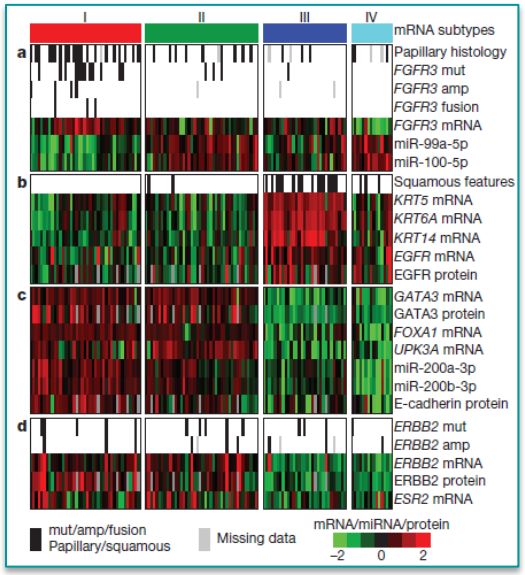




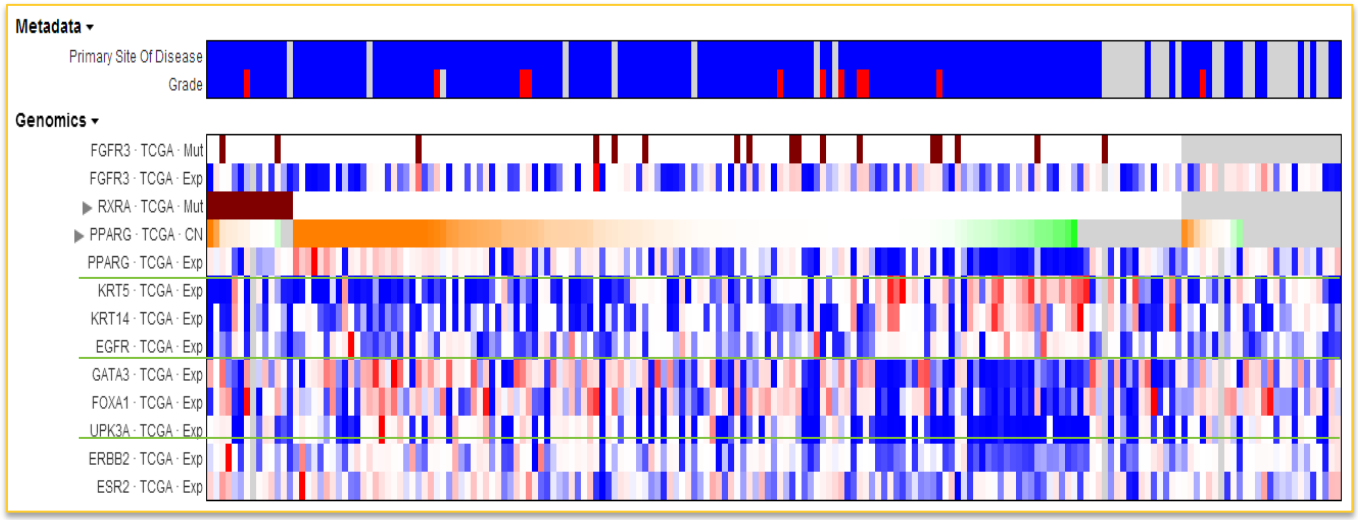
Bollag G et al. Nat Rev Drug Discov. 2012 Nov;11(11):873-86.

Assessing targets under pathway or disease context

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



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

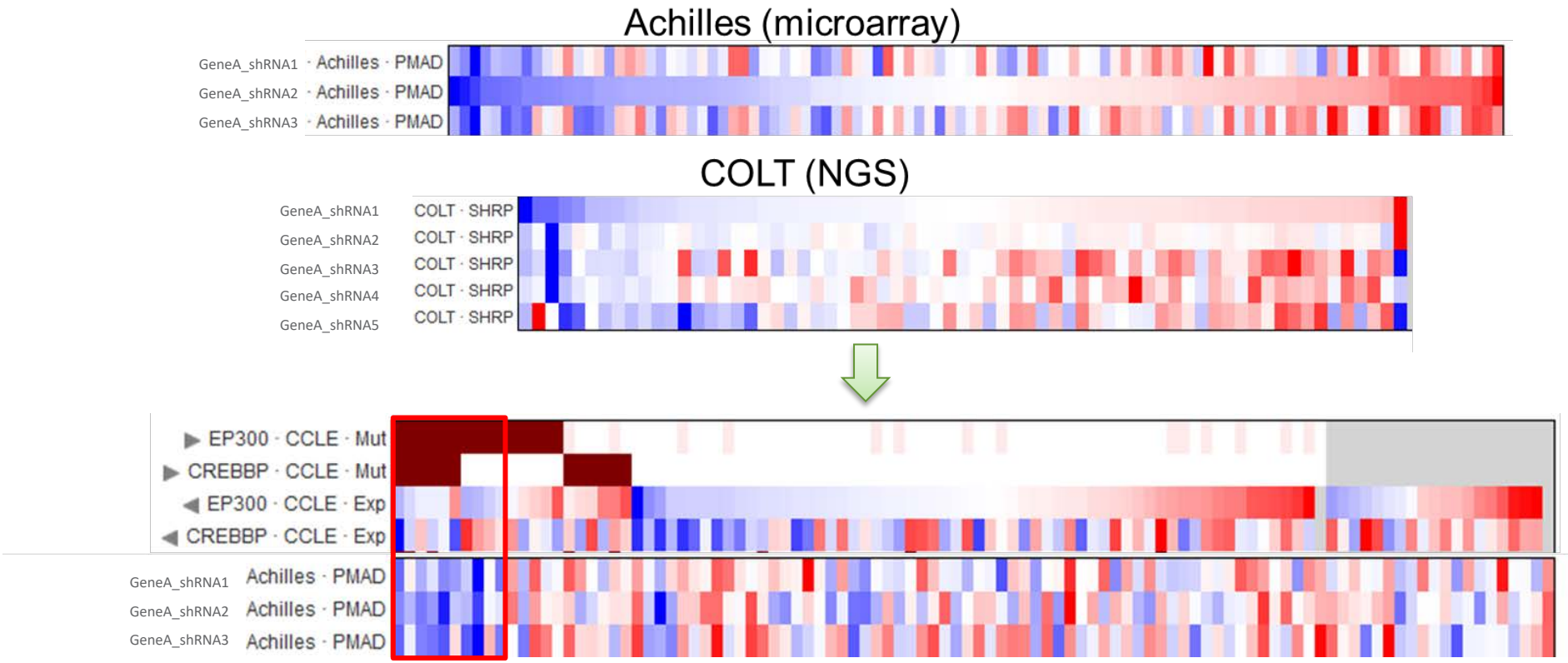


Class I & II

Class III&IV

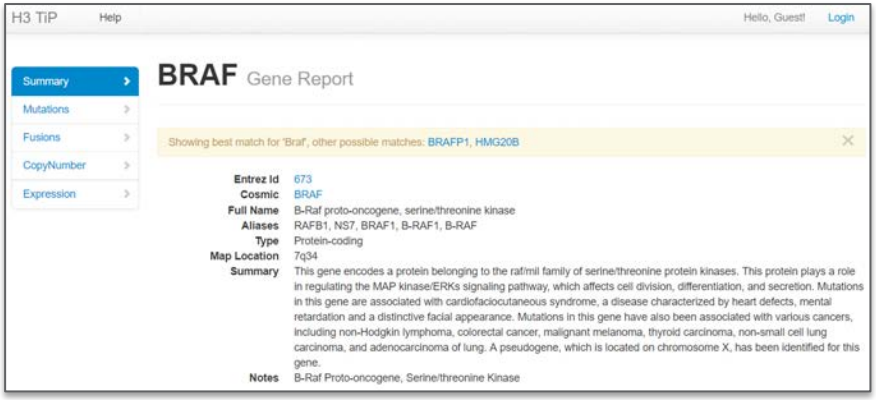
Assessing synthetic lethal targets using functional genomics

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

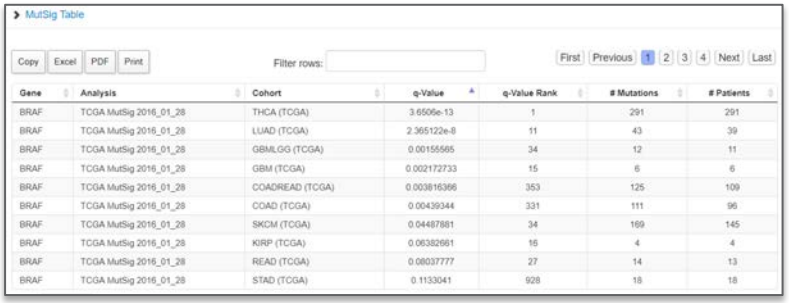


On-target analysis confirmed by ATARiS (Shao DD, et al. Genome research, 23, 665–78)

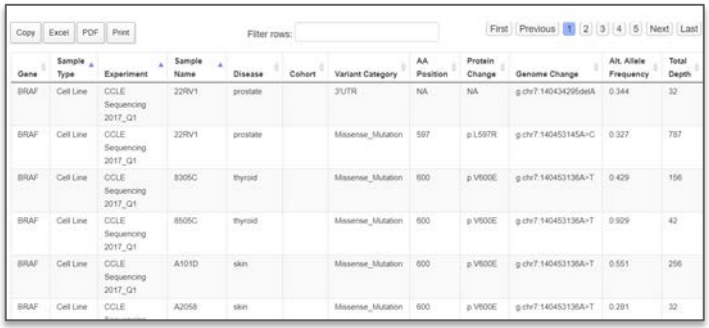
Gene Report – Deliver the Data as A Biologist see it



Mutation – Visualization of types of mutations, MutSig Signifcance, and Mutation Details



Summary of Genomics Aberrations



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

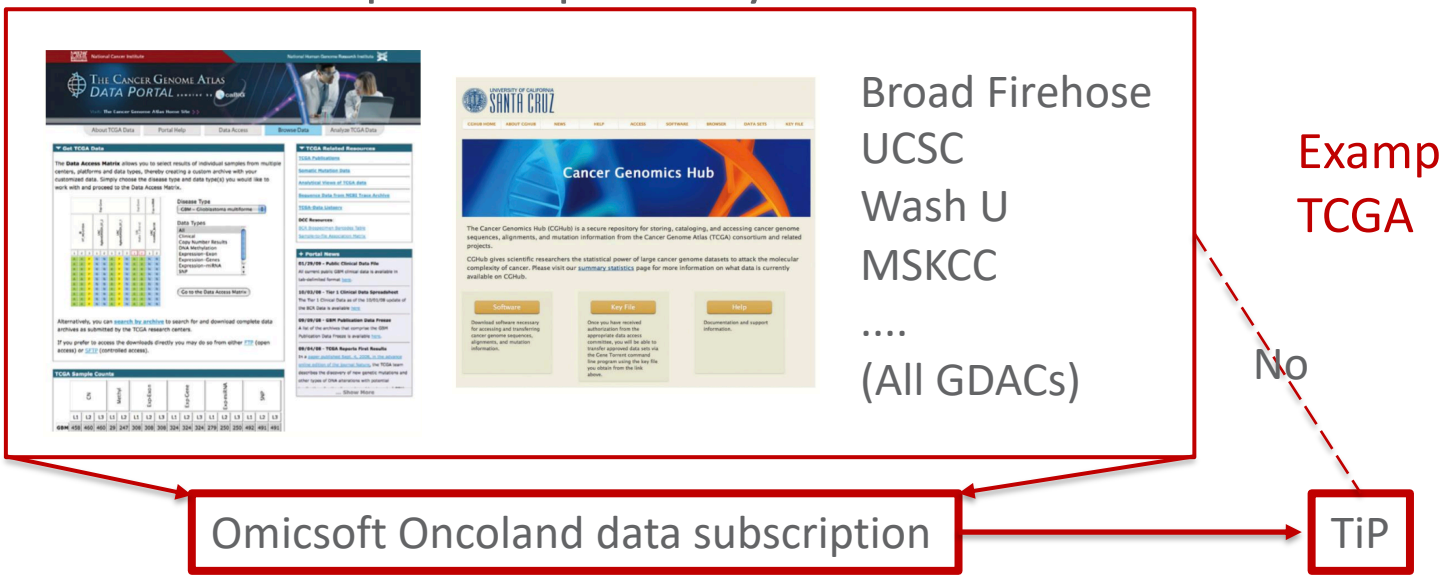
1. Internal resource investment should focus on core business
   - Data scientists enable data-driven drug discovery and development
   - Be the conductor!!

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



- 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

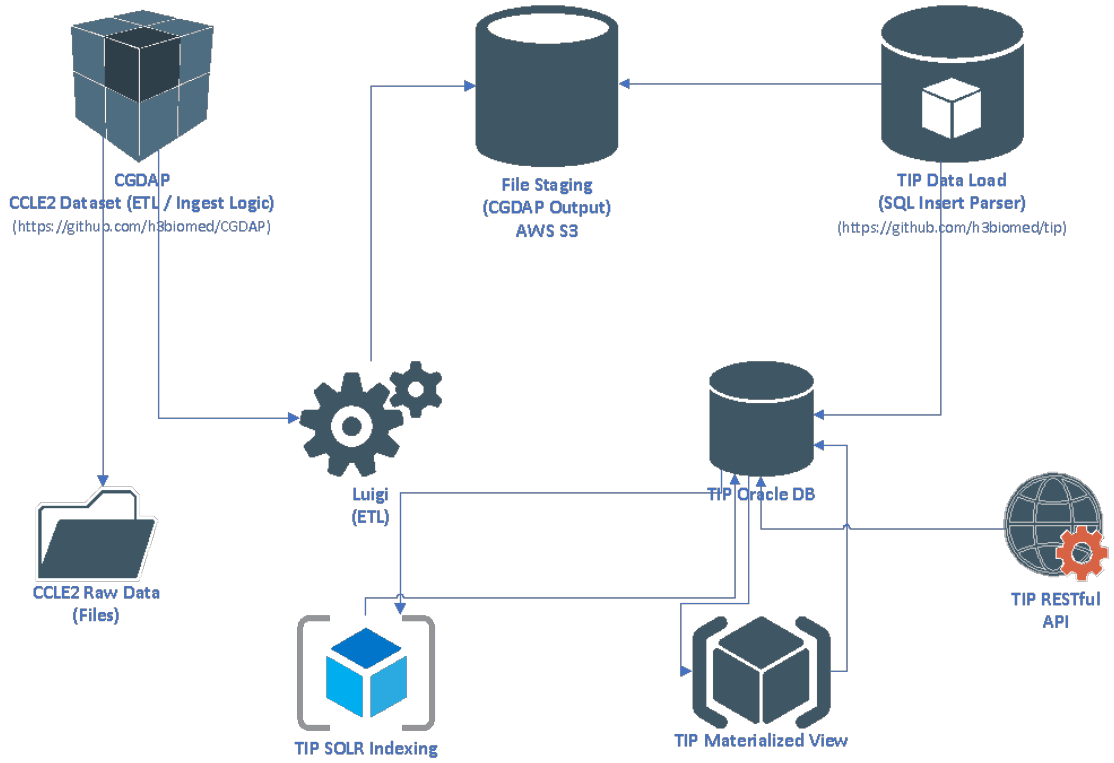
- Alex Ramos
- Jacob Feala
- Chia-Ling Huang
- Stephen Kottman

- 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