



The Biomedical Evidence Graph

ITCR Update for Face-to-Face
Santa Cruz, CA *June 1, 2017*

Josh Stuart (UCSC), Kyle Ellrott (OHSU)

BMEG Aims

- **Aim 1** - Develop **automated cancer evidence discovery pipelines**. Formatted for graph analysis.
- **Aim 2** - Determine **genotype-phenotype recognizers**. ID top methods with public challenges.
- **Aim 3** - **Evidence graph and network analysis** (e.g. connectivity, co-occurrence, phenotype ID)
- **Aim 4** - Online **use case apps** (e.g. recommenders, infer targets)

Challenges of Interdisciplinary Omics Analyses

- Complex pipelines
 - And many method choices at every step
 - Methods are computationally expensive
 - Every method different requirements
- Diverse sources of data needed
 - Scattered
 - Too large to move around
- Too complex to understand
 - Need interpretation
 - Need intuitive overviews
 - Need integration for “big picture”

Challenges of Interdisciplinary Omics Analyses

- **Complex pipelines**
 - And many method choices at every step
 - Methods are computationally expensive
 - Every method different requirements

- **Diverse sources of data needed**
 - Scattered
 - Too large to move around

- **Too complex to understand**
 - Need interpretation
 - Need intuitive overviews
 - Need integration for “big picture”

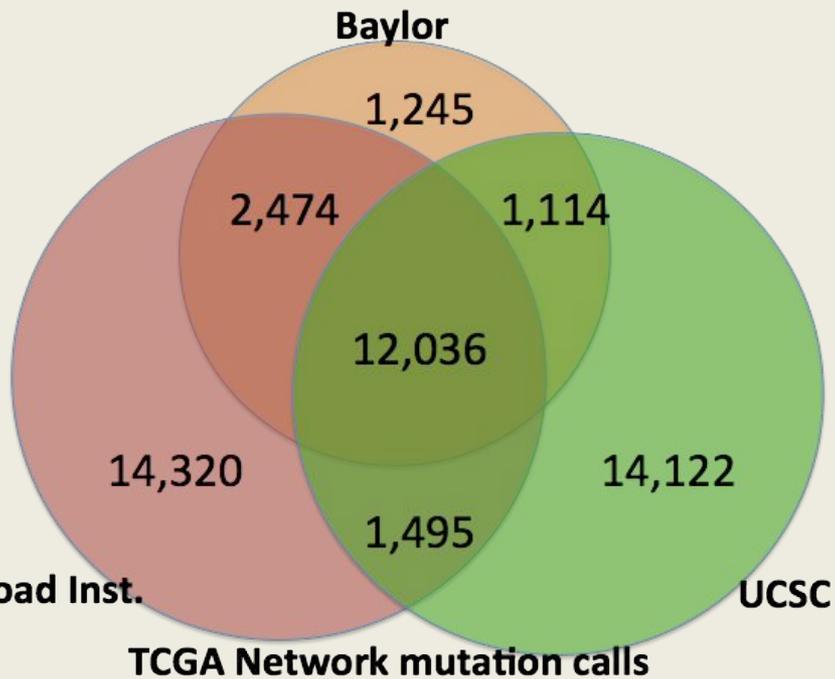
Challenges of Integrating Primary Analyses

Biomedical Evidence Graph (BMEG) Aims

- Complex pipelines
 - And many method choices at every step
 - Methods are computationally expensive
 - Every method different requirements
- Diverse sources of data needed
 - Scattered
 - Too large to move around
- Organize crowdsourced challenges to identify top methods
 - DREAM projects, living benchmarks
- Cloud ready methods for evidence gathering
 - Alignments and somatic variant detectors

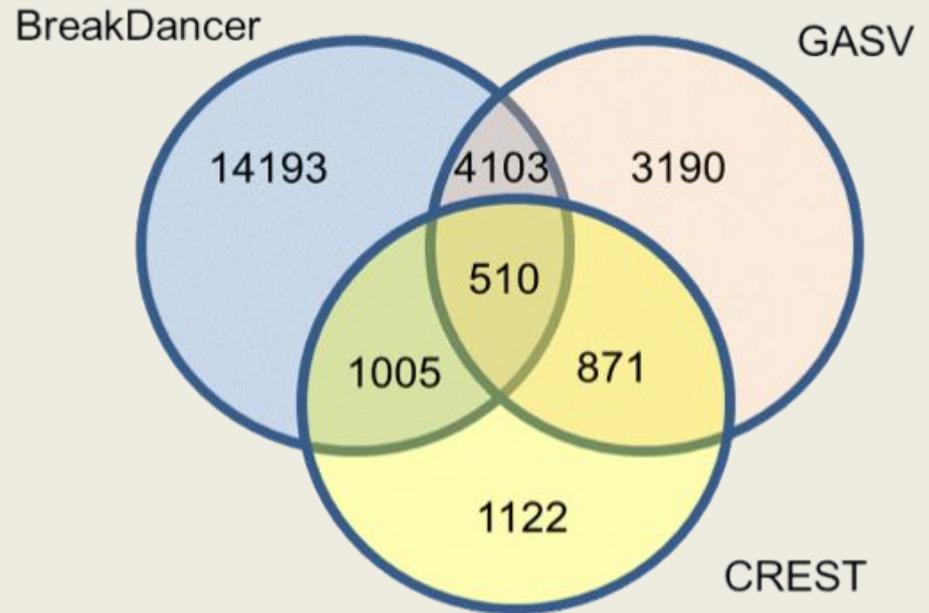
Mutation Callers Give Different Answers

SNVs



TCGA Network mutation calls

Renal Clear Carcinoma:
Somatic calls from 297 Samples



Singer Ma (UCSC)

What is the best method?

Have a Bake-Off.



Best Cookies



What is the best method?

Have a Bake-Off.



Best Cookies

and

Best Cooking Components



Somatic Mutation Evidence

1) Identifying Best Methods



The banner features a central graphic of five vertical DNA strands with red and blue segments, connected by a network of lines. To the right, the text reads 'ICGC-TCGA DREAM Somatic Mutation Calling Challenge – Tumour Heterogeneity & Evolution'. Below this, a row of logos includes: DREAM CHALLENGES (powered by Sage Bioinformatics), UNIVERSITY OF TORONTO, IBM, ICGC, UCSC, NSERC CRSNG, OICR (Oncology Institute of Cancer Research), Sage, Prostate Cancer Canada, GenomeCanada, The Cancer Genome Atlas, Understanding genomics to improve cancer care, Galaxy, and OHSU.

- ~400 registrants
- ~3000 entries on 14 genomes
- Completed 2014
- https://synapse.org/DREAM_SMC
- New insights gleaned (*Nat Methods* 2015).



Somatic Mutation Evidence

2) Method Documentation

3) Run on Large Scale Data

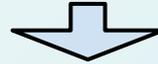
TCGA Unified Mutations (MC3)

- Need unified, comprehensive mutation calls for all TCGA samples for PanCanAtlas
- AWG calls vary from group to group
- Different calling techniques
- Different filtering methods
- Stopped work when papers published (incomplete data)
- Inconsistencies from tumor type to tumor type.



MC3 Process With A Uniform Pipeline

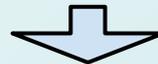
10K TCGA Exomes



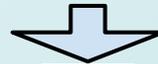
GATK Preprocessing
(DREAM winner)



Coherent BAM collection



Consistent Mutation Calling
across all TCGA exomes



MAF Creation



PanCanAtlas

Methods

(all eval w/ DREAM

& dockerized)

Broad - Mutect

Baylor - MuSE (1.0 RC)

WashU

VarScan 2.3.9

Somatic Sniper 1.0.5

Pindel 0.2.5b8

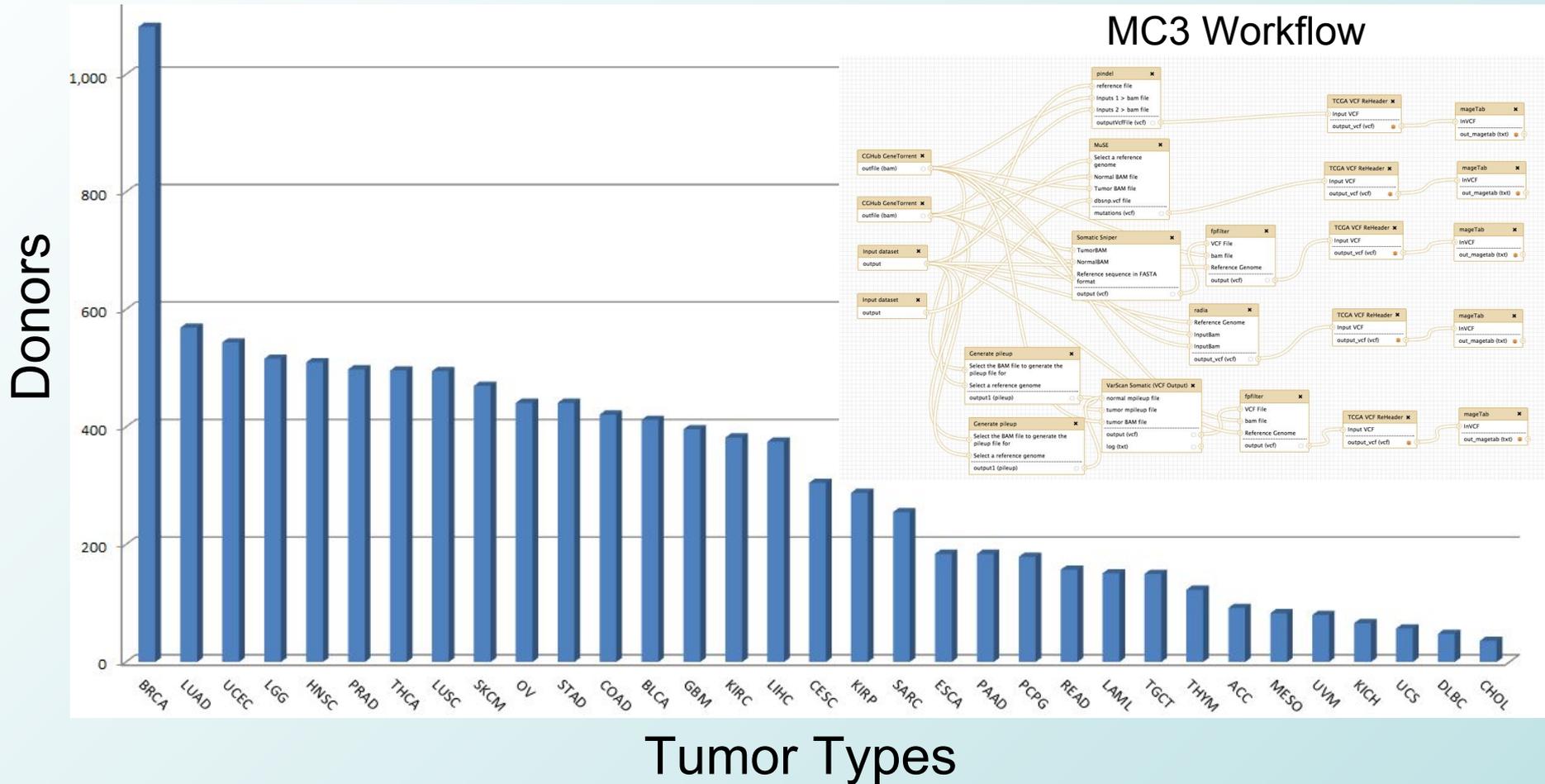
UCSC - Radia 1.1.5

MC3 Doubled Sample Coverage

<u>TYPE</u>	<u>TCGA AWGs</u>	<u>MC3</u>
Tumor Samples	5480	10176
Variant Calls	1,396,139	3,397,307
Methods	Varied by group	Consistent 7

Loaded into BMEG DB

MC3 Results, Posted in Summer, 2016

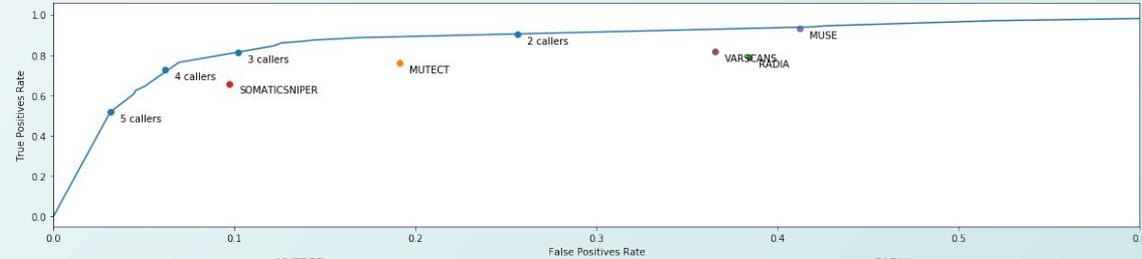


Donors: 10,485

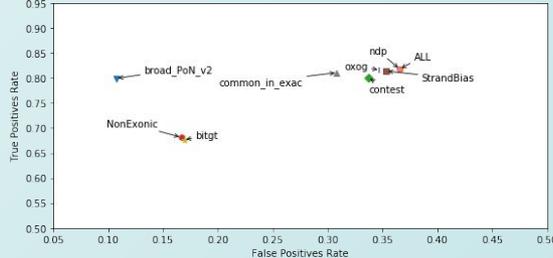
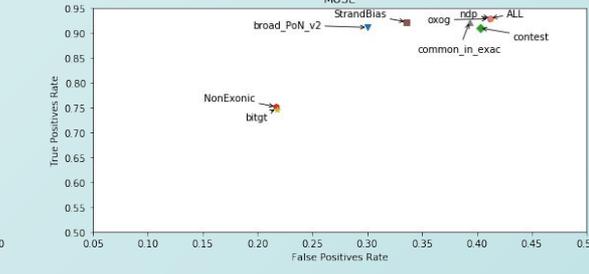
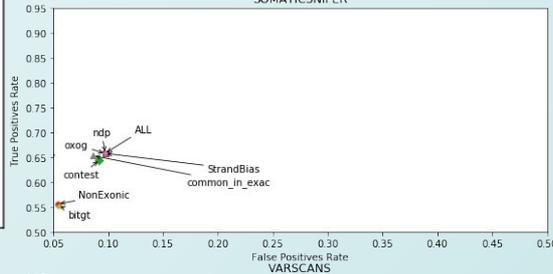
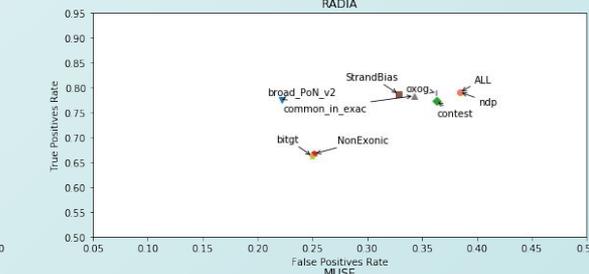
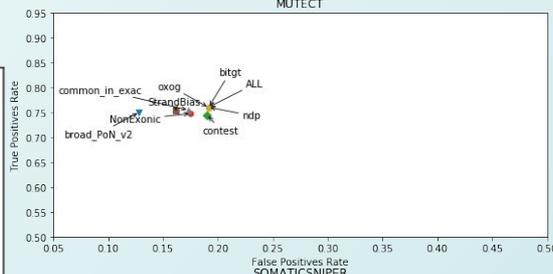
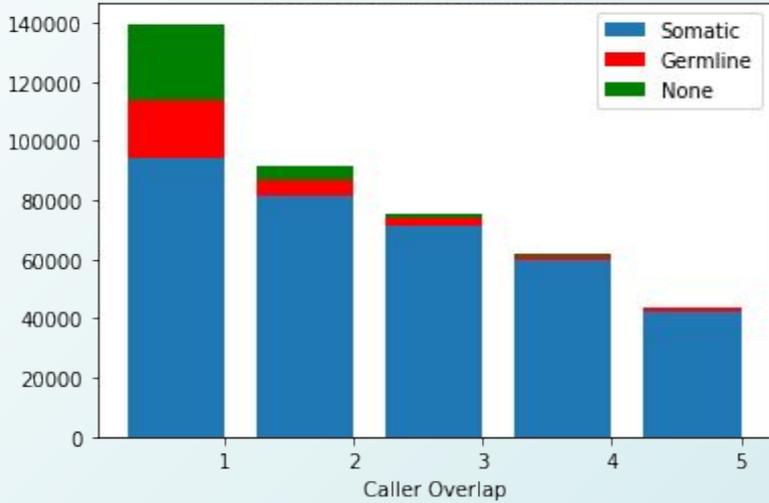
Open Access MAF file released

<https://www.synapse.org/#!/Synapse:syn5917256>

MC3 Analysis



Error Profile vs Caller Overlap



Automating Somatic Mutation calling for Ten Thousand Tumor Exomes. Submitted.

DREAM: From Competition to Science

1) DREAM Challenge Completion



2) Method Documentation



3) Mass Deployment



4) Use for real Science



DREAM: From Competition to Science

Getting Stuck Here. Why?

1) DREAM Challenge Completion



2) Method Documentation



for real
ce



DREAM: From Competition to Science

Getting Stuck Here. Why?

1) DREAM
Challenge
Completion



2) Method
Documentation



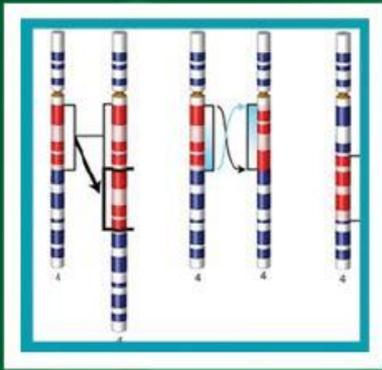
for real
ice



Rethink challenge incentives

- score code only when re-runnable.

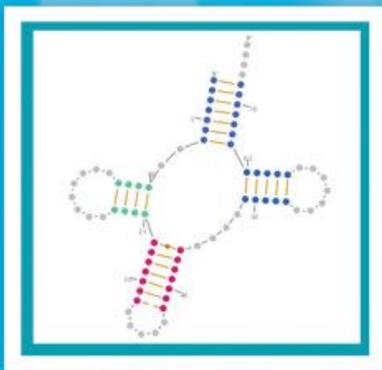
'Next Gen' DREAM Challenges Will Produce *Re-runnable Code*



ICGC-TCGA DREAM Somatic Mutation Calling Challenge – Tumour Heterogeneity & Evolution



<https://synapse.org/SMCHet>



ICGC-TCGA DREAM Somatic Mutation Calling Challenge – RNA



https://synapse.org/SMC_RNA

SMC-HET and SMC-RNA



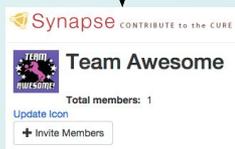
Global Alliance
for Genomics & Health



DREAM Participants



1 - Register at Synapse
https://www.synapse.org/SMC_RNA



2 - Go to cloud pilots to view training data and make containerized workflows

3 - Submit completed workflow back to Synapse



4 - DREAM organizers apply workflows to held out dataset



SMC-HET and SMC-RNA



Global Alliance
for Genomics & Health

Method only
Scored and Ranked
If ***WE CAN RUN IT***



4 - DREAM organizers
apply workflows to
held out dataset

3 - Submit
completed
workflow
back to
Synapse

2 - Go to cloud
pilots to view
training data
and make
containerized
workflows

DREAM Participants

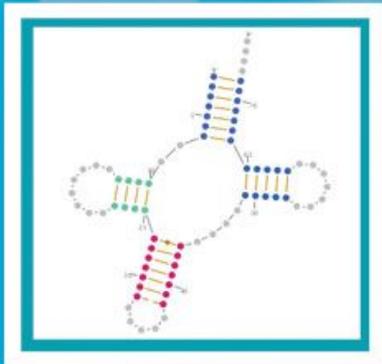
1 - Register at Synapse
https://www.synapse.org/SMC_RNA



DREAM Somatic Mutation RNA Challenge (SMC-RNA)

- Expression is a read-out of cellular pathways
- Examples of fusions as key oncogenic mechanism (BCR-ABL, TMPRSS2-ERG)
- Alternative splicing implicated in many tumor types
- Provide treatment clues

- Sub-challenge 1: Identify Isoforms
- Sub-challenge 2: Identify fusions



ICGC-TCGA DREAM Somatic Mutation Calling Challenge – RNA



The Cancer Genome Atlas



Understanding genomics to improve cancer care



SMC RNA Round 3 Isoform Quantification

Entry ID	Workflow	sim46	sim47	sim48	sim49	sim50
7150823	isoform_1471603661_merged.cwl	0.900685984	0.901220059	0.8983185	0.900216196	0.899260689
7150898	isoform_1471603893_merged.cwl	0.900685984	0.901220059	0.8983185	0.900216196	0.899260689
7185755	smcIsoform_1471960823_merged.cwl	0.937827748	0.938388059	0.935550804	0.938250032	0.936523443
7363252	smcIsoform_Hera_workflow_1476415974_merged.cwl	0.847368555	0.857134368	0.84790518	0.860188186	0.860869401
7367548	smcIsoform_Hera_workflow_1476549574_merged.cwl	0.860726563	0.854973128	0.845887071	0.859998273	0.857392585
8023084	winterfell_isoform_r2_rspd_b1	0.926925616	0.930314124	0.925143894	0.927371855	0.928692024
8057844	k27	0.901556743	0.901081506	0.898600064	0.900330374	0.898702192
8064391	smcIsoform_Hera_workflow_1485238785_merged.cwl	0.902614171	0.920109533	0.919696658	0.918259168	0.922281783
8066038	smcIsoform_Hera_workflow_1485267086_merged.cwl	0.902627384	0.92010953	0.919716034	0.918270795	0.913222666
8277921	Isoform - rsem w/ bowtie2	0.929763347	0.933696508	0.927963461	0.930224398	0.931750869
8277922	Isoform - rsem w/ bowtie2	0.929763347	0.933696508	0.927963461	0.930224398	0.931750869
8294423	kilo27	0.901720503	0.902152152	0.89875754	0.90068149	0.898642423
8473425	Salmon	0.898876397	0.897720817	0.896396752	0.897728471	0.89539482

SMC-RNA Round 3 Fusion Detection

Entry ID	Workflow	sim46	sim47	sim48	sim49	sim50
7185743	smcFusion_1471957621_merged.cwl	0.918919	0.888889	0.967033	0.903226	0.949495
7238196	CWL_jaffa_worklow_1473812025_merged.cwl	NA	0.666667	0.633333	0.685714	0.75
7252855	mps3_workflow_1474050212_merged_1474055180_merg ed.cwl	NA	NA	NA	0.275862	NA
7269246	FusionRnadt_workflow_1474369510_merged.cwl	0.30137	0.139535	0.191304	0.117647	0.243478
7378934	comb_workflow_v2_1476704303_merged.cwl	0.42515	0.175439	0.401826	0.12844	0.356364
7488045	smcFusion_Hera_workflow_1478021629_merged.cwl	0.265306	0.424242	0.280702	0.1	0.262295
7997358	CWLworkflow_round2_1483936483_merged.cwl	0.878378	0.809524	0.906977	0.896552	0.783505
8114525	STAR_SEQR0_1485986782_merged.cwl	0.900662	0.837209	0.955556	0.933333	0.94
8226744	smcFusion_Hera_workflow_1486535300_merged.cwl	0.333333	0.4375	0.285714	0.222222	0.294118
8281648	Winterfell Fusion Feb-2017	0.918919	0.863636	0.967033	0.866667	0.938776
8396803	Sierra	0.690141	0.470588	0.637363	0.27907	0.510204
8399080	STARSEQRv050_merged_1489019984_merged.cwl	0.932432	0.956522	0.977778	0.967742	0.970297
8517255	fusioncatcher_1	0.780142	NA	0.847826	0.705882	0.690476

The workflows generated by the SMC-RNA challenge have been released publically

CWL Workflows

Docker Images

Reference Files

The image displays three screenshots related to the SMC-RNA challenge:

- GitHub:** Shows the repository page for `smc-rna-challenge` under the organization `smc-rna-challenge`. It lists several Python-based challenge submissions, including `zhanghj-8555063`, `Ginny-7899344`, `bhaas-7185755`, `bhaas-8277921`, and `tnv-8614991`, all updated 5 days ago.
- Quay.io:** Shows the `smc-rna-challenge` repository page, displaying a grid of Docker images. The images are organized by user, including `beccyl-7238196-jaffa`, `genomehacker-9610017-g...`, `genomehacker-9610018-g...`, `jasper1918-9610272-stars...`, `zhanghj-9609594-romeo`, and `zhanghj-9609594-romeo_...`.
- Synapse:** Shows the file page for the SMC-RNA challenge, displaying a list of reference files. The files are organized into folders, including `Entry_references`, `Ginny-7217266`, `Ginny-7217268`, `Ginny-7488045`, `Ginny-7899344`, `Ginny-7899426`, `Ginny-8050015`, `Ginny-8064391`, `Ginny-8066038`, `Ginny-8226744`, `Ginny-9607675`, `Ginny-9609145`, `Ginny-9609498`, `Ginny-9609499`, `Ginny-9609501`, `beccyl-7238196`, `beccyl-7997358`, `bhaas-7180105`, and `bhaas-7185742`.

Even the Simulator Has Now Been Handed Over to the Community...

Contestants identified issues with simulator. Admins Debugged

Feature requests posted to forum, ideas for more realism.

Solution: Made simulator available to all to modify.

Community Challenge Phase: Use Community Simulator!

GitHub

The screenshot shows the GitHub repository page for Sage-Bionetworks / rnaseqSim. The repository has 88 commits, 2 branches, 0 releases, 4 contributors, and is licensed under Apache-2.0. The current branch is master. The repository is watched by 8 people, starred by 4, and forked by 1. The commit history shows several recent changes:

Commit	Message	Time
fastq_create	Deleted.	2 months ago
fastq_create_polyester	Reordering for modular.	7 months ago
fusion_create	Modular (#2)	3 months ago
genome_create	Modular (#2)	3 months ago
model_isoforms	Modular (#2)	3 months ago

Challenges of Interdisciplinary Omics Analyses

- Complex pipelines
 - And many method choices at every step
 - Methods are computationally expensive
 - Every method different requirements
- Diverse sources of data needed
 - Scattered
 - Too large to move around
- Too complex to understand
 - Need interpretation
 - Need intuitive overviews
 - Need integration for “big picture”

Challenges of Interdisciplinary Omics Analyses

- Complex pipelines
 - And many method choices at every step
 - Methods are computationally expensive
 - Every method different requirements
- Diverse sources of data needed
 - Scattered
 - Too large to move around
- Too complex to understand
 - Need interpretation
 - Need intuitive overviews
 - Need integration for “big picture”

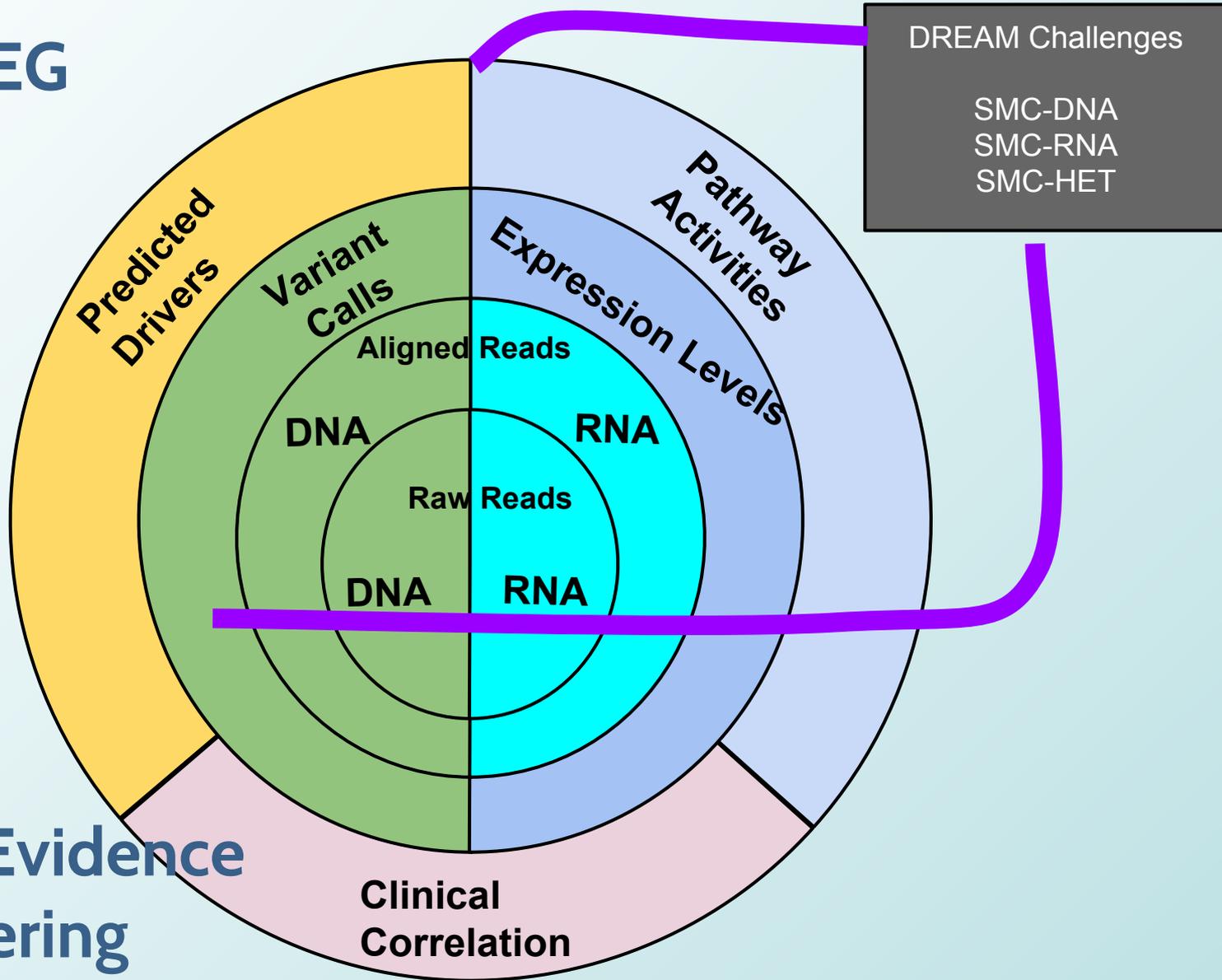
Challenges of Integrative / Interconnected Analyses

Biomedical Evidence Graph (BMEG) Aims

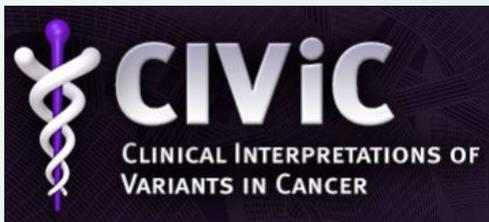
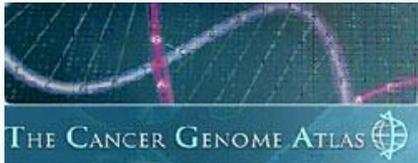
- Enable integrative / interconnected analyses
 - Create layers of linked evidence
 - Predictive signatures
 - Correlations to infer subtypes
 - Networks to infer n-of-1 mechanisms
- Diverse sources of data needed
 - Scattered
 - Too large to move around
- Too complex to understand
 - Need interpretation
 - Need intuitive overviews
 - Need integration for “big picture”

BMEG

Layered Evidence Gathering & Generation



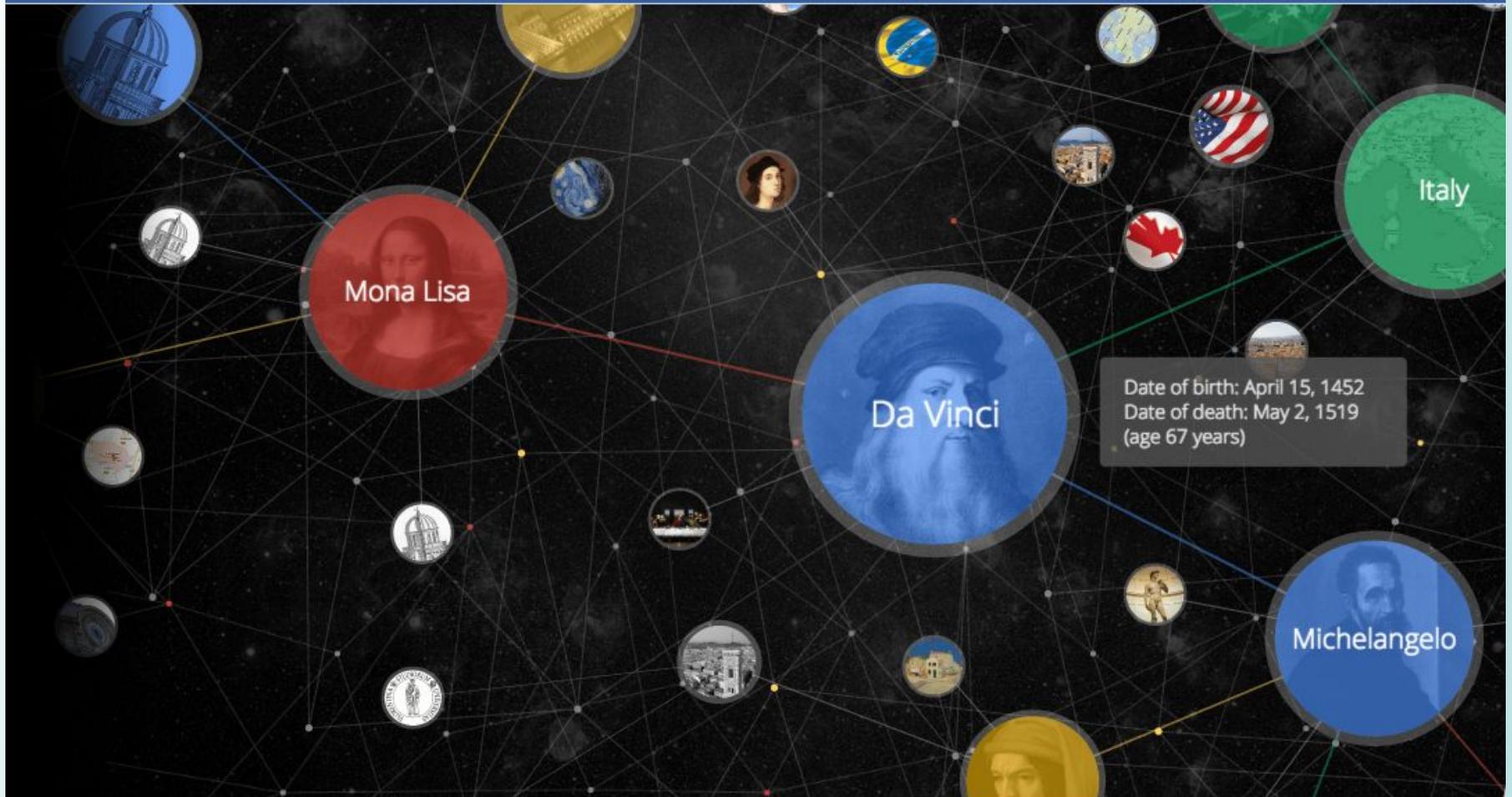
There is lots of data out there



The Google Knowledge Graph

Google Inside Search

[Home](#) [How Search Works](#) [Tips & Tricks](#) [Features](#) [Search Stories](#) [Playground](#) [Blog](#) [Help](#)



The Google Knowledge Graph



leonardo da vinci



All

Images

Books

Videos

News

More

Settings

Tools

About 8,970,000 results (0.69 seconds)

Leonardo da Vinci - Wikipedia

https://en.wikipedia.org/wiki/Leonardo_da_Vinci

Leonardo di ser Piero da Vinci more commonly **Leonardo da Vinci** or simply Leonardo, was an Italian polymath whose areas of interest included invention, ...

[Vitruvian Man](#) · [Mona Lisa](#) · [The Last Supper](#) · [Lady with an Ermine](#)

Leonardo da Vinci - Artist, Mathematician, Inventor, Writer - Biography ...

www.biography.com/people/leonardo-da-vinci-40396

Nov 17, 2015 - Artist, Mathematician, Inventor, Writer(1452–1519) ... **Leonardo da Vinci** was a leading artist and intellectual of the Italian Renaissance who's known for his enduring works "The Last Supper" and "Mona Lisa." ... Born on April 15, 1452, in **Vinci**, Italy, **Leonardo da Vinci** was the epitome ...

Leonardo da Vinci - Facts & Summary - HISTORY.com

www.history.com/topics/leonardo-da-vinci

Leonardo da Vinci (1452-1519) was a painter, architect, inventor, and student of all things scientific. His natural genius crossed so many disciplines that he epitomized the term "Renaissance man." Today he remains best known for his art, including two paintings that remain ...

Leonardo Da Vinci - The complete works

www.leonardoda-vinci.org/

Leonardo Da Vinci - Homepage. The complete works, large resolution images, ecard, rating, slideshow and more! One of the largest **Leonardo Da Vinci** ...

Home | Leonardo Da Vinci - The Genius - Museum of Science, Boston

<https://www.mos.org/leonardo/node/1>

Who Was **Leonardo Da Vinci**? While **Leonardo da Vinci** is best known as an artist, his work as a scientist and an inventor make him a true Renaissance man.

Leonardo da Vinci Paintings, Drawings, Quotes, Biography

www.leonardodavinci.net/

Leonardo da Vinci was a true genius who graced this world with his presence from April 15, 1452 to May 2, 1519. Like Athens in the age of Pericles, ...



More images

Leonardo da Vinci



Mathematician

Leonardo di ser Piero da Vinci, more commonly Leonardo da Vinci or simply Leonardo, was an Italian polymath whose areas of interest included invention, painting, sculpting, architecture, science, music, ... [Wikipedia](#)

Born: April 15, 1452, Anchiano

Died: May 2, 1519, Clos Lucé, Amboise, France

Siblings: Bartolomeo da Vinci, Giovanni Ser Piero, More

Parents: Piero Fruosino di Antonio da Vinci, Caterina da Vinci

Quotes

View 7+ more

Simplicity is the ultimate sophistication.

As a well-spent day brings happy sleep, so a life well spent brings happy death.

Learning never exhausts the mind.

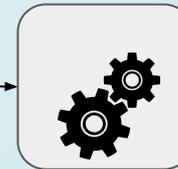
Scalable Graph Database Framework

Option - Client Library

- Python
- Javascript
- GO



GAIA



Graph Analysis
Integration and
Analysis

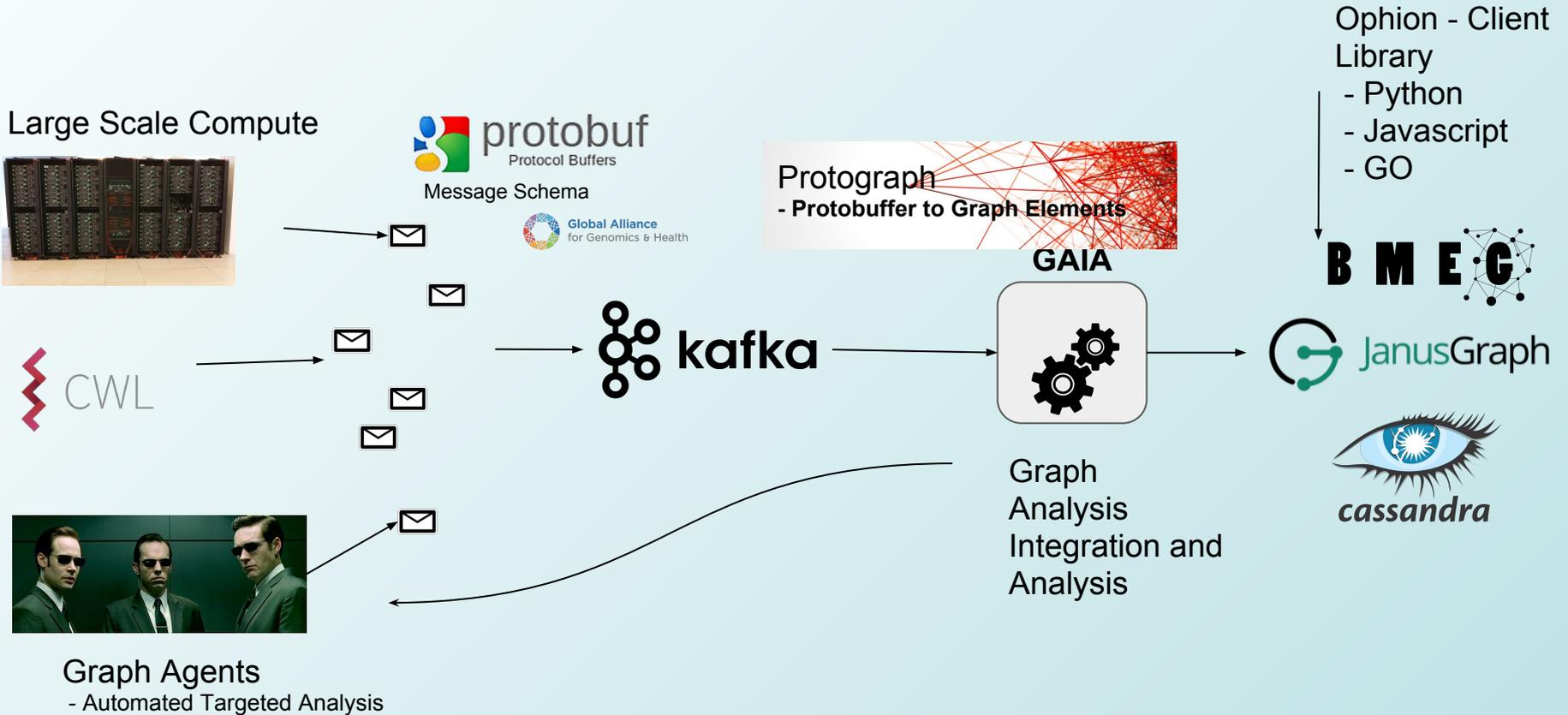


Large Scale Compute

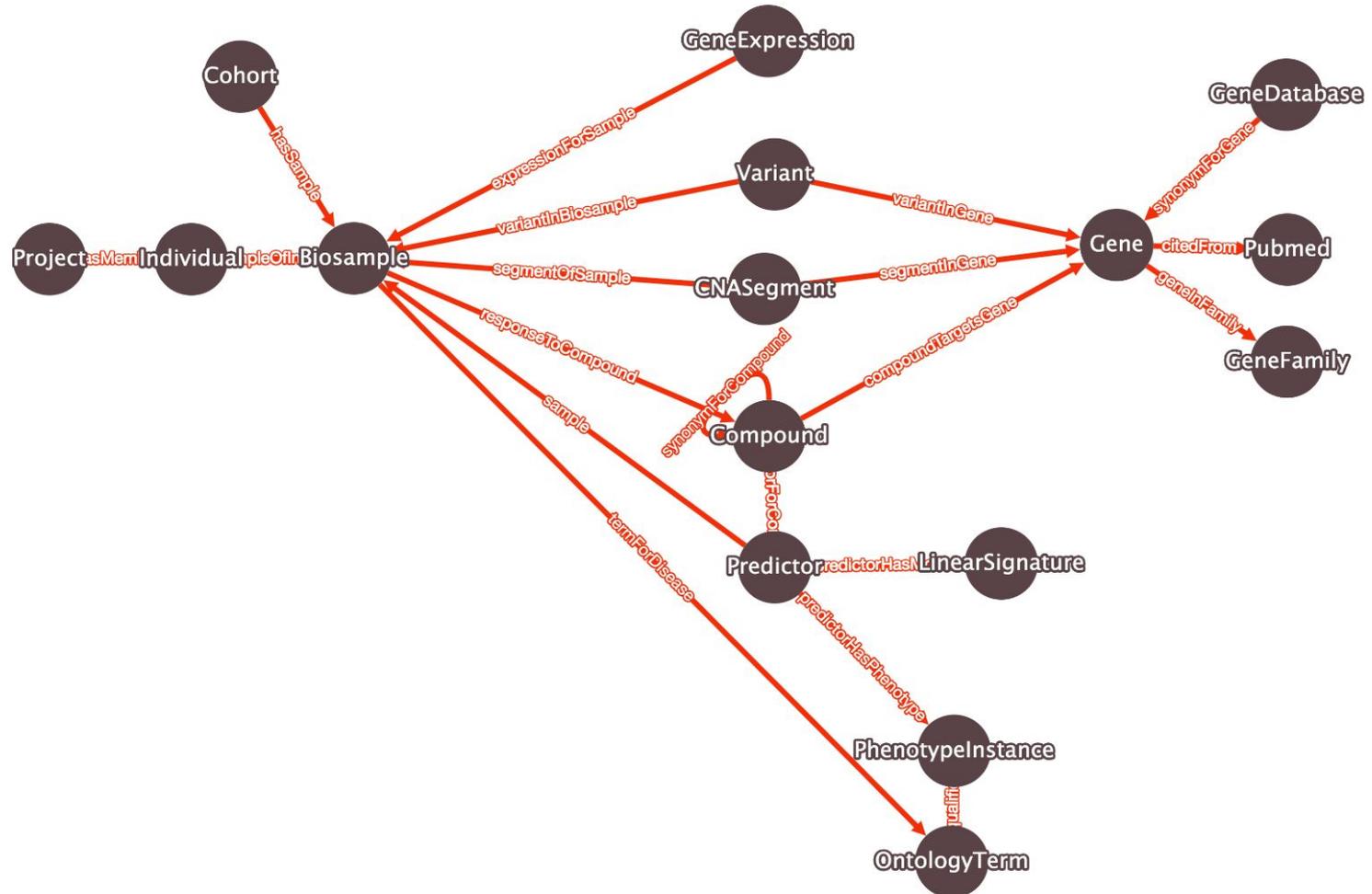


Graph Agents

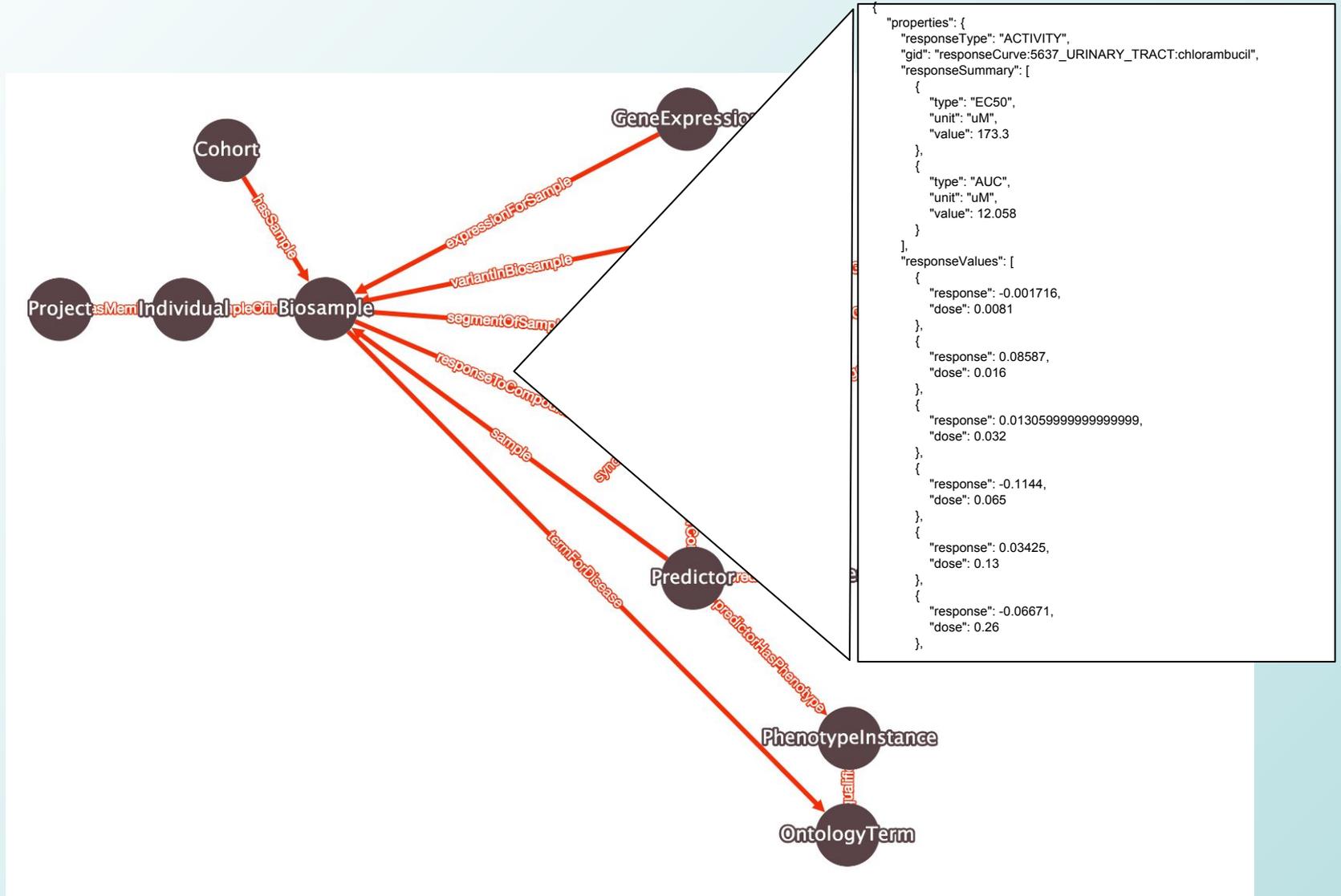
- Automated Targeted Analysis



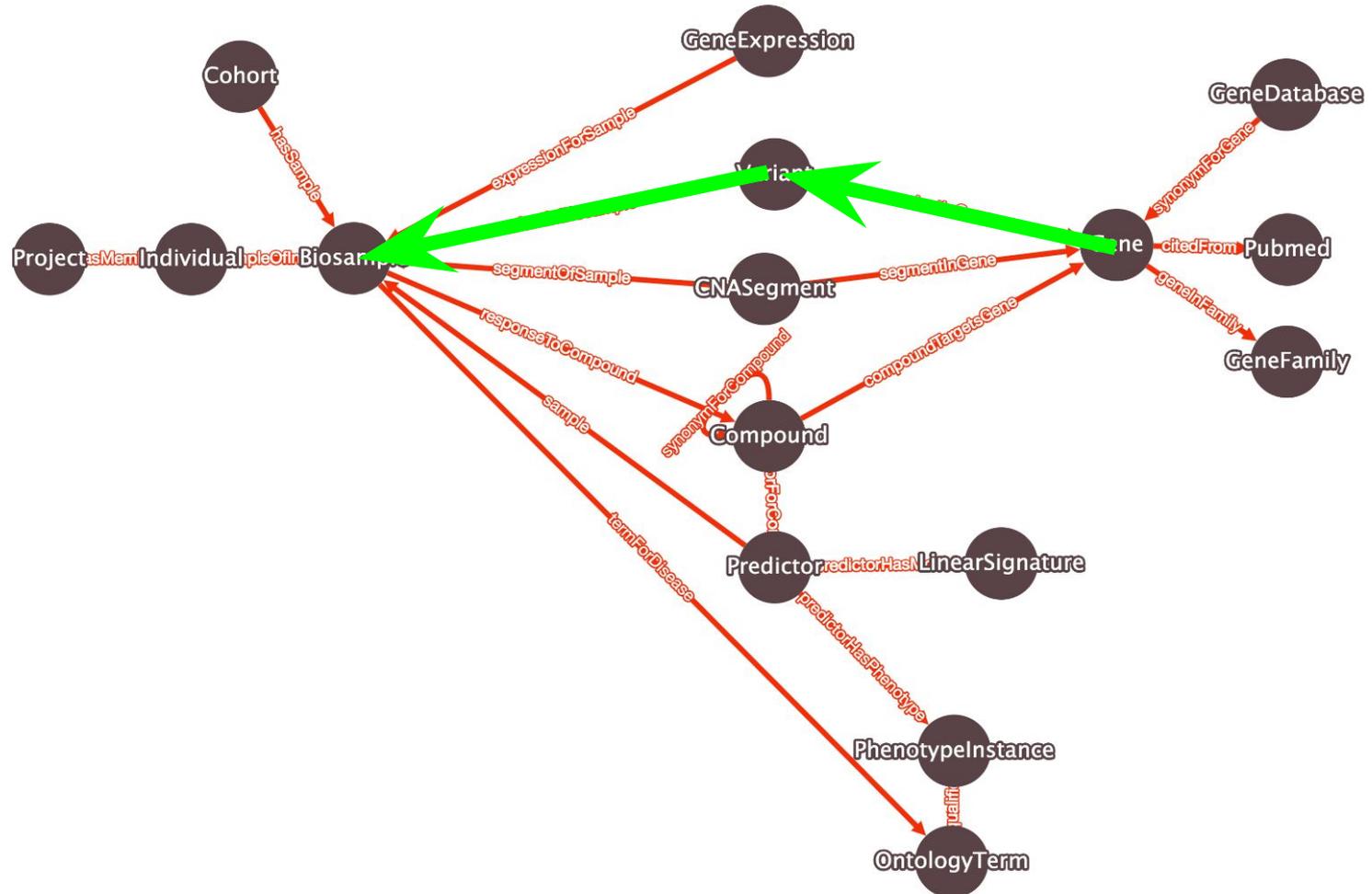
BMEG: Public Graph of Evidence



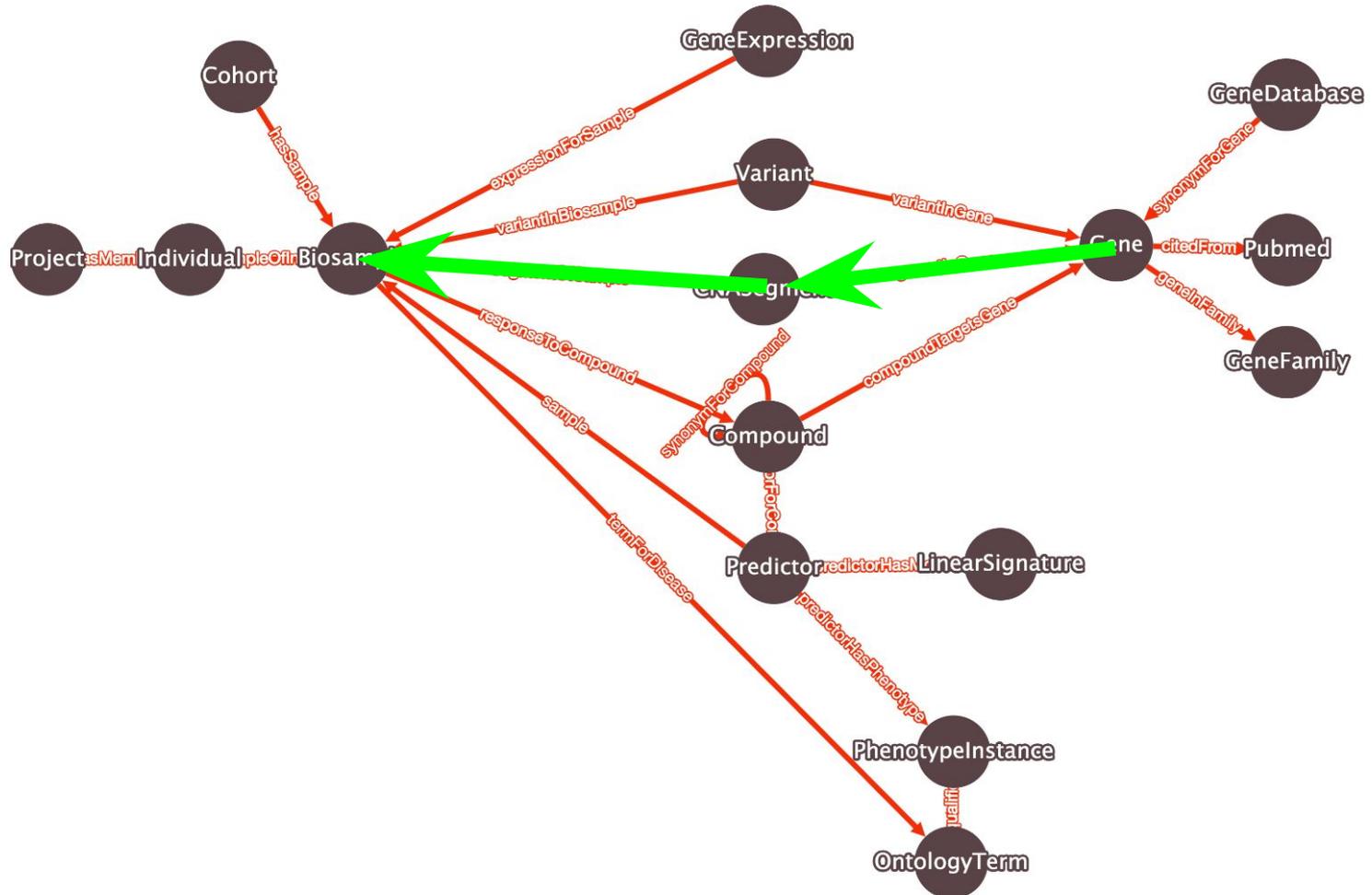
BMEG: A dense property graph



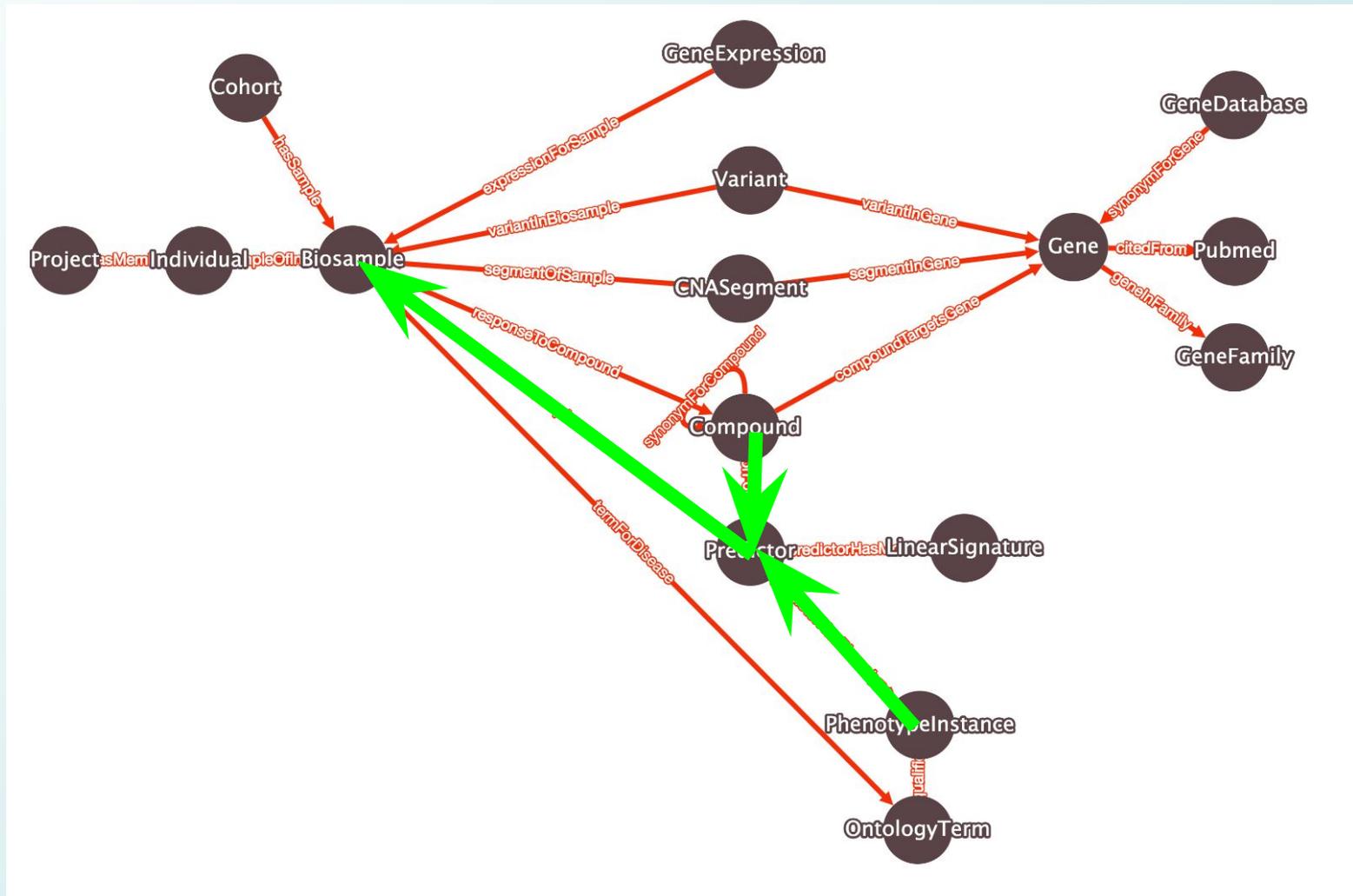
Find Samples by Variant



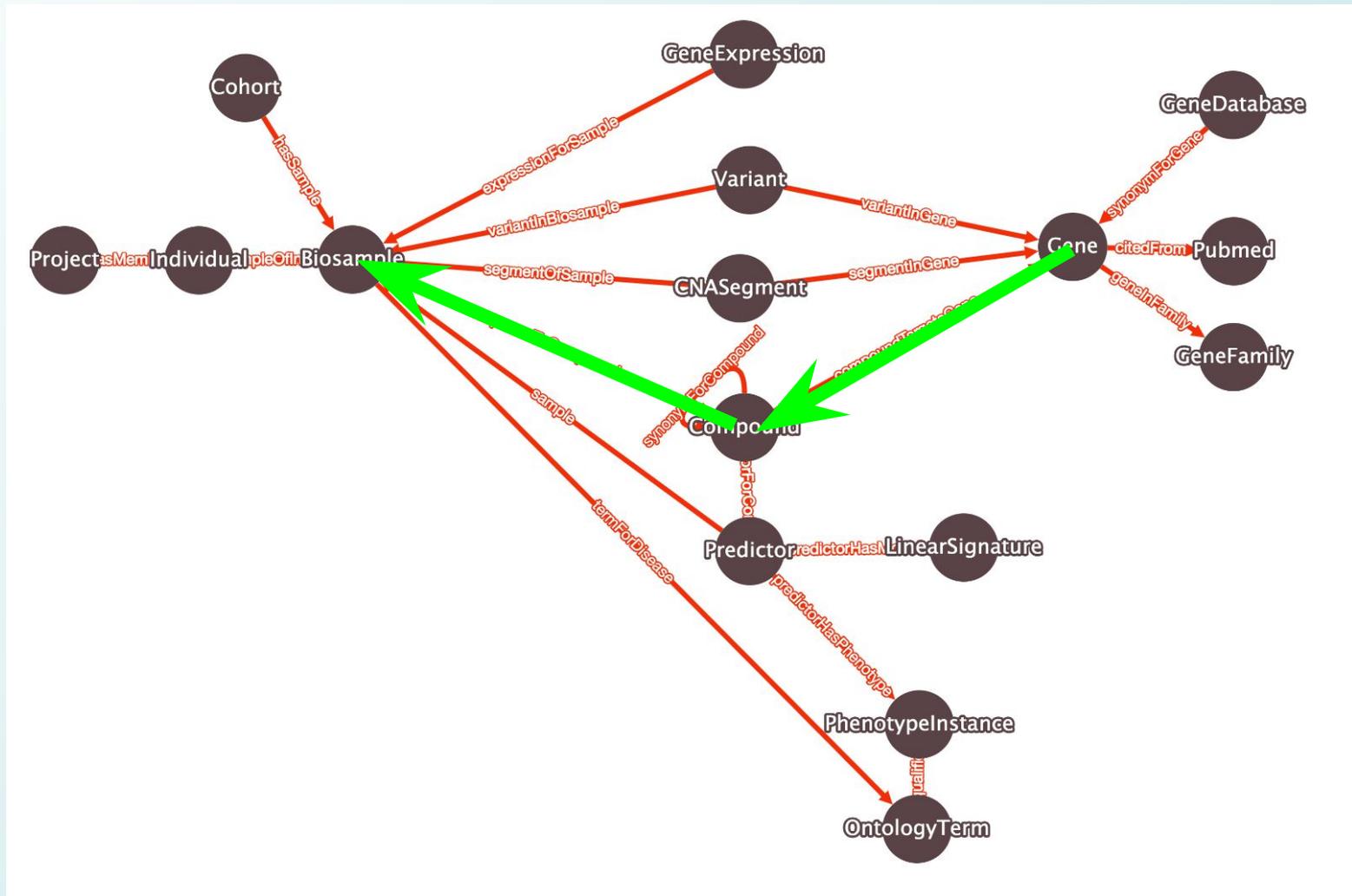
Find Samples by CNA Event



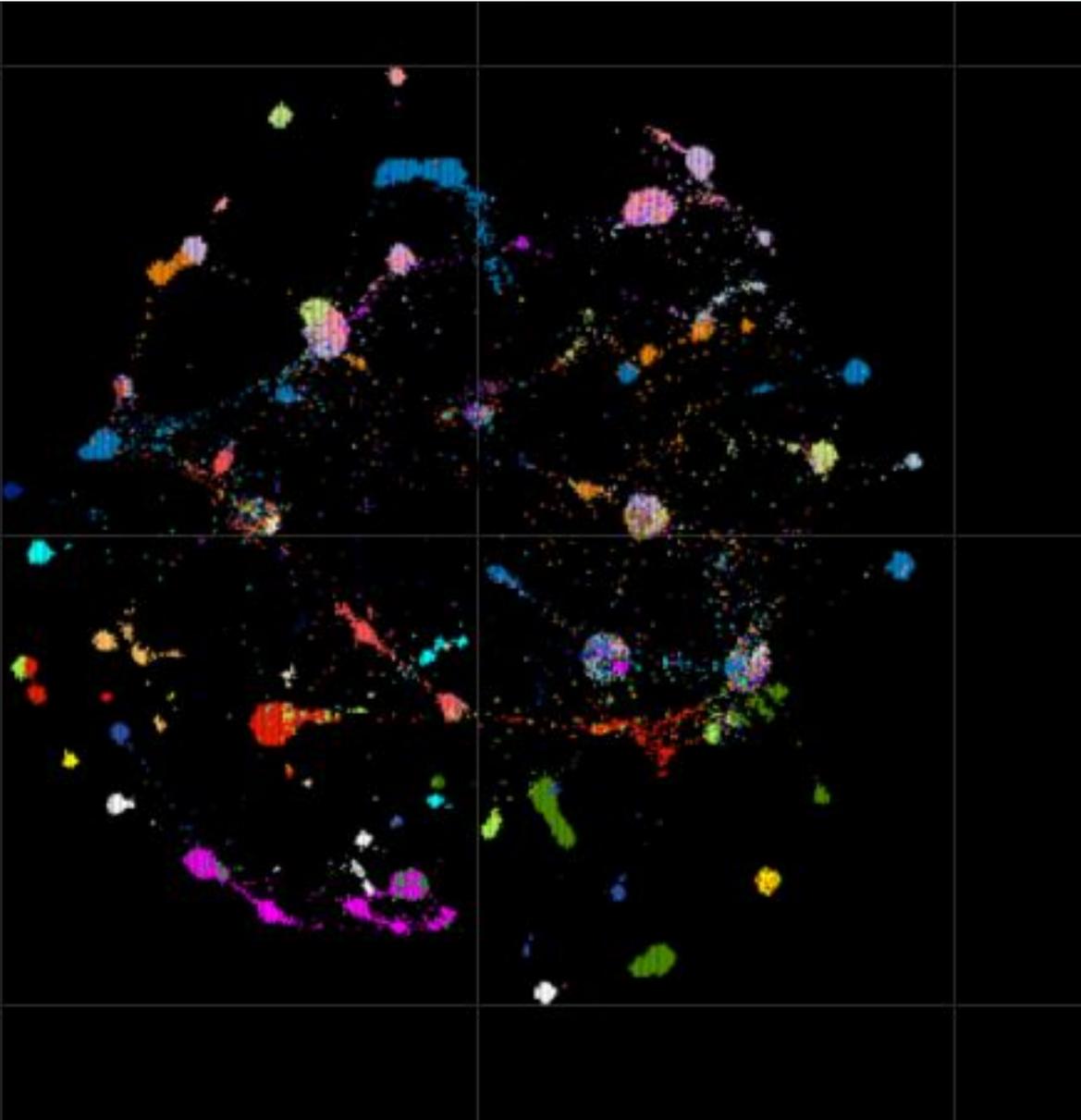
Find Samples predicted to be sensitive to a compound



Find Samples with a particular IC50 for a compound that targets a particular gene



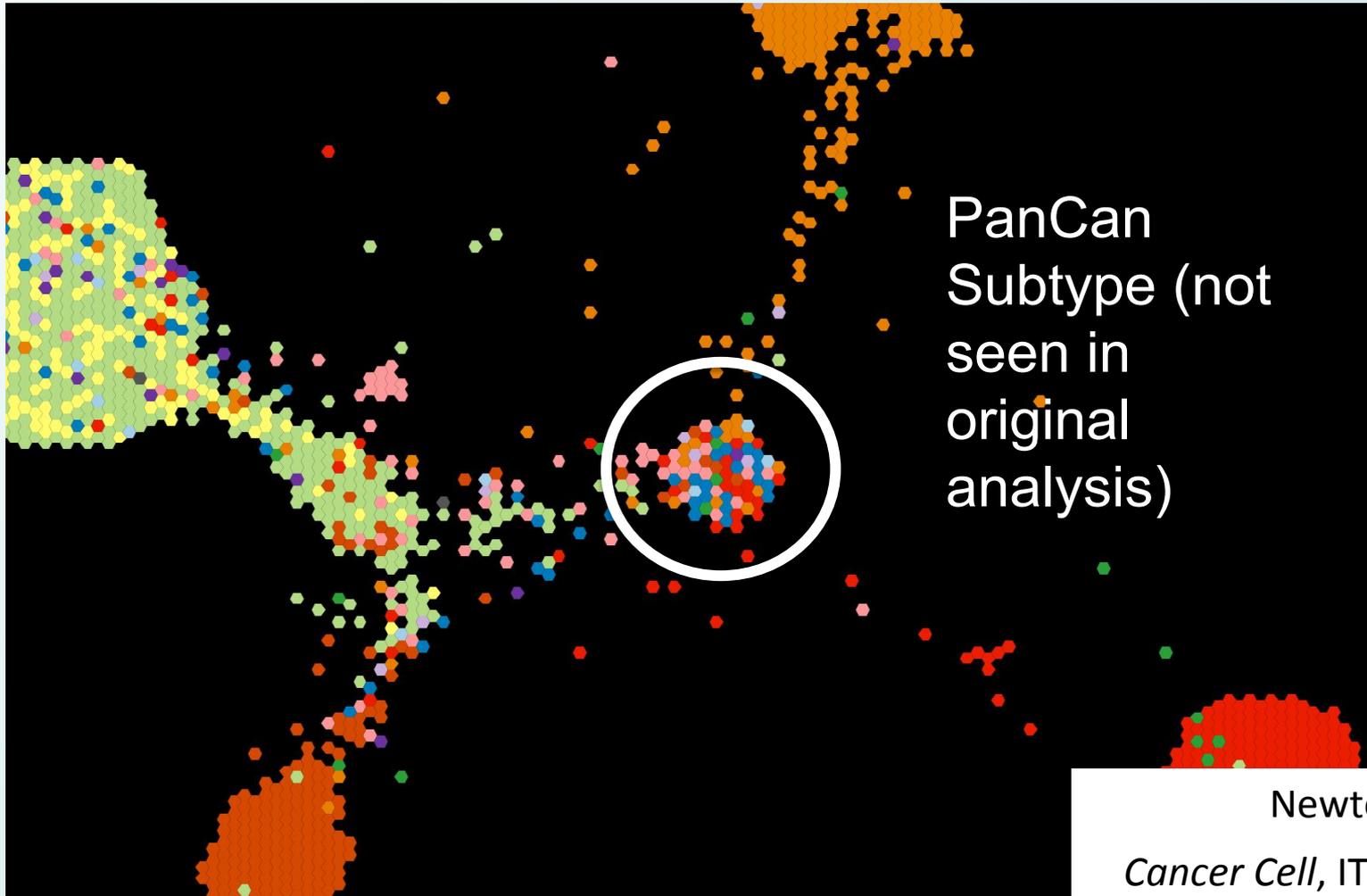
PanCan-33 TumorMap



Colors show
Tissue of origin.

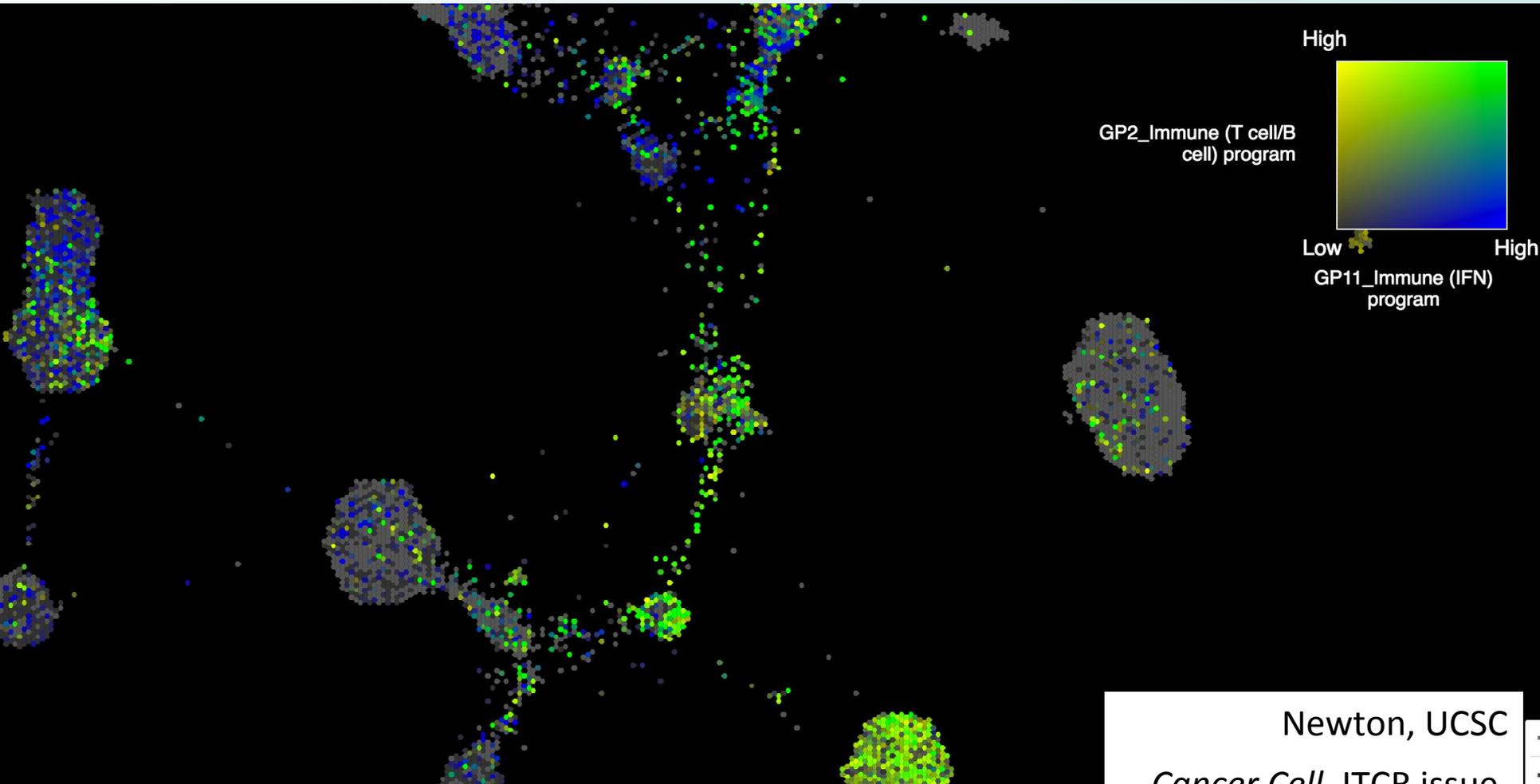
Newton, UCSC
Cancer Cell, ITCR issue,
in review

Integrated map reveals pancan subtypes



Newton, UCSC
Cancer Cell, ITCR issue,
in review

Enriched for t/B and IFN immune (D. Wolf's) programs

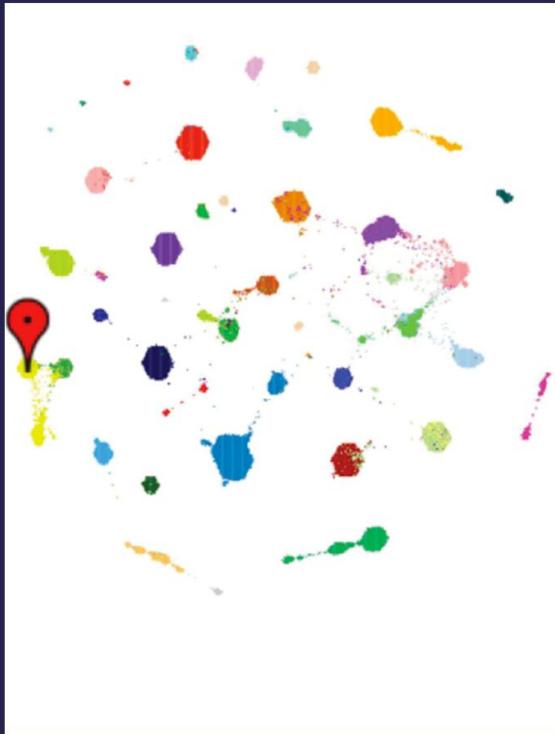


Newton, UCSC
Cancer Cell, ITCR issue,
in review

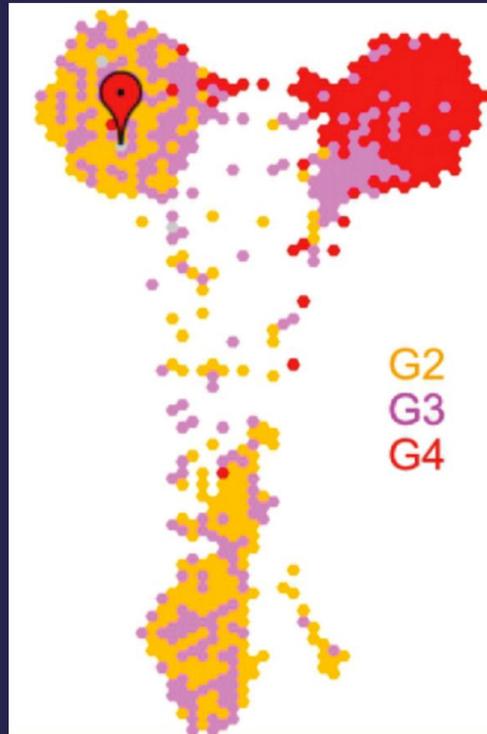


Treehouse

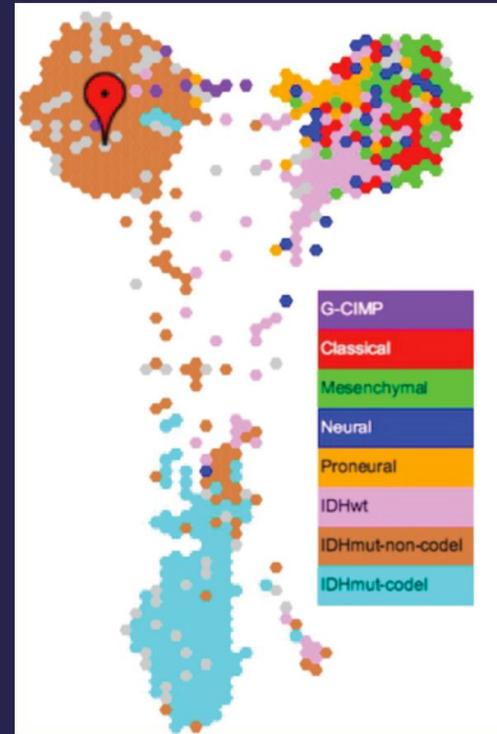
CHILDHOOD CANCER INITIATIVE



Bird's eye view



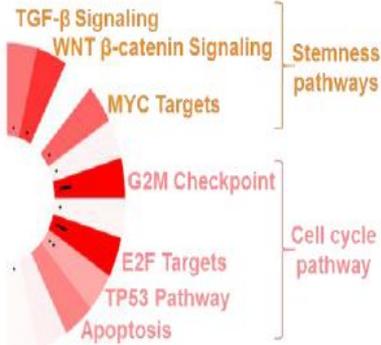
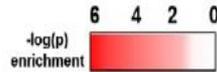
Zoom in on the glioma region
(tumors now colored by grade)



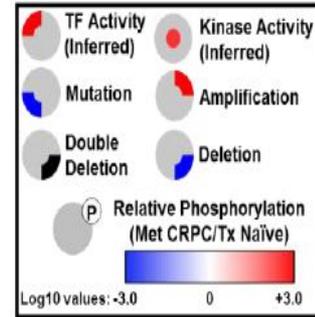
Zoom in on the glioma region
(tumors now colored by glioma subtype)

pCHIPS for Patient RA40

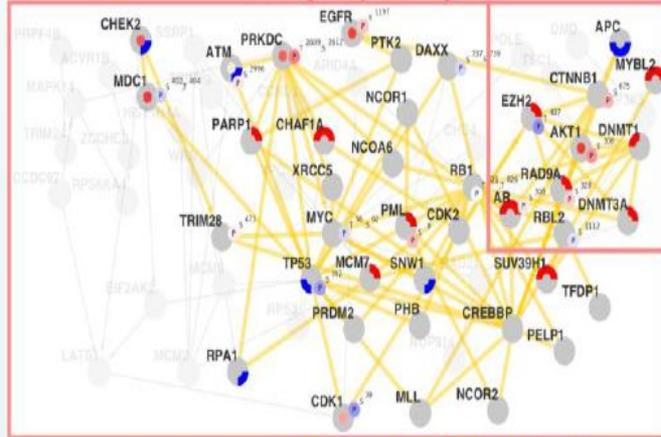
Nuclear Receptor pathway
Nuclear Receptor Signaling
PI3K-AKT-mTOR pathway
PI3K-AKT-mTOR Signaling



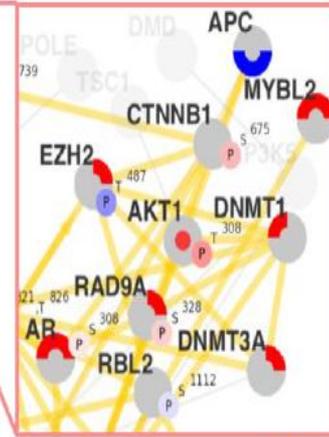
KEY: Patient RA40



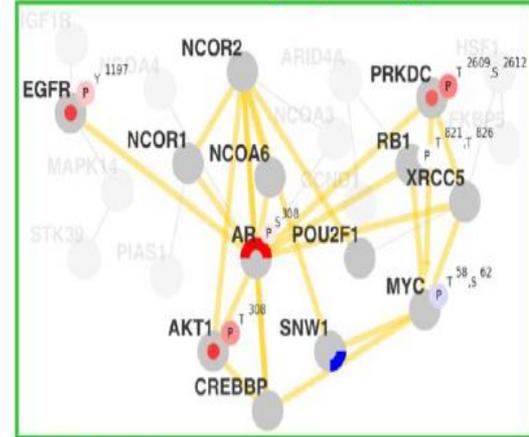
B Cell cycle pathway



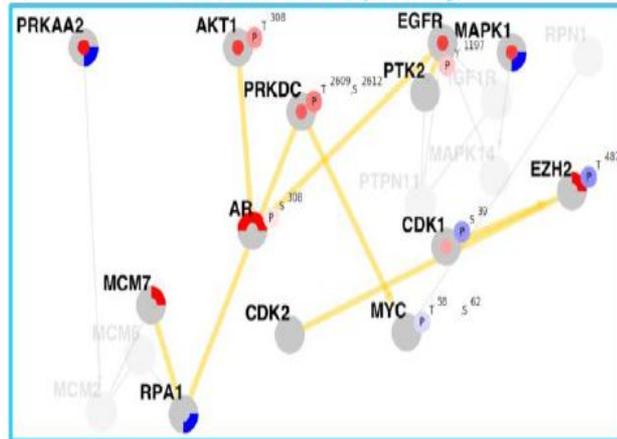
C



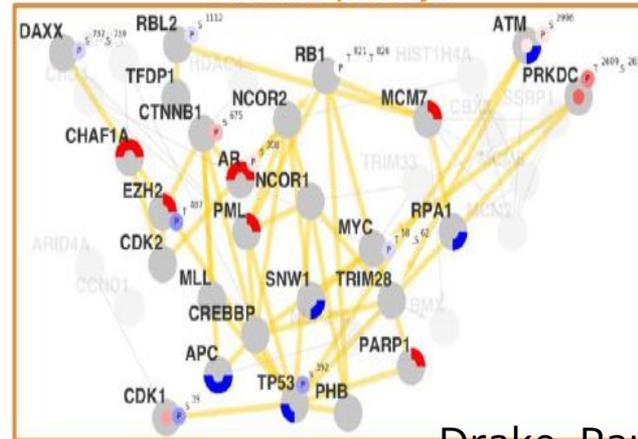
D Nuclear receptor pathway



E PI3K-AKT-mTOR pathway

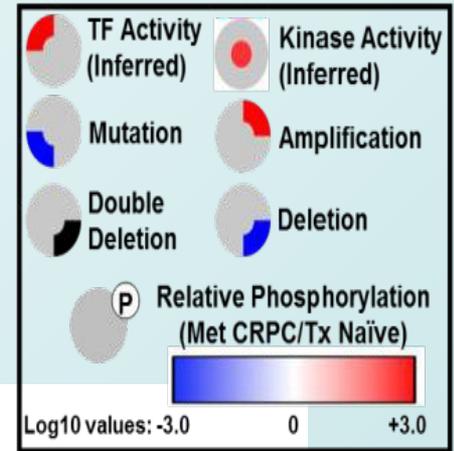


F Stemness pathways

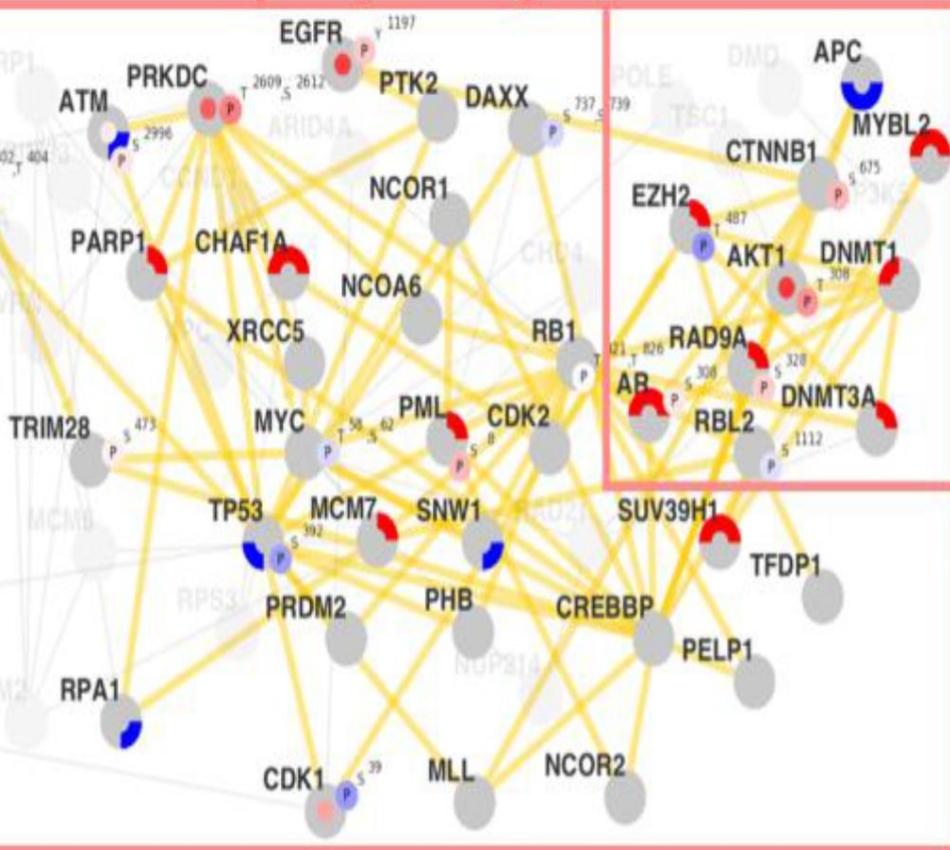


pCHIPS for Patient RA40

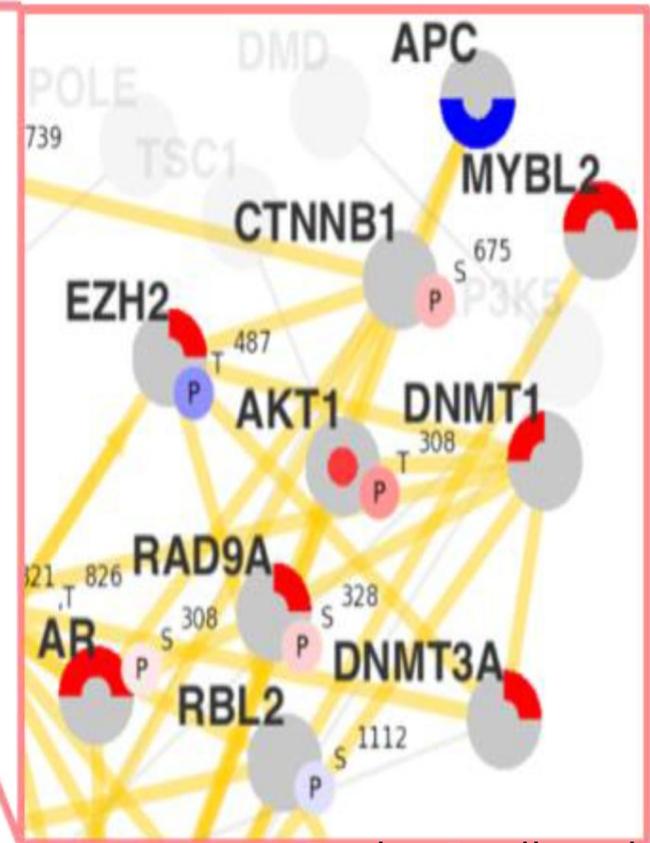
KEY: Patient RA40



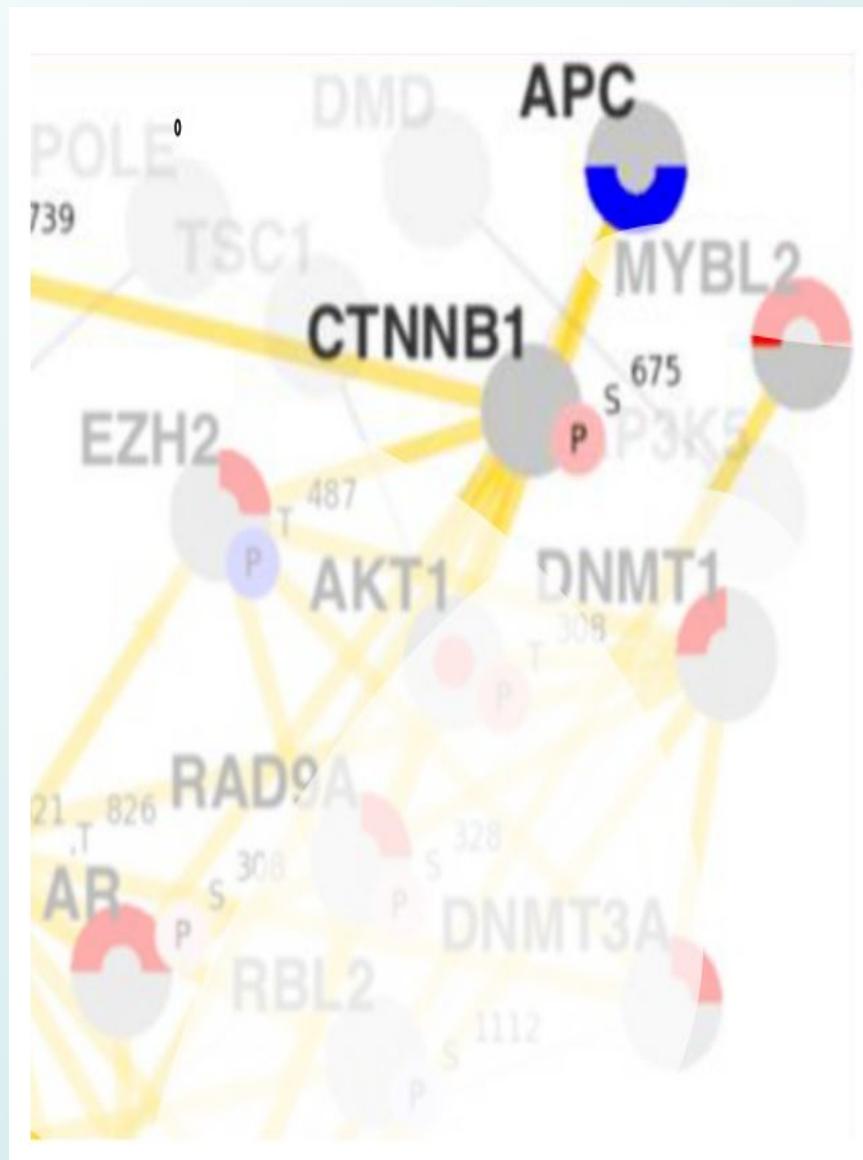
Cell cycle pathway



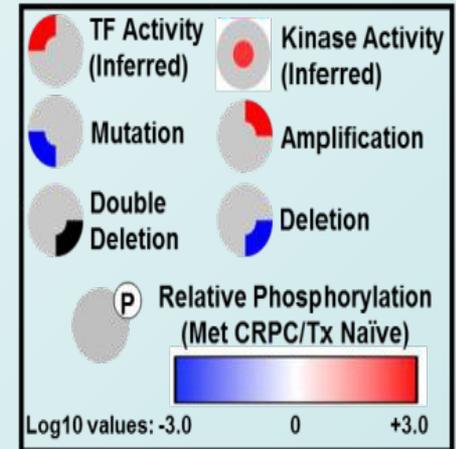
C



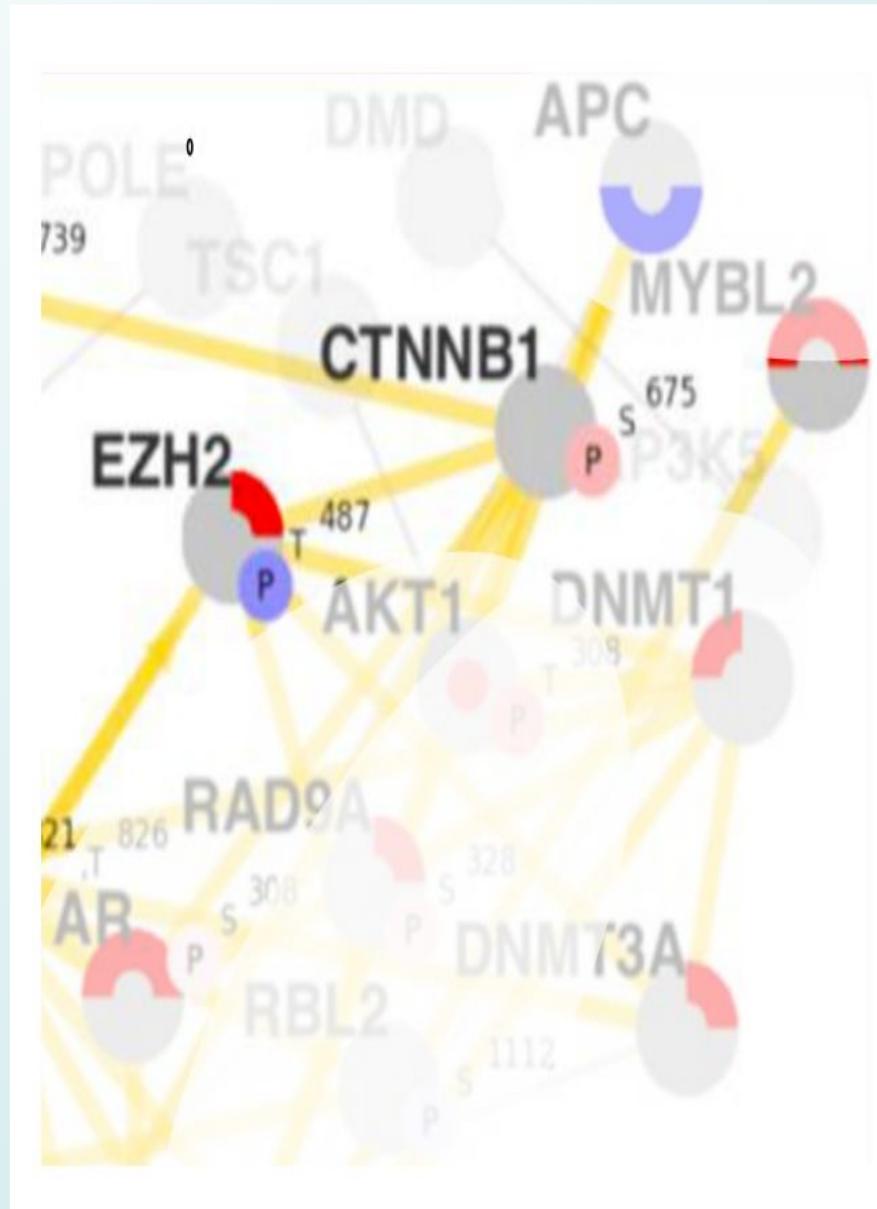
pCHIPS for Patient RA40



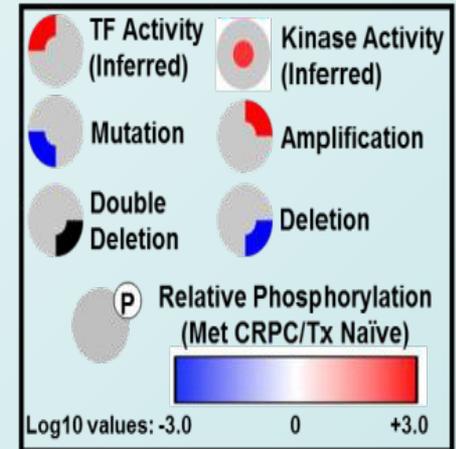
KEY: Patient RA40



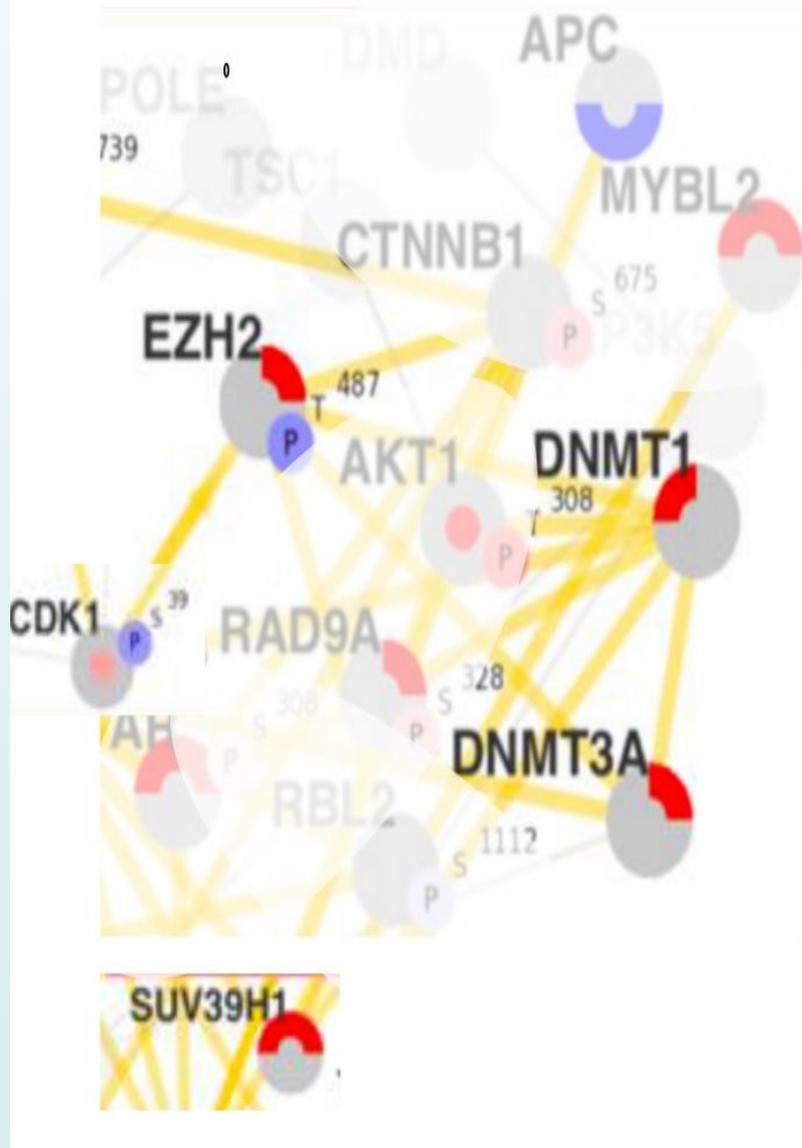
pCHIPS for Patient RA40



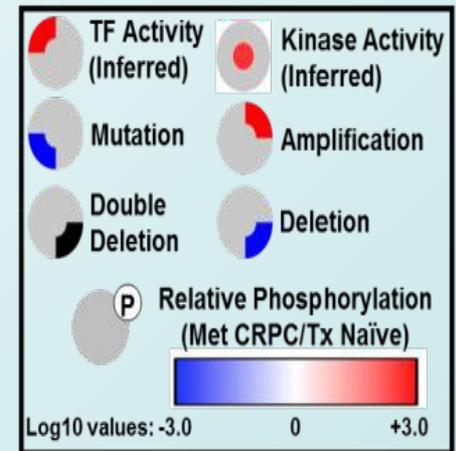
KEY: Patient RA40



pCHIPS for Patient RA40

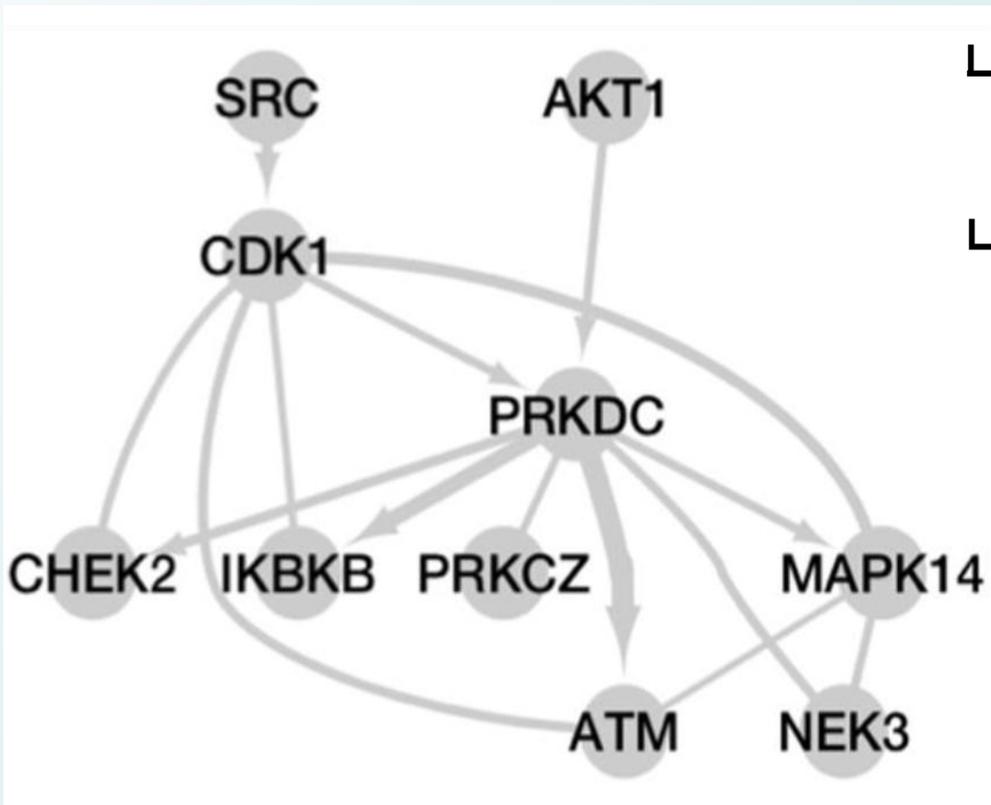


KEY: Patient RA40



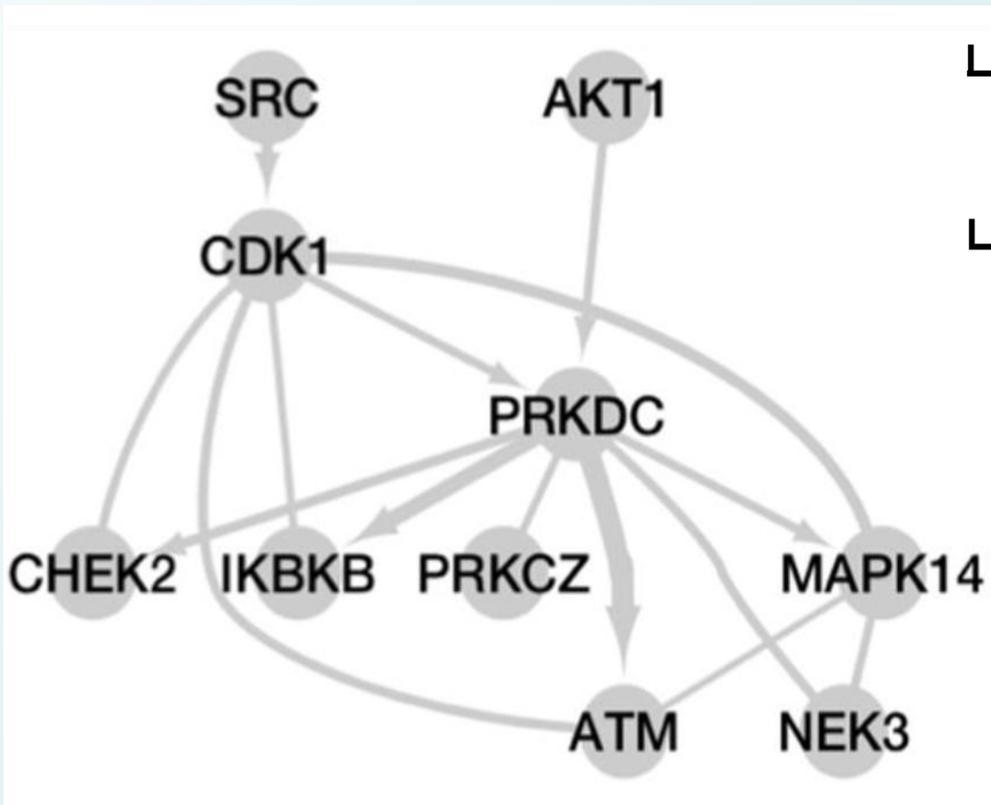
pCHIPs: Network-based selection of targets and target combinations for individual patients

B

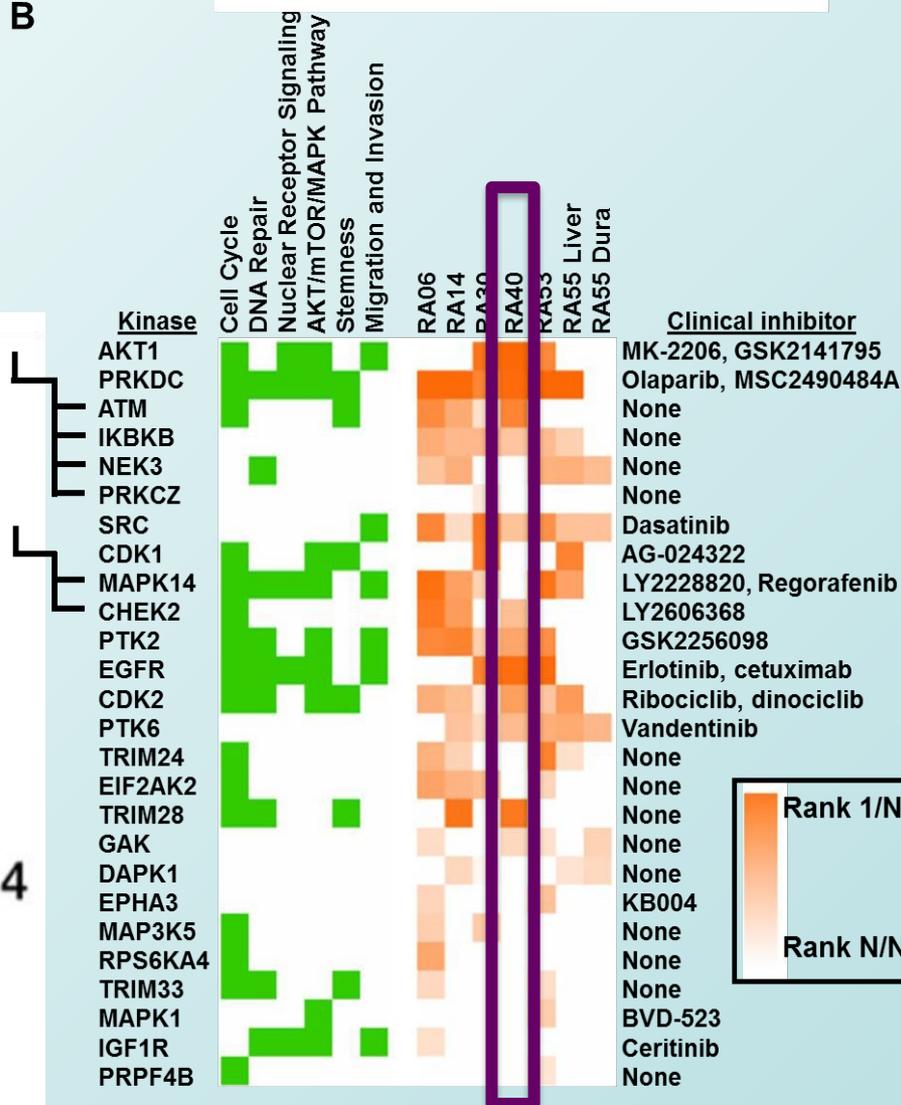


pCHIPs: Network-based selection of targets and target combinations for individual patients

Patient 40

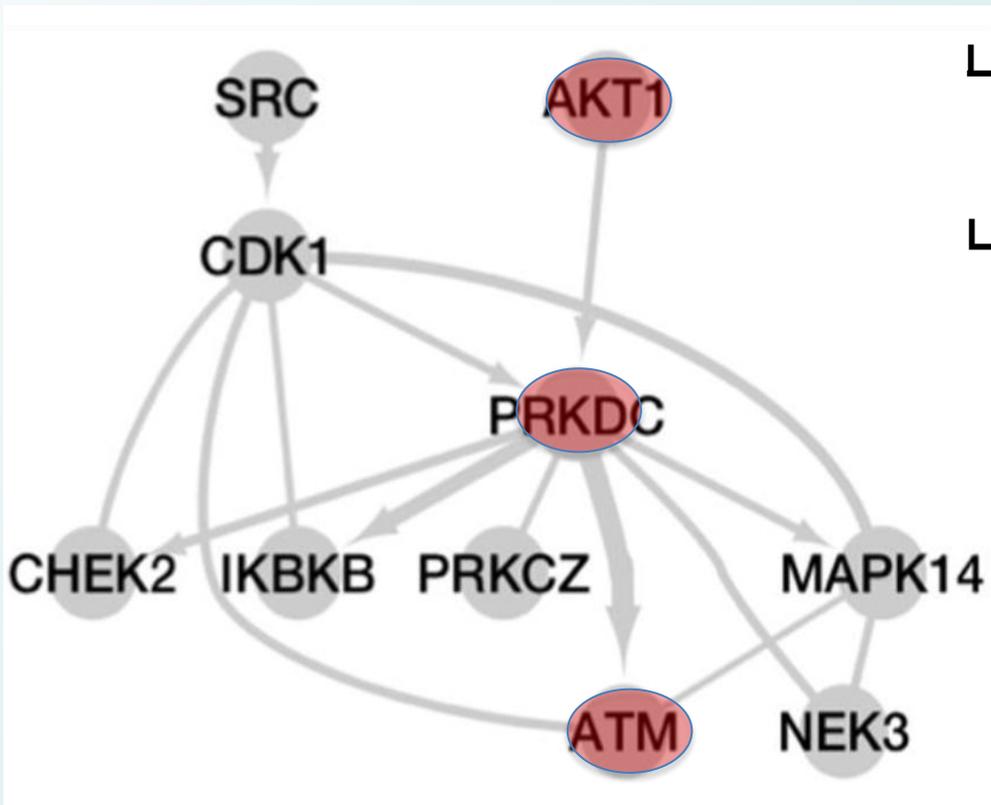


B

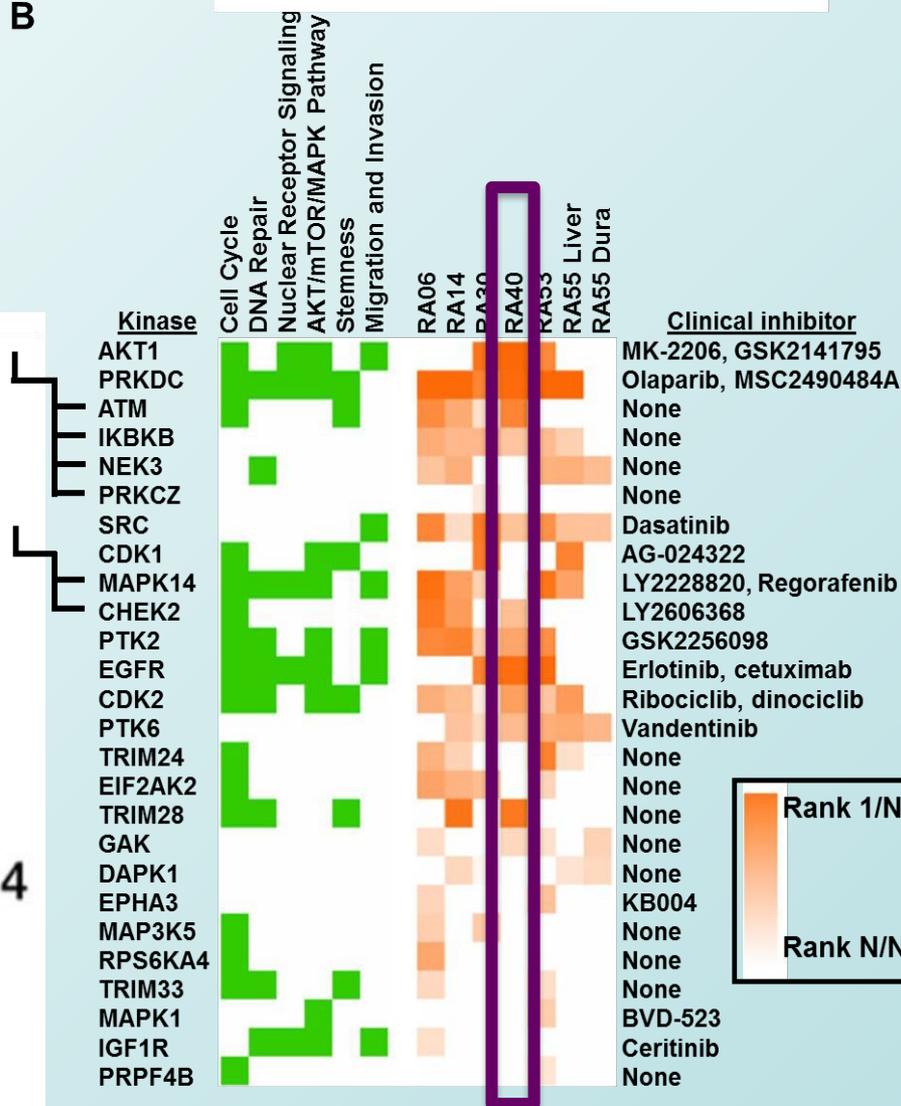


pCHIPs: Network-based selection of targets and target combinations for individual patients

Patient 40

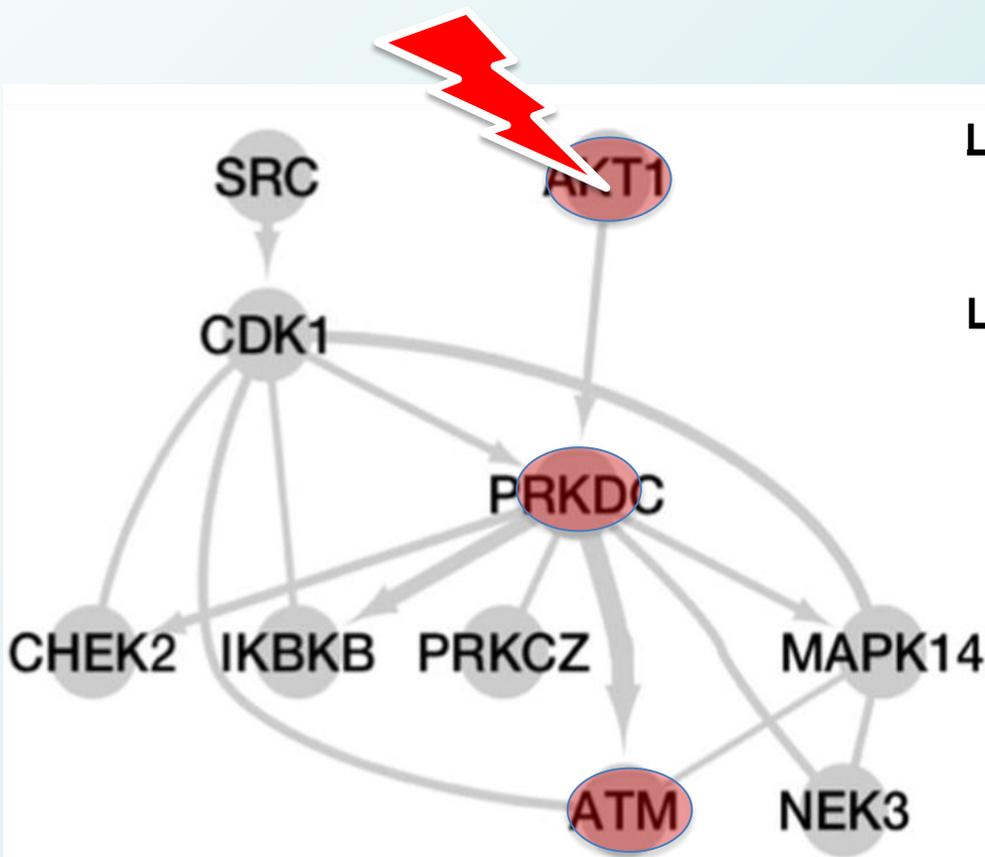


B

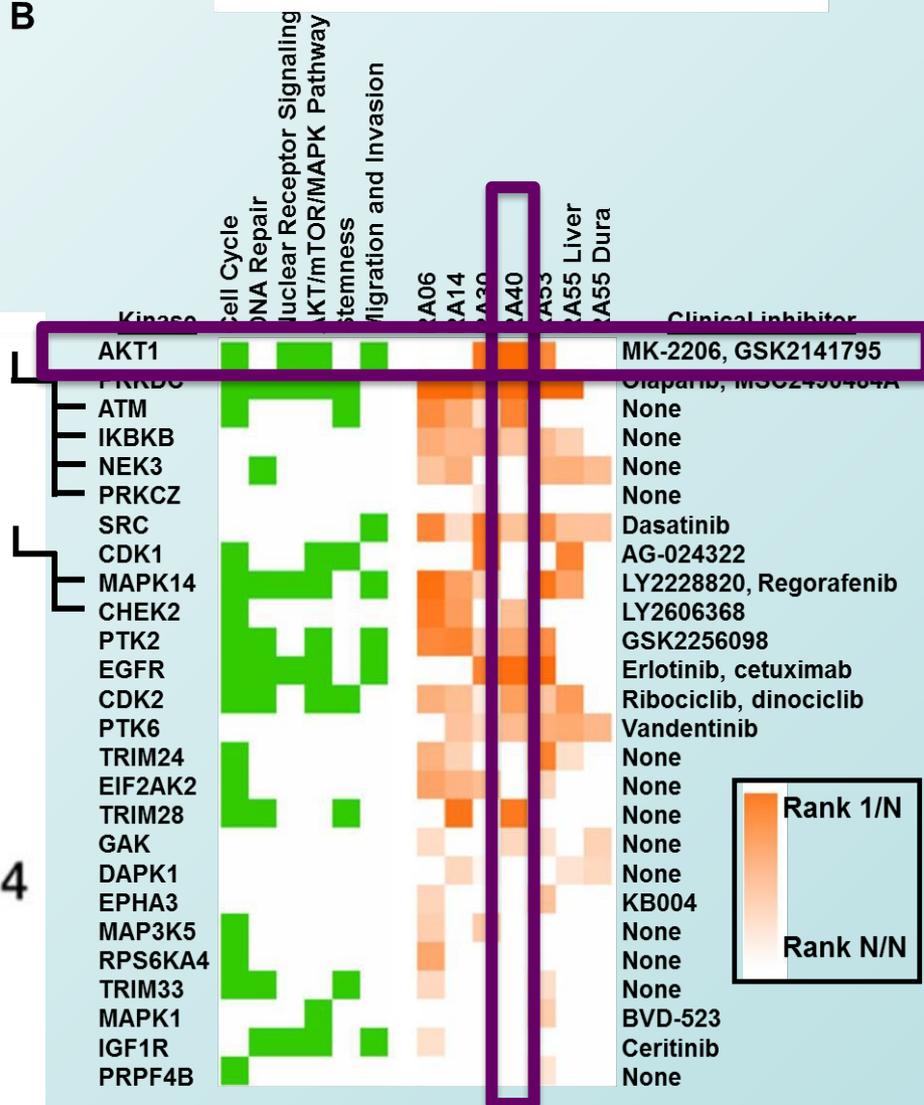


pCHIPs: Network-based selection of targets and target combinations for individual patients

Patient 40

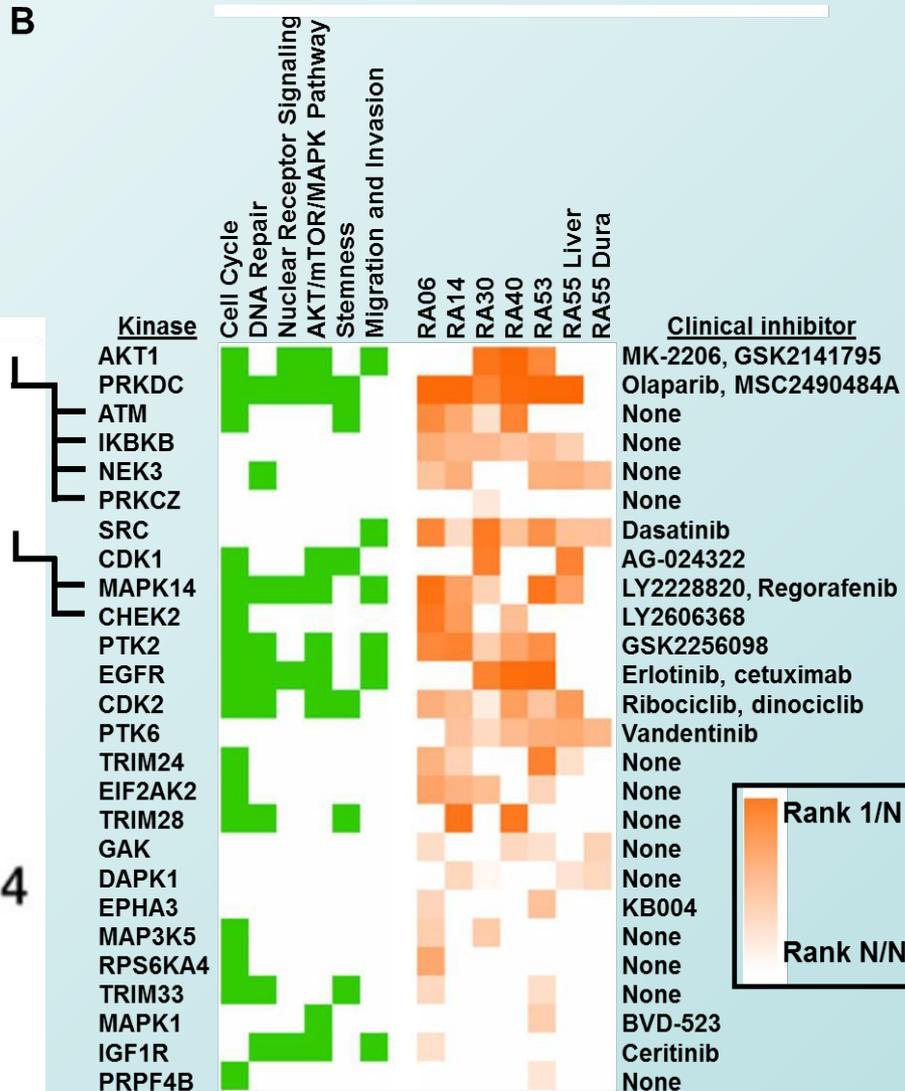
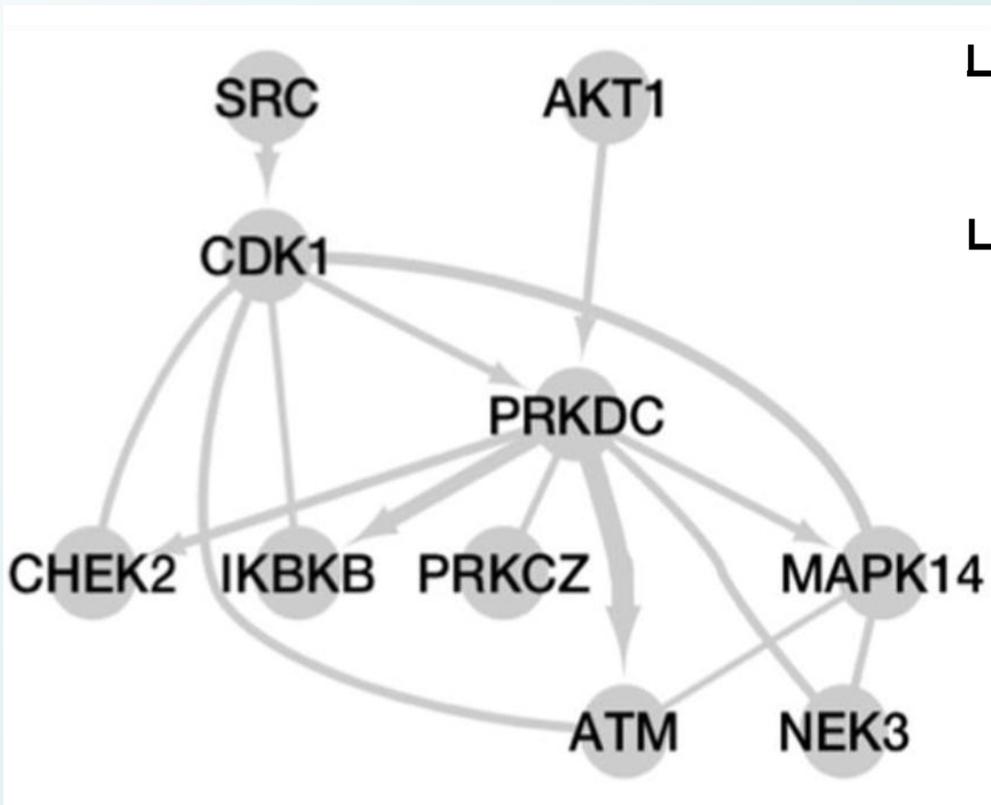


B



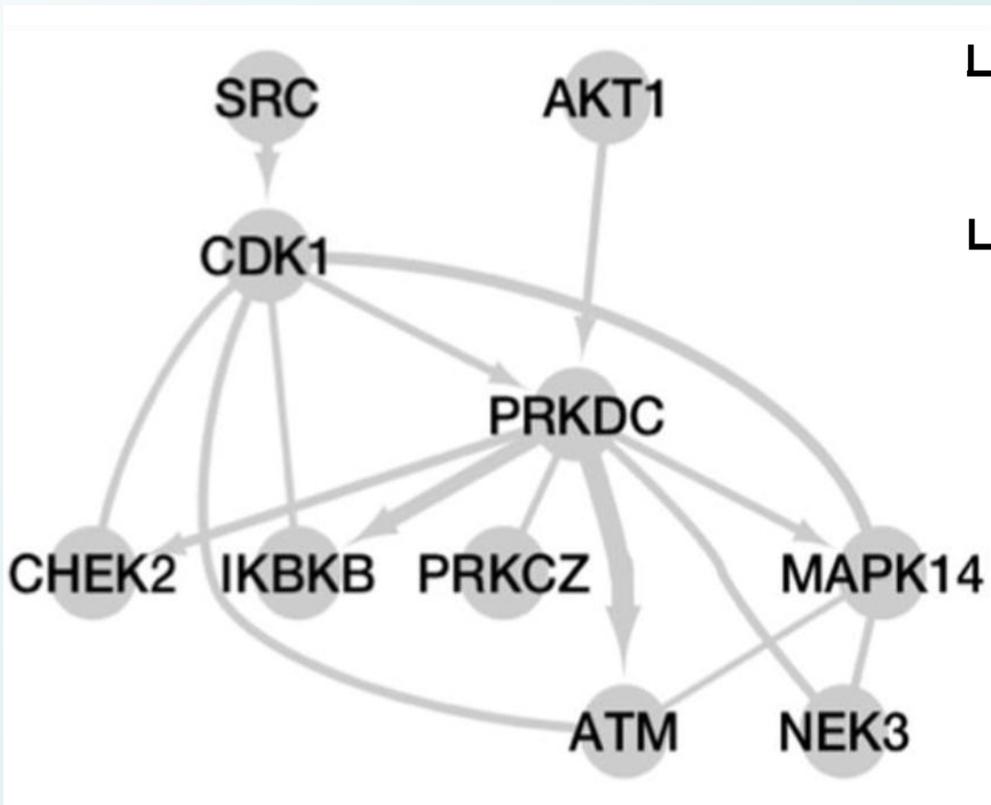
pCHIPs: Network-based selection of targets and target combinations for individual patients

B

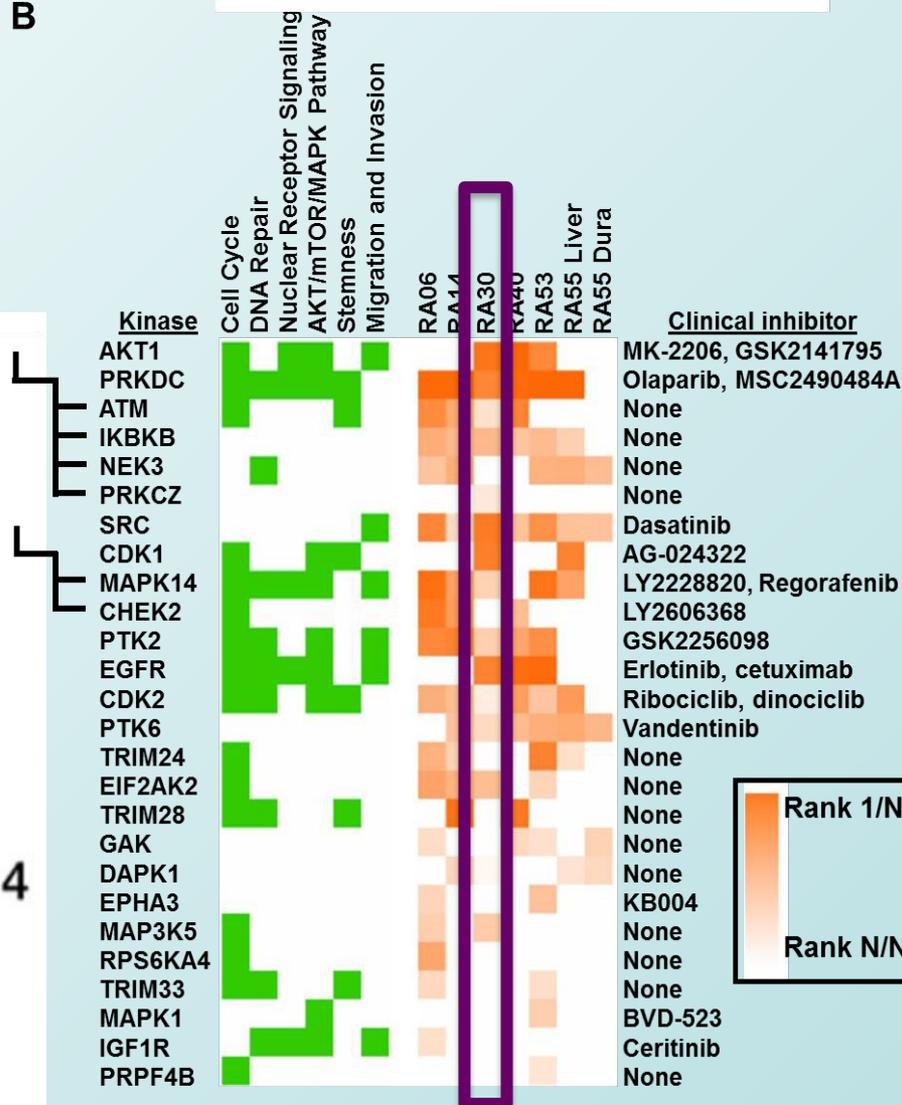


pCHIPs: Network-based selection of targets and target combinations for individual patients

Patient 30

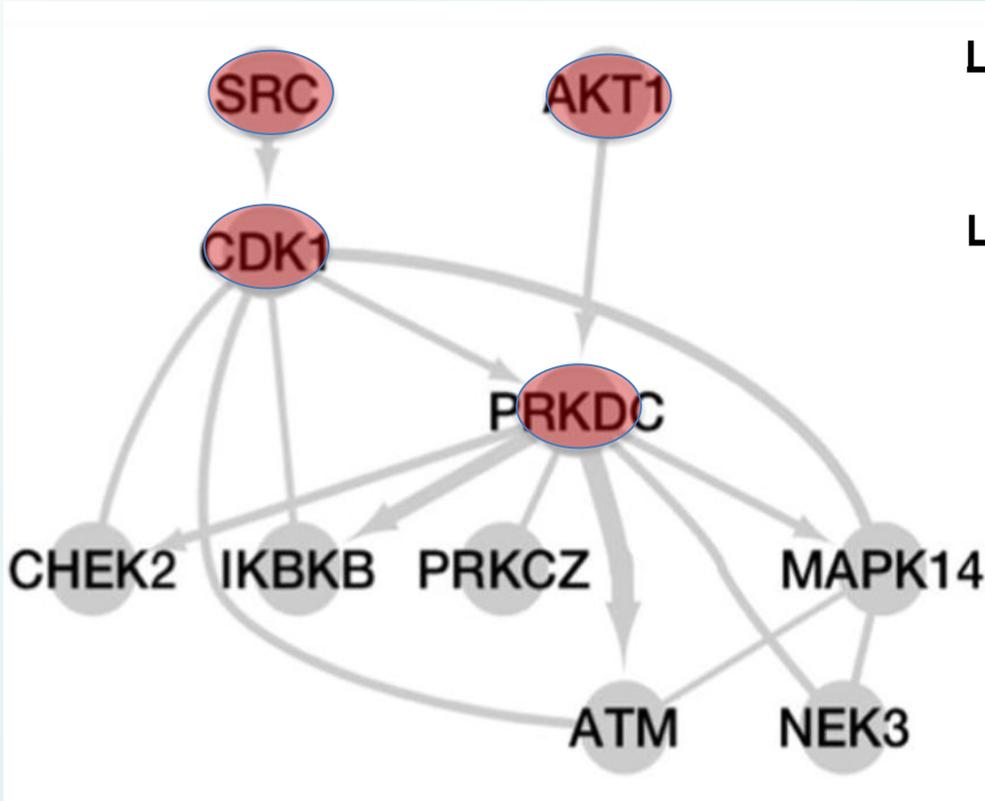


B

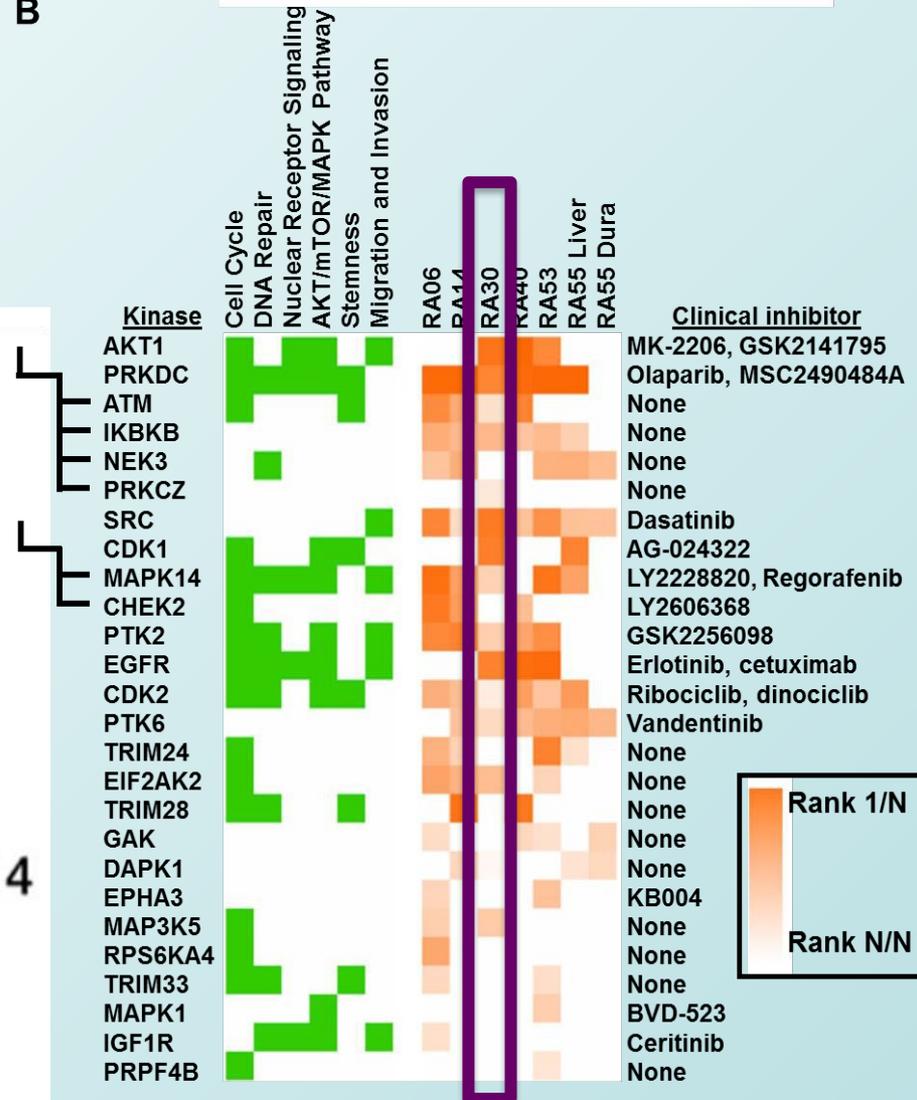


pCHIPs: Network-based selection of targets and target combinations for individual patients

Patient 30

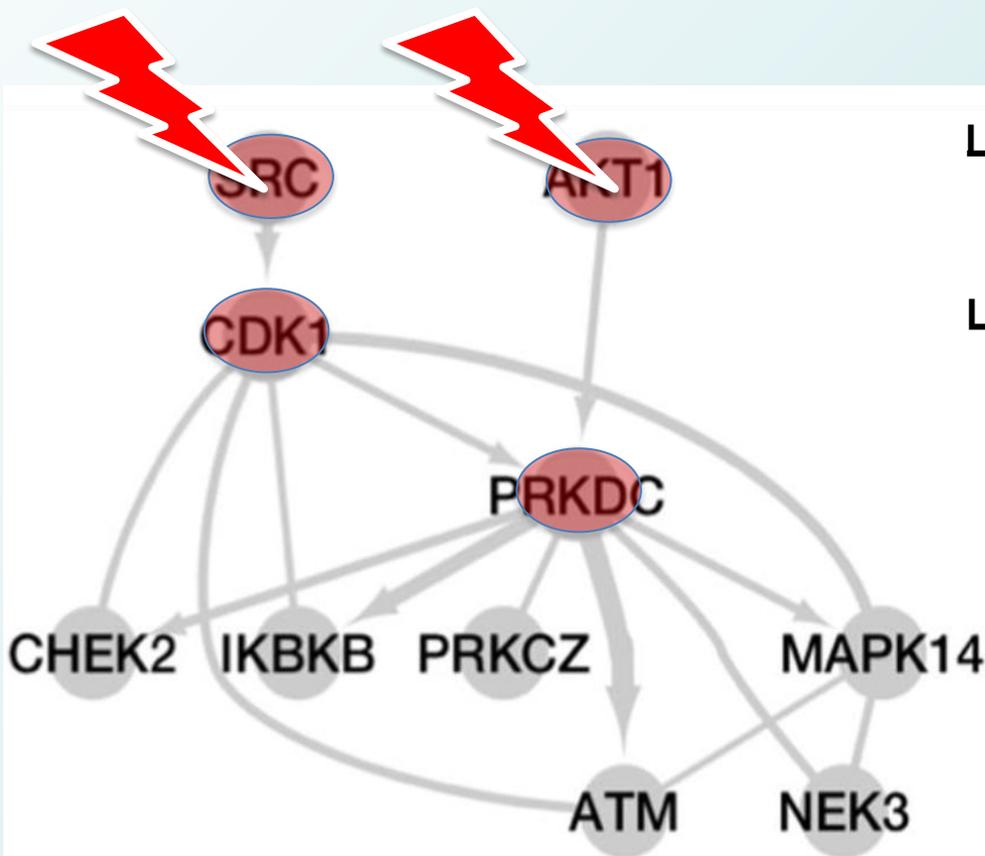


B

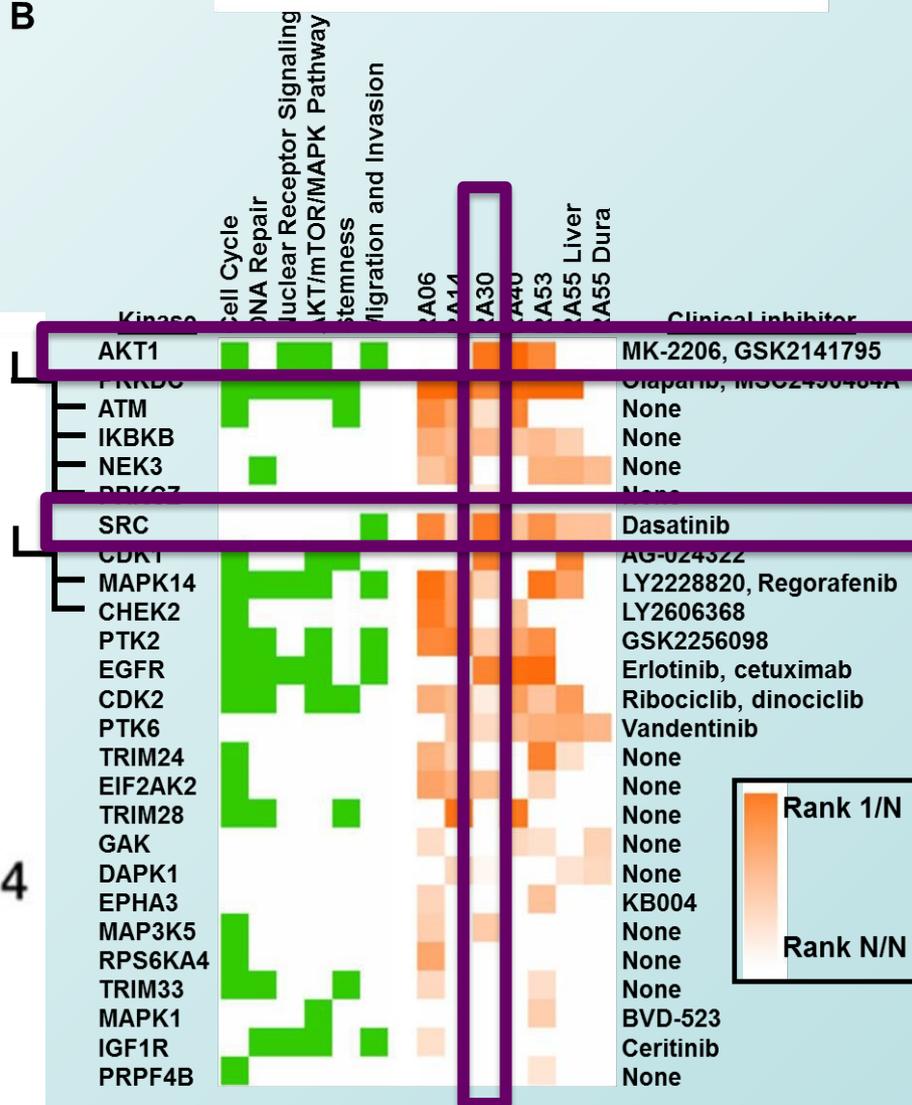


pCHIPs: Network-based selection of targets and target combinations for individual patients

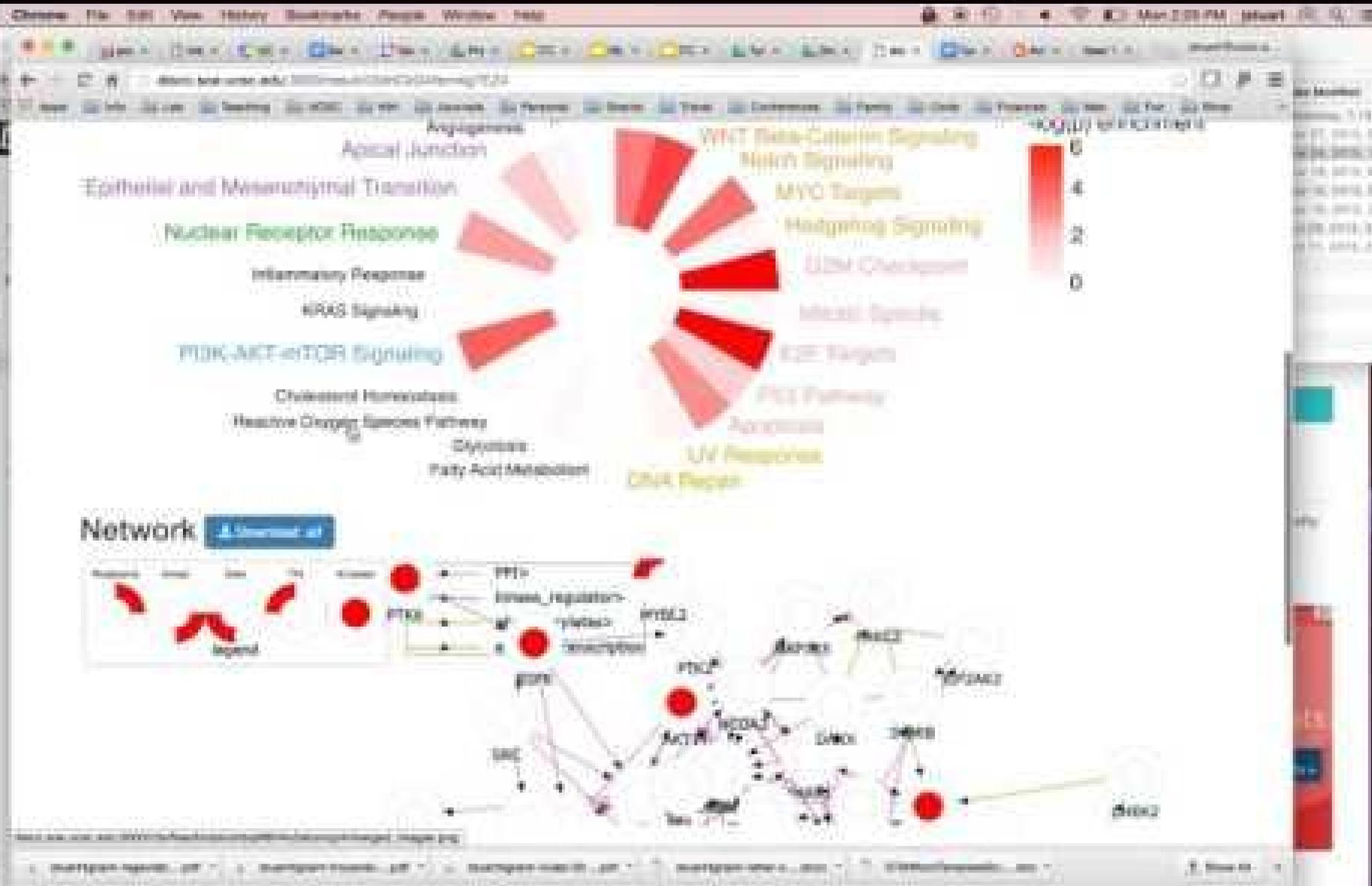
Patient 30



B



BMEG Portal of Patient-Specific Hallmarks (pCHIPS)



BMEG Progress Summary

- New DREAM brain-sourcing challenges produce re-runnable, cloud-ready code
 - Compare methods, reusable, rerunnable code
- Established calls for PanCan consortia.
 - Data & code adopted to GDC
- Graph of knowledge to capture high-level information
- Introspective inference methods
 - Granger causality-based gene network inference
 - ID patient-specific altered pathway diagrams (pCHIPS)

BMEG Year 5 Research Plan

- Implementing several use-cases
 - E.g. **tumor recommender**
 - Link TCGA tumors to predicted **drug sensitivities** (CCLE)
 - Connect **non- to coding variants** (from ICGC PCAWG)
 - Load high-level **imaging data**
- Distributed and **Automated Introspective Machine-Learning**
- Load **top DREAM results on SMC-Het & SMC-RNA**
- Ongoing Collaborative Projects
 - GDAN analyses - PanCanAtlas, Driver Discovery, ALCHEMIST, Exceptional Responders, LUSC survival
 - DREAM - Plan for **next RNA-Seq deconvolution challenge**

BMEG Team

UCSC Group

Chris Wong



David Haan



Pablo Cordero



Teresa Swatloski



James Durbin



Vlado Uzunangelov



OHSU Group

Kyle Ellrott



Ryan Spangler



Allison Creason



Alexander Buchanan



Malisa Smith



Jeena Lee





David Haussler

UCSC Genome Browser Staff UCSF / Buck Institute for Aging

- David Haussler
- Mark Diekins
- Melissa Cline
- Jorge Garcia
- Erich Weiler

Ted Goldstein



UCSF Medbook Team

- Ted Goldstein

UCSC Cancer Genomics

- Jing Zhu
- Sofie Salama
- Teresa Swatlowski
- Brian Craft

Jing Zhu



Olena Morozova



UCSC Tree House Project

- Olena Morozova
- Melissa Cline

Acknowledgments

Chris Benz,



- Chris Benz, Buck
- Christina Yau, Buck
- Denise Wolf, UCSF
- Laura van't Veer, UCSF
- Eric Collisson, UCSF

Collaborators

- Katie Hoadley, UNC
- Rehan Akbani, MDACC
- Justin Guinney, Sage
- Larsson Omberg, Sage

PCF

ORACLE



TREEHOUSE FOLKS

Olena Morozova, UC Santa Cruz
Atul Butte, CIAPM
Marco Marra, BC Cancer Agency
David Haussler, UC Santa Cruz
Ted Goldstein, UC Santa Cruz
Sofie Salama, UC Santa Cruz
Linda Rosewood, UC Santa Cruz
Ann Pace, UC Santa Cruz
Melissa Cline, UC Santa Cruz
Jing Zhu, UC Santa Cruz
Yulia Newton, UC Santa Cruz

Olena
Morozova



Yulia Newton



Clinical Collaborators

S. Rod Rassekh, BC Children's Hospital
Rebecca Deyell, BC Children's Hospital
Stephen Yip, BC Cancer Agency
Leonard Sender, Children's Hospital of Orange County
Sabine Mueller, UC San Francisco
Adam Resnick, Children's Hospital of Philadelphia
Alejandro Sweet-Cordero, Stanford



TCGA Pan-Can AWG

MDACC

John Weinstein
Rehan Akbani

Lauren Byers
Han Liang
Roel Verhaak
Gordon Mills



MSKCC

Chris Sander
Giovanni Ciriello
Anders Jacobsen

Baylor

Mark Hamilton
David Wheeler

USC

Peter Laird
Hui Shen



Broad

Andrew Cherniak
Matt Meyerson
Gaddy Getz
Rameen Beroukhim
Scott Carter
Travis Zack
Mike Lawrence
Angela Brooks



Sage Bionetworks

Adam Margolin
Larsson Ohmberg

Harvard

Raju Kucherlapati

ISB

Ilya Shmulevich
Sheila Reynolds



UCSC

Vlado Uzunangelov
Sam Ng
Evan O. Paull
Kyle Ellrott
David Haussler
Jing Zhu



UPF Barcelona

Nuria Lopez-Bigas
David Tamborero
Abel
Gonzalez-Perez

UNC

Katherine Hoadley
Chuck Perou



NCI / NHGRI

Zhong Chen
Carter Van Waes



WashU

Li Ding
Cyriac Kandoth
Beifang Liu
Mike McLellan

Brown

Ben Raphael
Max Leiserson



UBC

Gordon Robertson
Andy Chu



UCSF/Buck Inst

Chris Benz
Eric Collisson
Christina Yau
Denise Wolf

BMEG Overview Video → [[Play Youtube Video](#)]

Additional

Future Challenges

- Deconvolve tumors
 - Subclones, evolution, migration
 - ID all cells in tumor environment (new DREAM deconvolution challenge)
 - Leverage single cell data

- Translate omics to patients
 - Reproducible / reliable *patients-like-me*
 - Recommenders to identify relevant evidence for submitted tumor specimens
 - E.g. implicated drugs, activated pathways
 - End-to-end sequence to prediction (DNA & RNA)

- Contribute to early detection
 - Precursor mutations that lead to later stage mutations.

Quest for the best: not just a dream but a **DREAM** challenge



ICGC-TCGA (SMC) Challenge Series for Cancer

SMC-DNA

- Identify the best methods mutation calling methods (wrapping up)

SMC-RNA

- Identify the best methods for fusion gene finding from RNA-seq data
- (launching Jan, 2015)

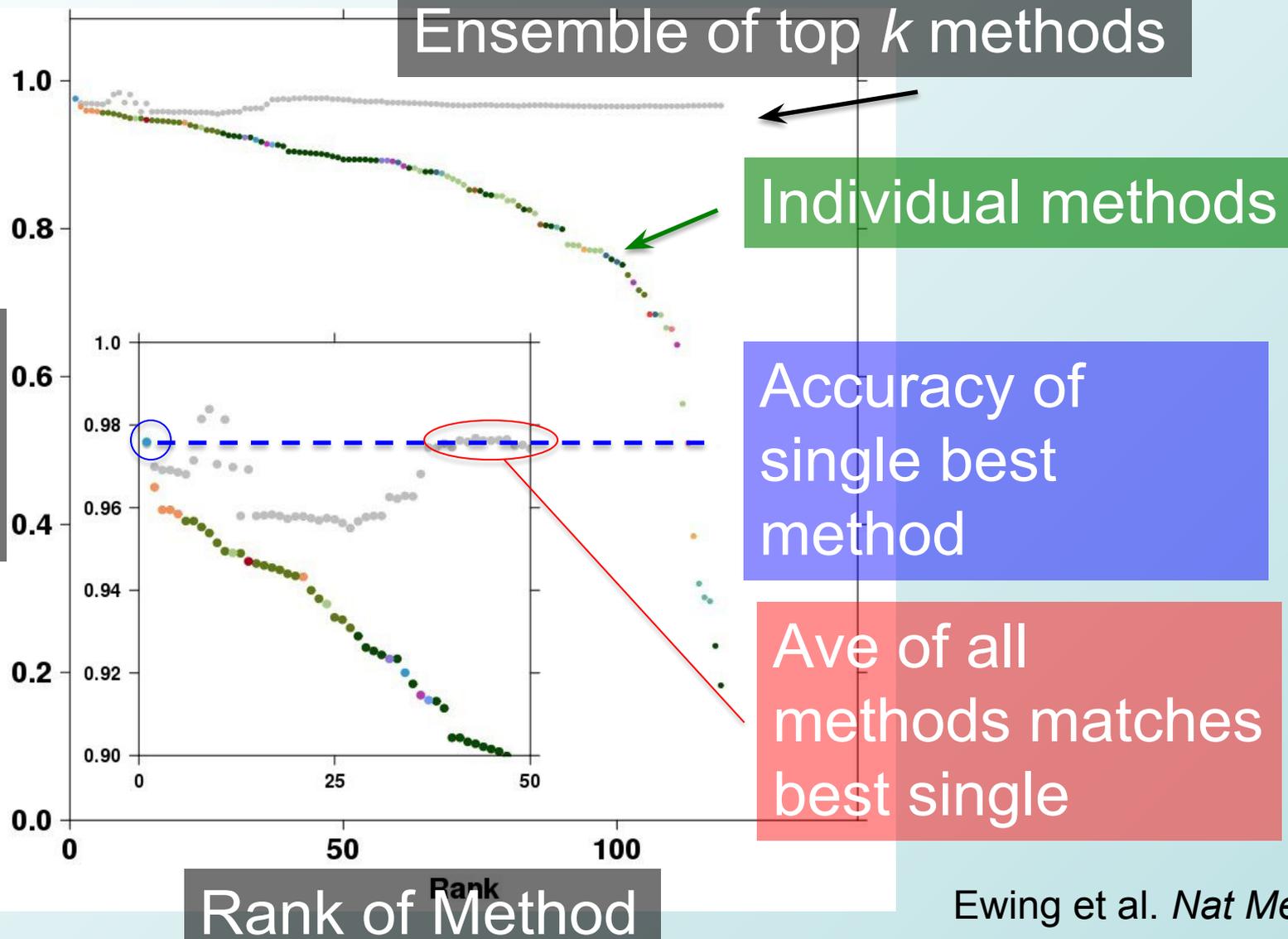
SMC-HET

- ID best methods for deconvolving subclonal heterogeneity from VAFs.
- Estimate sub-clone proportions
- Predict lineage tree
- (launching Jan 2015)

Results of DREAM-SMC

- Participation At Closing Time:
 - o **345** contestants
 - o 948 entries on 4 *in silico* genomes
- On-going post-challenge submissions (*living benchmark*)
- Key insights into simulating cancer genomes (BamSurgeon)

Wisdom of the Crowds for DREAM-SMC



Distinguish True Variation from Artifact

GTTACTGTCGTTGTAATACTCCAC**G**ATGTC
GTTACTGTCGTTGTAATACTCCACGATGTC
GTTACTGTCGTTGTAATACTCCACGATGTC
GTTACTGTCGTTGTAATACTCCAC**A**ATGTC
GTTACTGTCGTTGTAAT**g**CTCCACGATGTC
GTTACTGTCGTTGTAATACTCCAC**A**ATGTC
GTTACTGTCGTTGTAATACTCCACGATGTC
GTTACTGTCGT**G**GTAATACTCCAC**a**ATGTC
GTTACTGTCGTTGTAATACTCCAC**a**ATGTC
GTTA**a**TGTCGTTGTAATACTCCACGATGTC
GTTACTGTCGTTGTA**c**TACTCCACGATGTC
GTTACTGTCGTTGTAATACTCCAC**a**ATGTC



sequencing errors

SNV

A needle in a human genome ~~haystack~~ ^{needlestack}

- A human genome has **23** chromosomes.
- **6 billion** individual DNA basepairs per genome.
- A **single basepair** error can be a disease mutation.



BMEG's DREAM-SMC Paved the Way for MC3

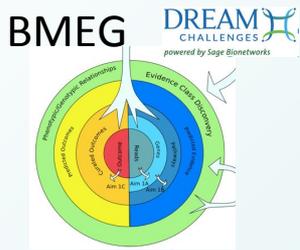
Method Name	Team	Sensitivity	Precision	Balanced Accuracy
MuTect - L10	Broad SMC	0.9672	0.9836	0.9754
Dream_Set1_MuSE_Setting8	Wang-Wheeler-HG SC	0.9208	0.9778	0.9493
Strelka_aln-0.5.7_recSNV_filters_default	SMC_Admins	0.8325	0.953	0.8928
somaticsniper_bwa_2	SMC_Admins	0.9341	0.8317	0.8829
radia	UCSC	0.884	0.8815	0.8828
varscan-default-params	SMC_Admins	0.9859	0.1206	0.5532

Preprocessing of winning method (GATK cleaning)
adopted for all MC3 methods

All MC3 callers (and more) evaluated in challenge

Methods containerized after challenge completion

Overview of BMEG Pipelines



Data Sources

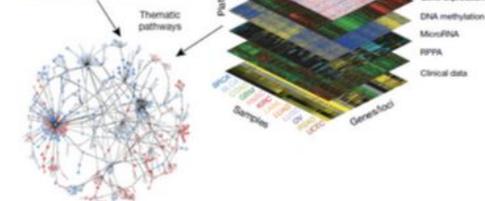
TCGA/ICGC/
SU2C / CKCC

Top Methods

Pathways



High-Level Genomics Data



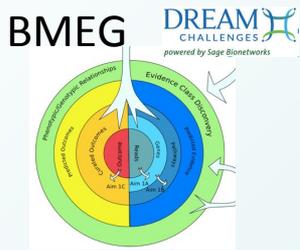
Cell Lines

Drug

Sensitivity



Overview of BMEG Pipelines



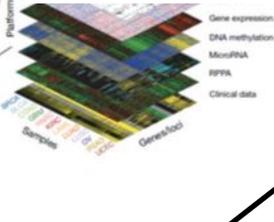
Data Sources

TCGA/ICGC/
SU2C / CKCC

Pathways



High-Level Genomics Data



Cell Lines

Drug Sensitivity



Top Methods

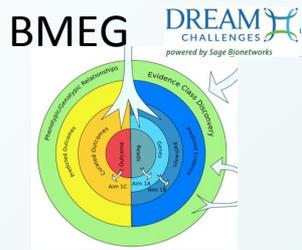


Outcomes / Phenotype Inferences

Aggressive

Gene Essentiality

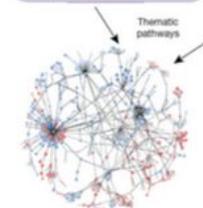
Overview of BMEG Pipelines



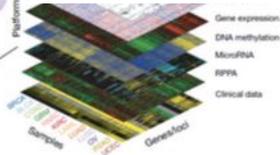
Data Sources

TCGA/ICGC/
SU2C / CKCC

Pathways

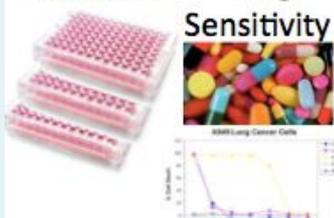


High-Level Genomics Data



Cell Lines

Drug Sensitivity



Top Methods



Outcomes / Phenotype Inferences

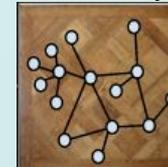
Aggressive

Gene Essentiality

Predictive Signature DB



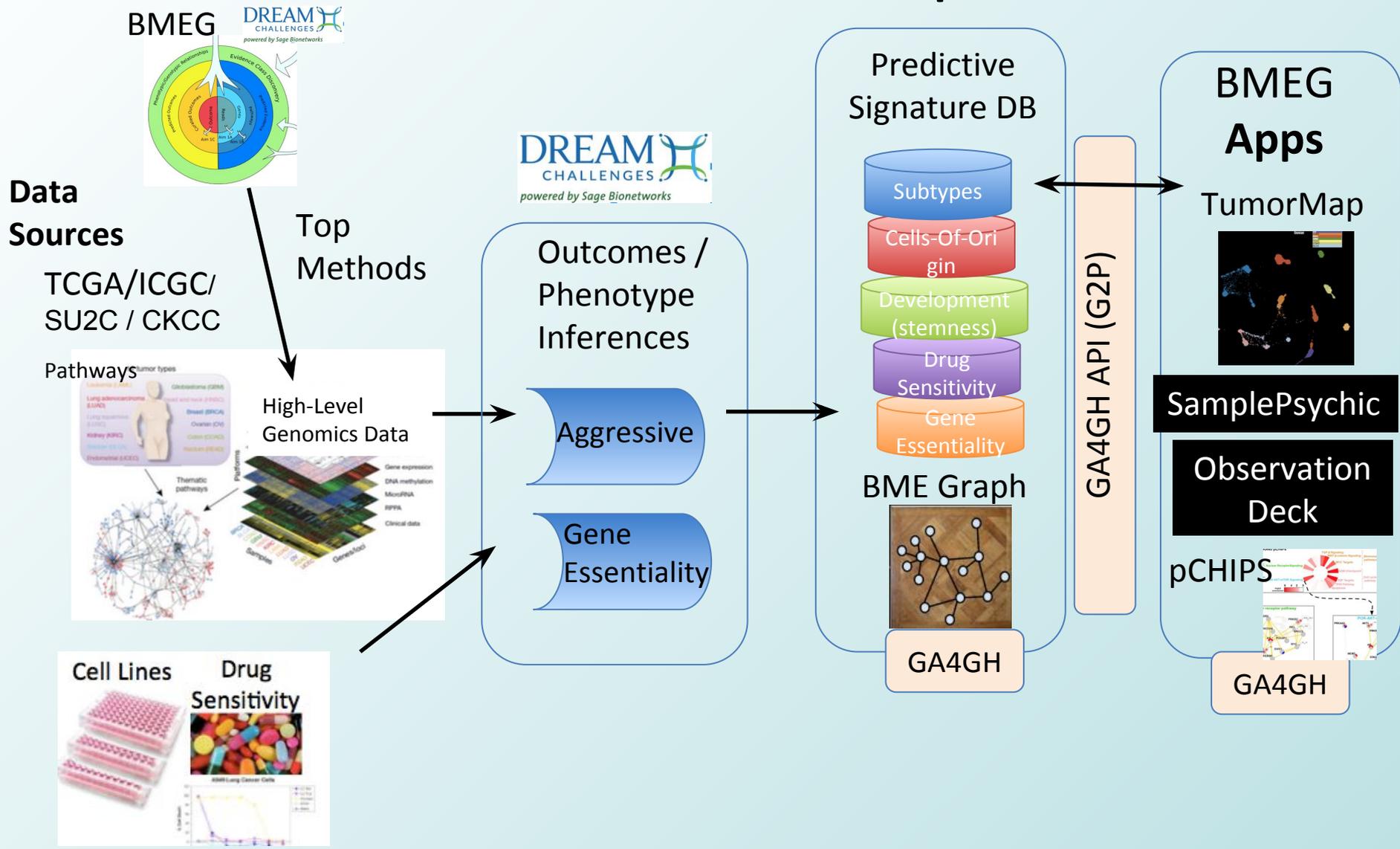
BME Graph



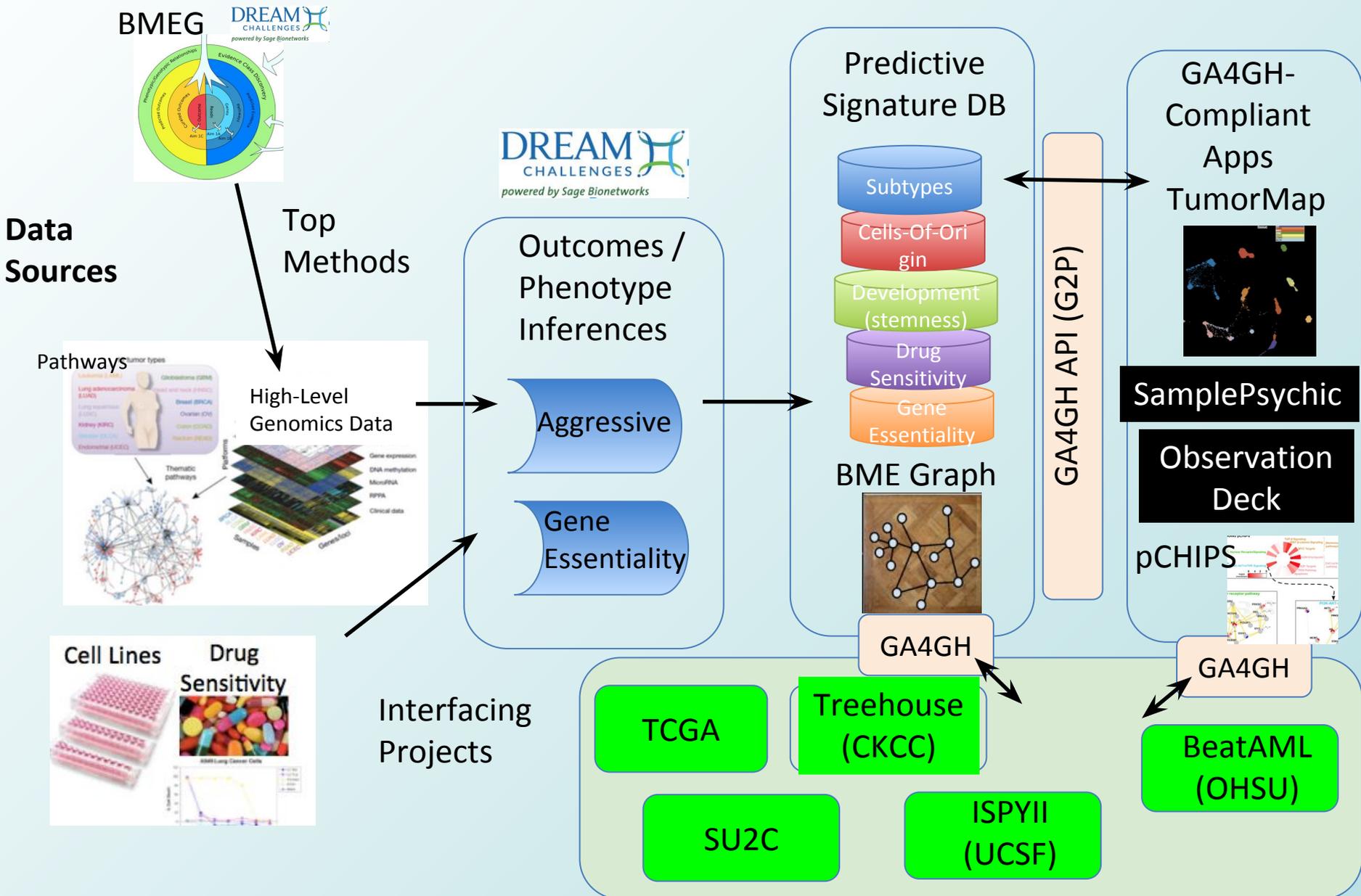
GA4GH

GA4GH API (G2P)

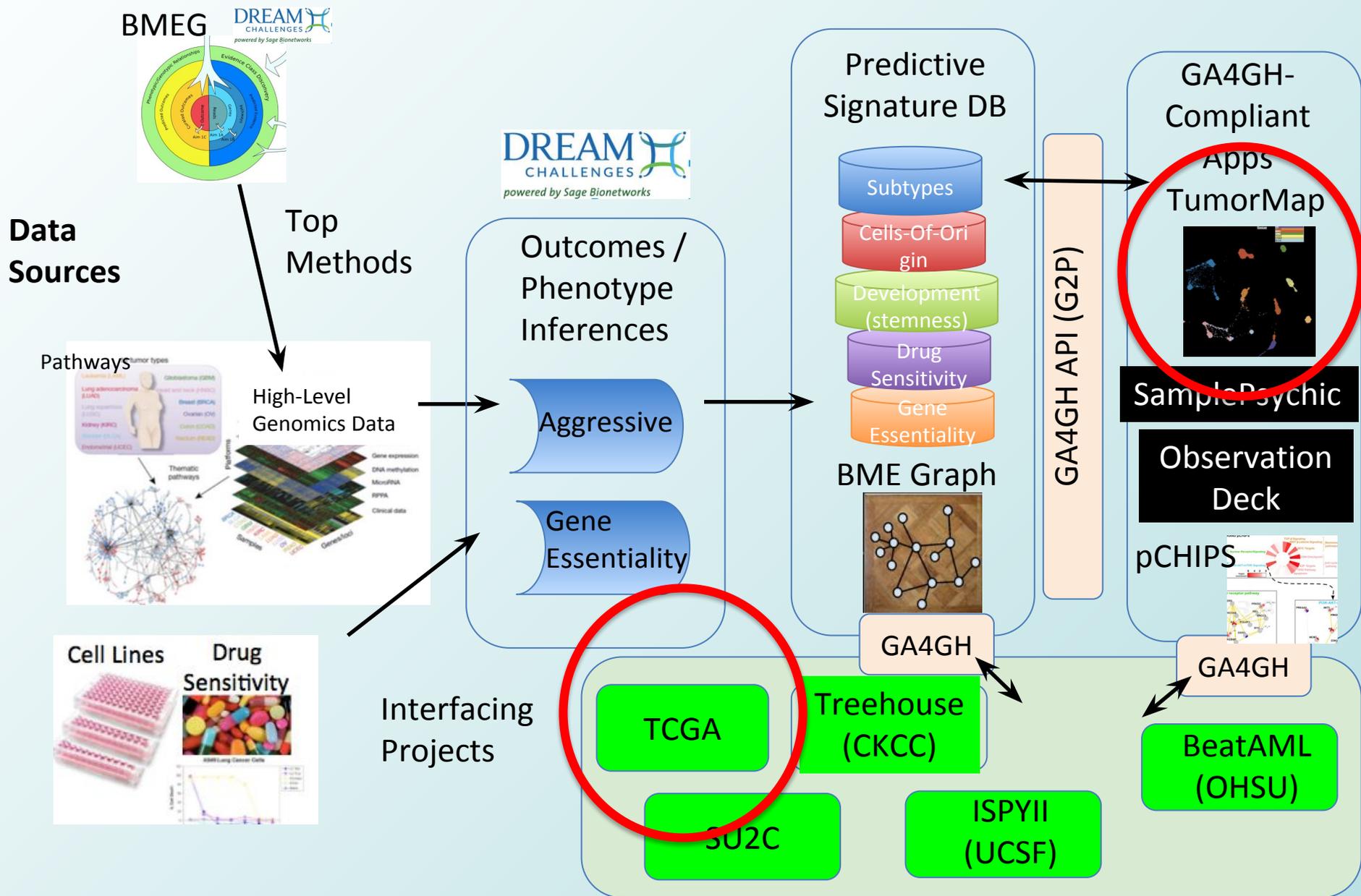
Overview of BMEG Pipelines



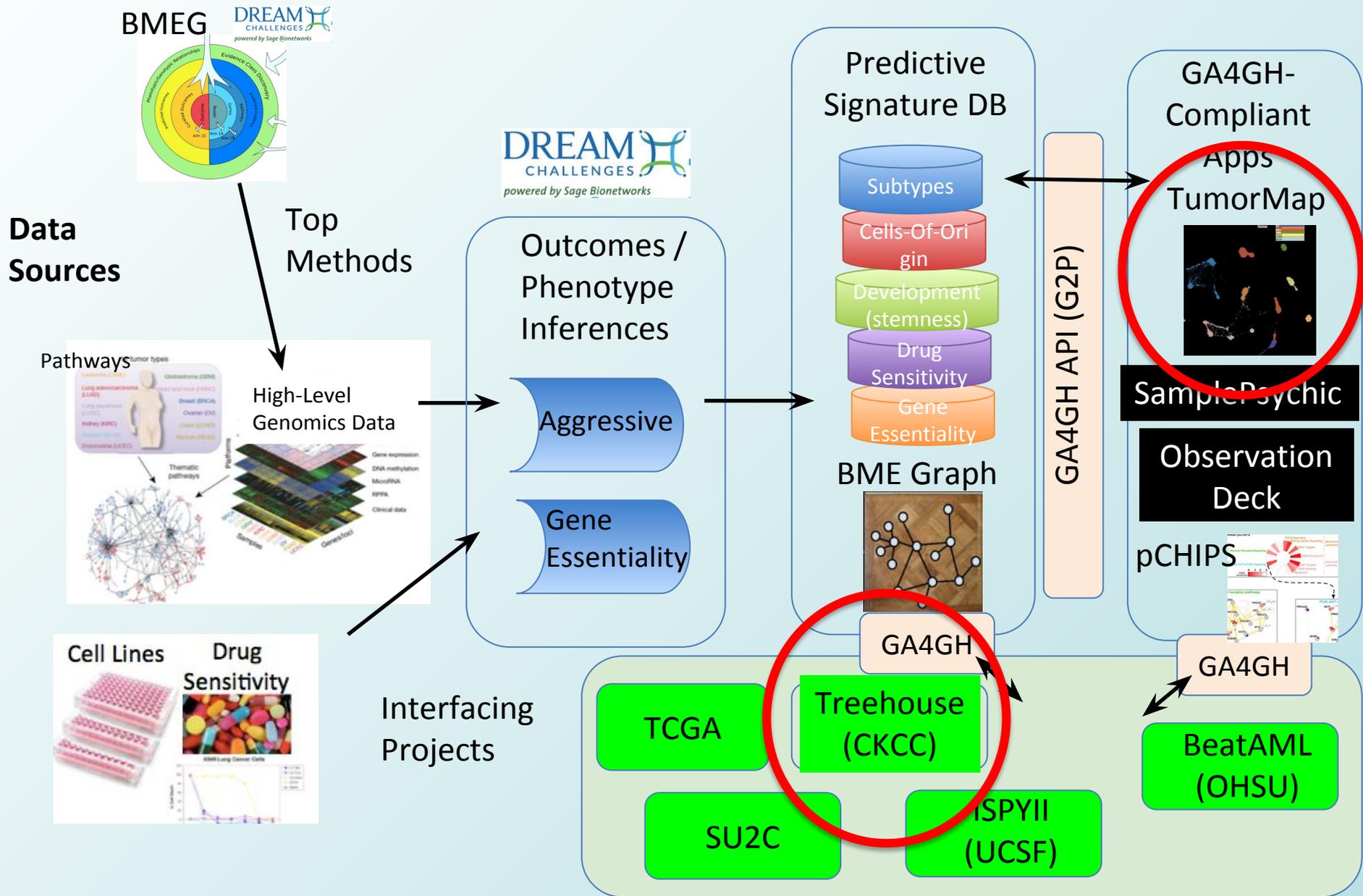
Overview of BMEG Pipelines



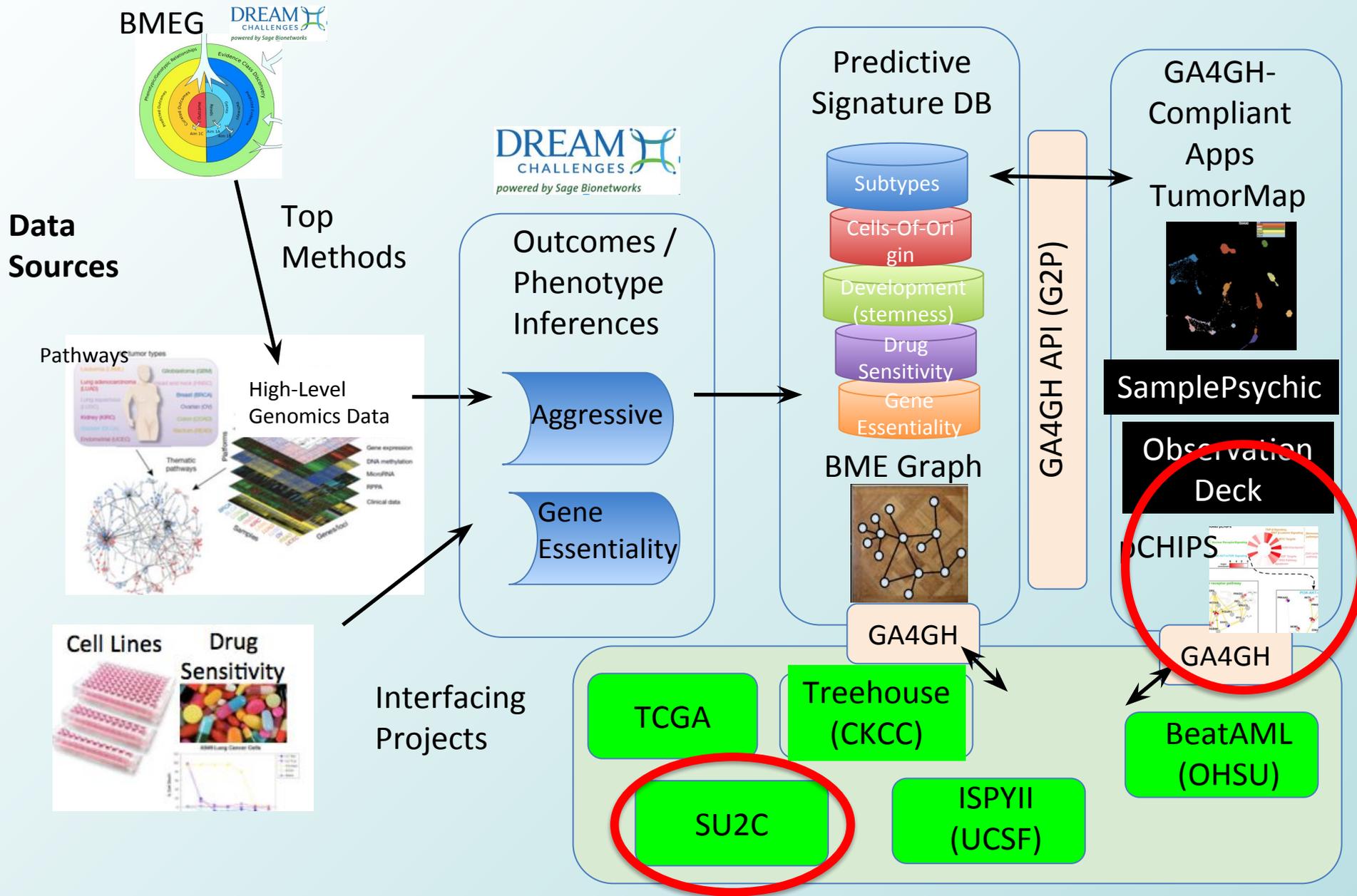
Overview of BMEG Pipelines



Overview of BMEG Pipelines



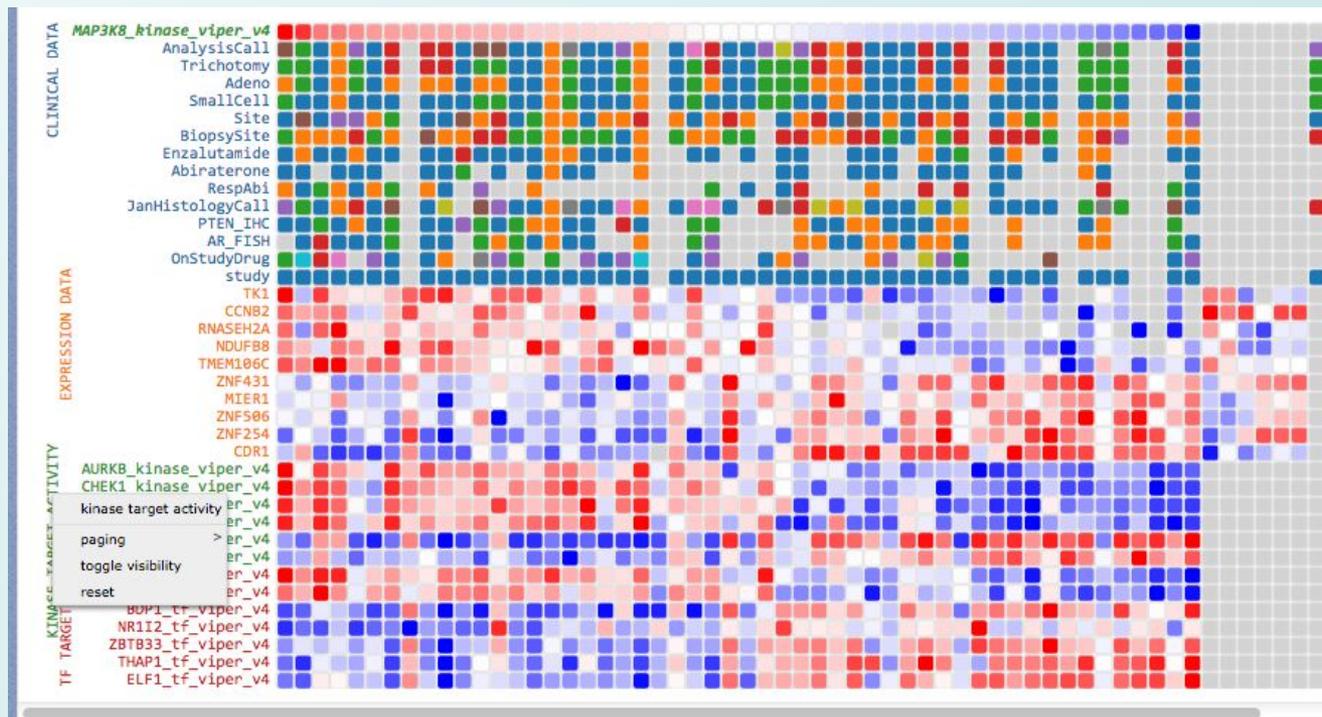
Overview of BMEG Pipelines



Demo: DNA Damage Genes -> Gemcitabine Sensitivity

Link: [MOV File on Google](#)

BMEG ObservationDeck



Demo: Query the TumorMap

Link: [MOV File on Google](#)

Demo: Query pCHIPS

Link: [MOV File on Google](#)

<https://www.youtube.com/watch?v=WU3yd5zjnt8>

