*AMARETTO for network biology and medicine:

linking diseases, drivers, targets and drugs

via multi-omics, clinical, imaging and perturbation data fusion

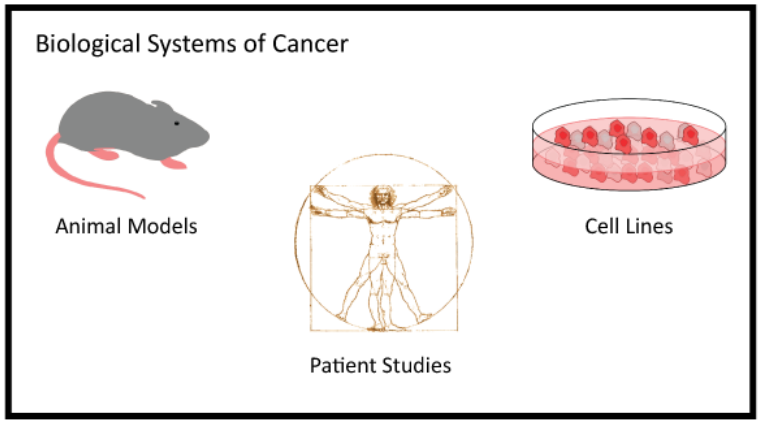
Nathalie Pochet, Ph.D.
Assistant Professor, Harvard Medical School
Associate Scientist, Brigham and Women's Hospital
Associate Member, Broad Institute of MIT and Harvard

http://portals.broadinstitute.org/pochetlab/
http://portals.broadinstitute.org/pochetlab/amaretto.html
npochet@broadinstitute.org, npochet@bwh.harvard.edu

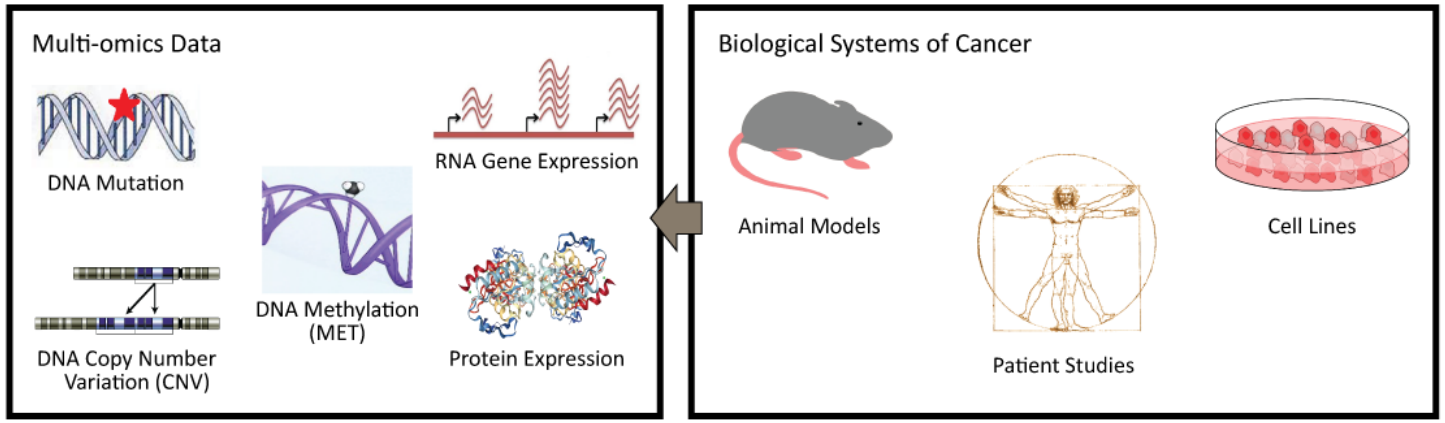
Funded by NIH NCI ITCR

The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer

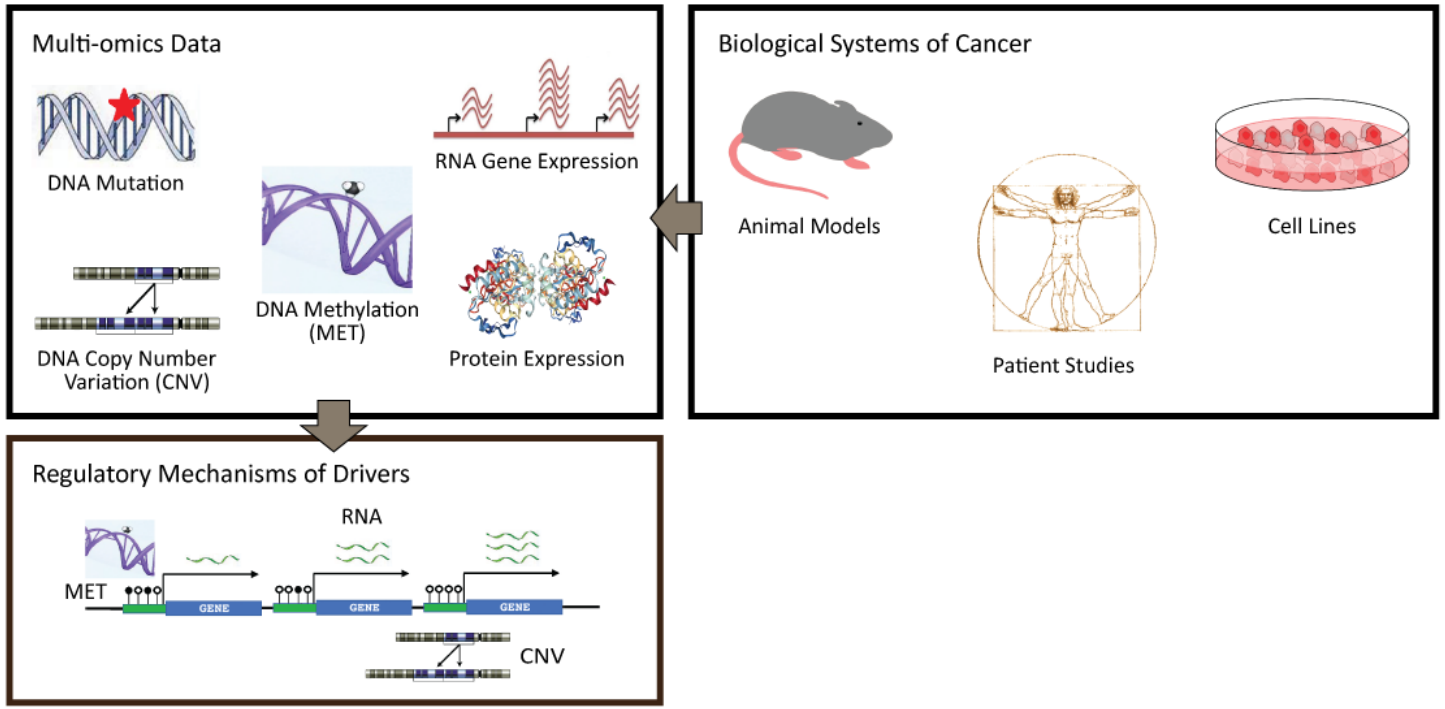
The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer



The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer



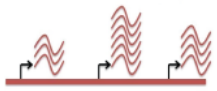
The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer



The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer

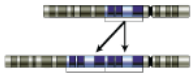
Multi-omics Data

DNA Mutation

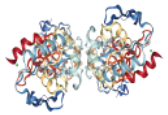


RNA Gene Expression

DNA Methylation (MET)



DNA Copy Number Variation (CNV)

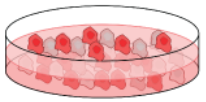


Protein Expression

Biological Systems of Cancer

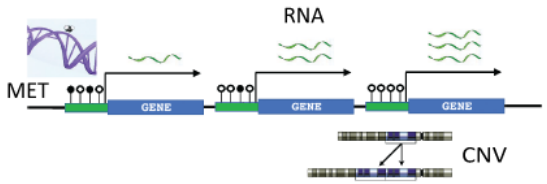
Animal Models

Patient Studies

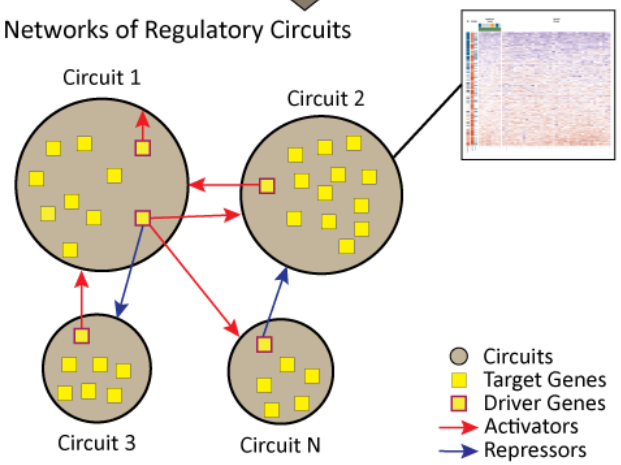


Cell Lines

Regulatory Mechanisms of Drivers

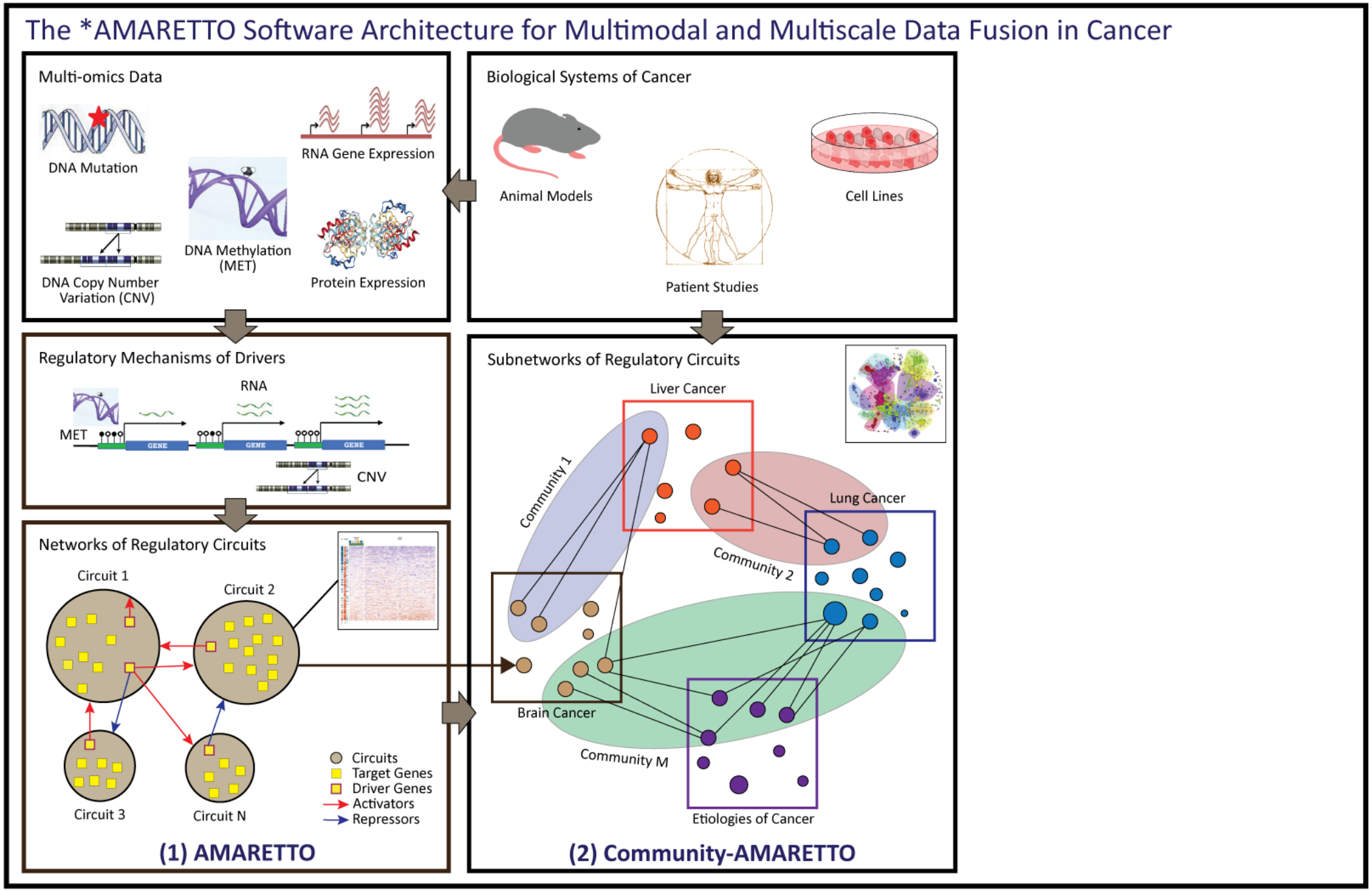


Networks of Regulatory Circuits



(1) AMARETTO

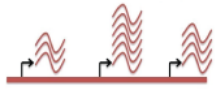
The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer



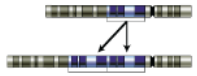
The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer

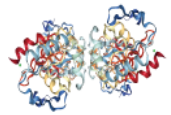
Multi-omics Data

DNA Mutation


RNA Gene Expression

DNA Methylation (MET)

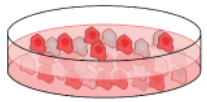

DNA Copy Number Variation (CNV)


Protein Expression

Biological Systems of Cancer

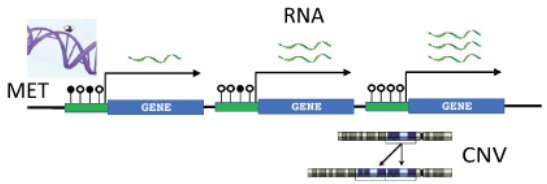
Animal Models

Patient Studies

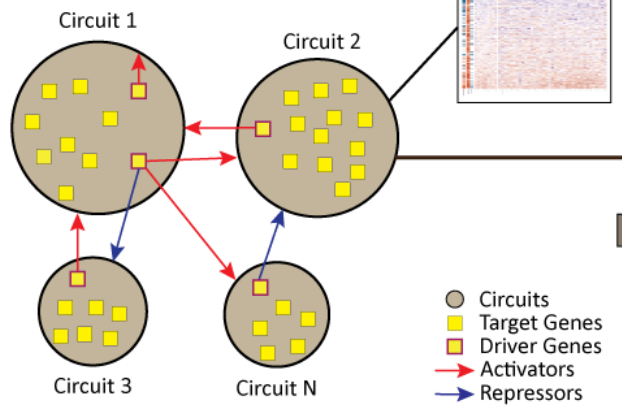

Cell Lines

Clinical and Experimental Outcomes

Regulatory Mechanisms of Drivers

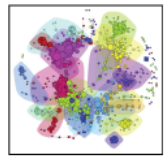


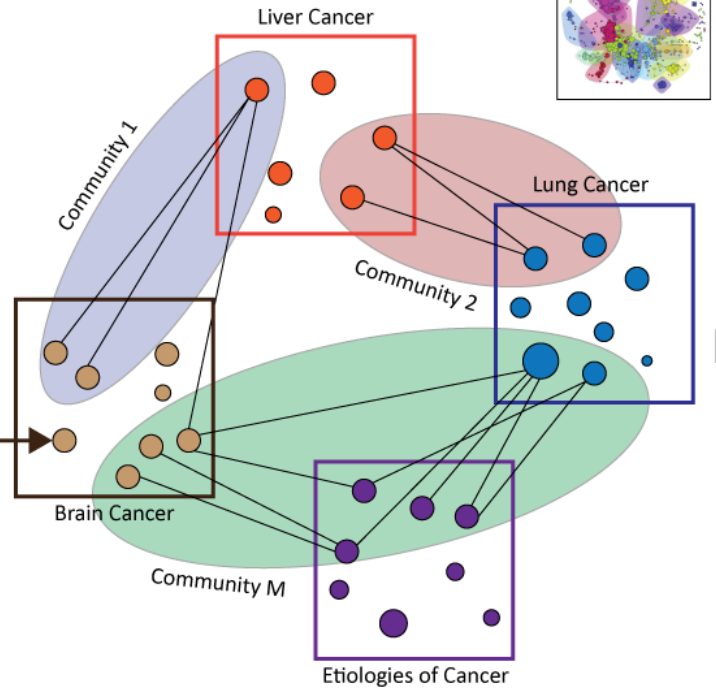
Networks of Regulatory Circuits



(1) AMARETTO

Subnetworks of Regulatory Circuits



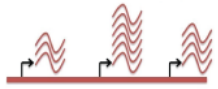


(2) Community-AMARETTO

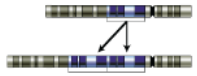
The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer

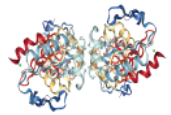
Multi-omics Data

DNA Mutation


RNA Gene Expression

DNA Methylation (MET)

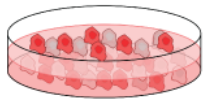

DNA Copy Number Variation (CNV)


Protein Expression

Biological Systems of Cancer

Animal Models

Patient Studies

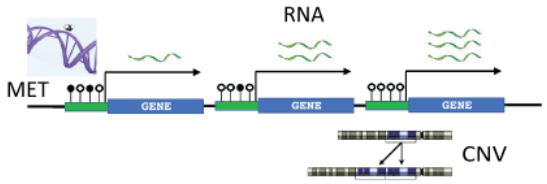

Cell Lines

Clinical and Experimental Outcomes

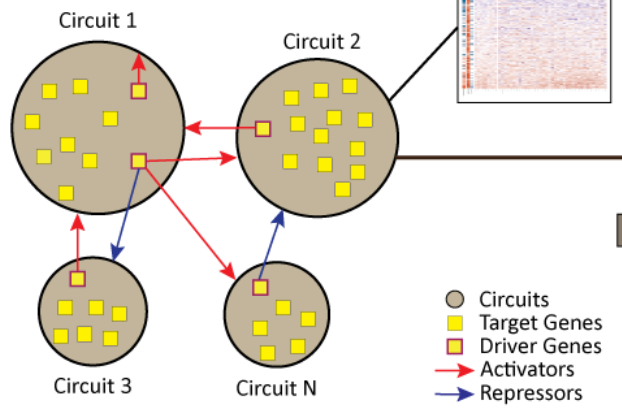
Clinical Disease Biomarker Discovery

Molecular Disease Biomarker Discovery

Regulatory Mechanisms of Drivers

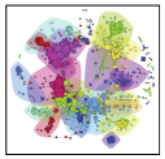


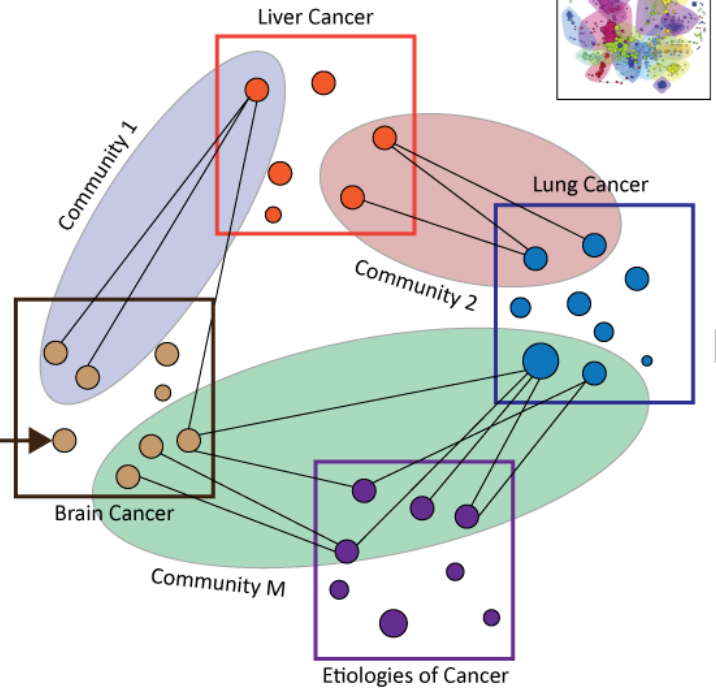
Networks of Regulatory Circuits



(1) AMARETTO

Subnetworks of Regulatory Circuits



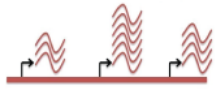


(2) Community-AMARETTO

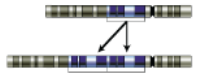
The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer

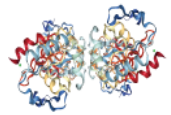
Multi-omics Data

DNA Mutation


RNA Gene Expression

DNA Methylation (MET)

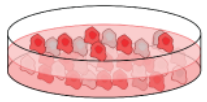

DNA Copy Number Variation (CNV)


Protein Expression

Biological Systems of Cancer

Animal Models

Patient Studies

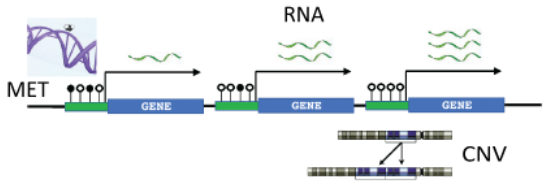

Cell Lines

Clinical and Experimental Outcomes

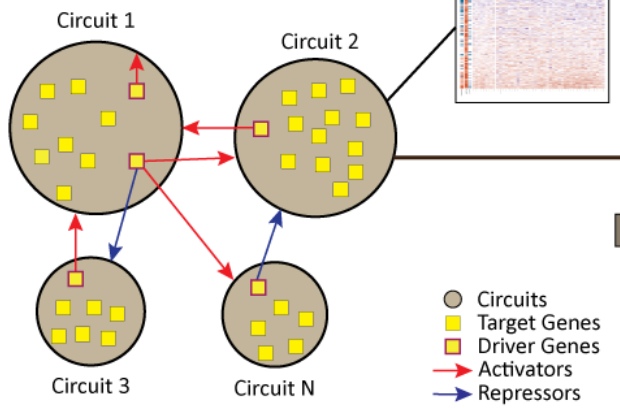
Clinical Disease Biomarker Discovery

Molecular Disease Biomarker Discovery

Regulatory Mechanisms of Drivers

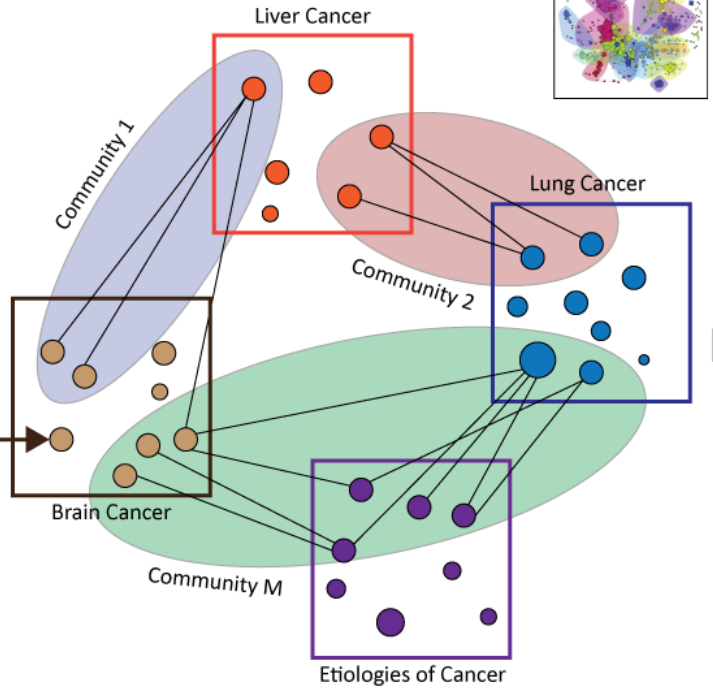


Networks of Regulatory Circuits



(1) AMARETTO

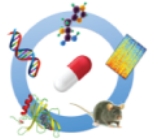
Subnetworks of Regulatory Circuits



(2) Community-AMARETTO

Genetic Perturbations for Driver Discovery

Chemical Perturbations for Drug Discovery

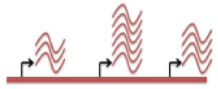


(3) Perturbation-AMARETTO

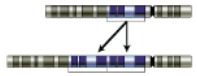
The *AMARETTO Software Architecture for Multimodal and Multiscale Data Fusion in Cancer

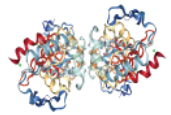
Multi-omics Data

DNA Mutation


RNA Gene Expression

DNA Methylation (MET)

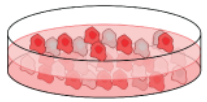

DNA Copy Number Variation (CNV)


Protein Expression

Biological Systems of Cancer

Animal Models

Patient Studies

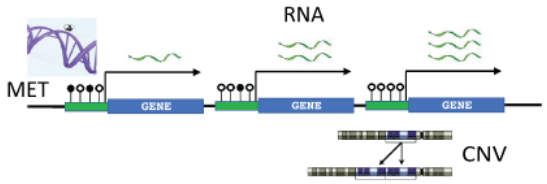

Cell Lines

Clinical and Experimental Outcomes

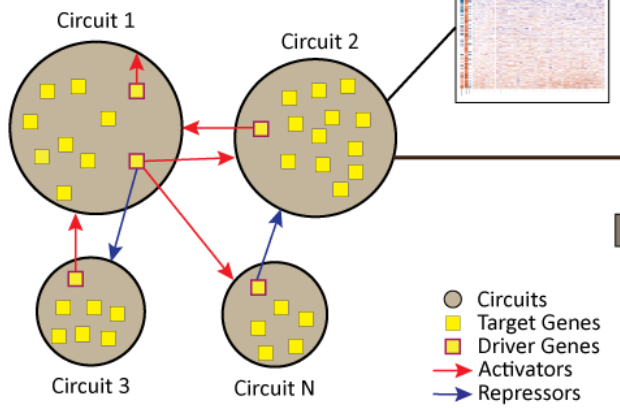
Clinical Disease Biomarker Discovery

Molecular Disease Biomarker Discovery

Regulatory Mechanisms of Drivers

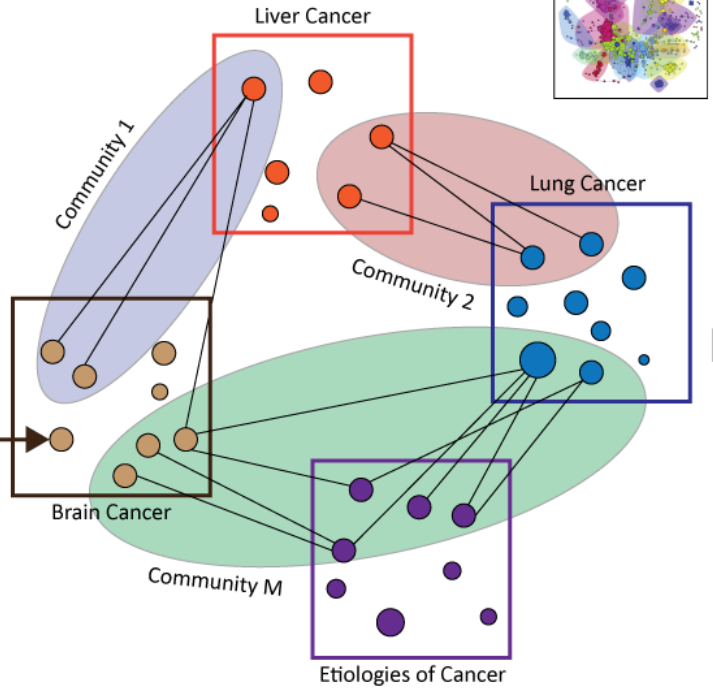


Networks of Regulatory Circuits



(1) AMARETTO

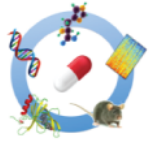
Subnetworks of Regulatory Circuits



(2) Community-AMARETTO

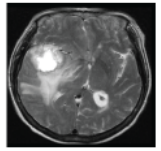
Perturbation-AMARETTO

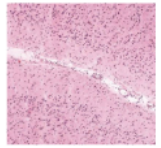
Genetic Perturbations for Driver Discovery


Chemical Perturbations for Drug Discovery

(3) Perturbation-AMARETTO

Imaging-AMARETTO


Radiography Imaging Biomarker Discovery


Histopathology Imaging Biomarker Discovery

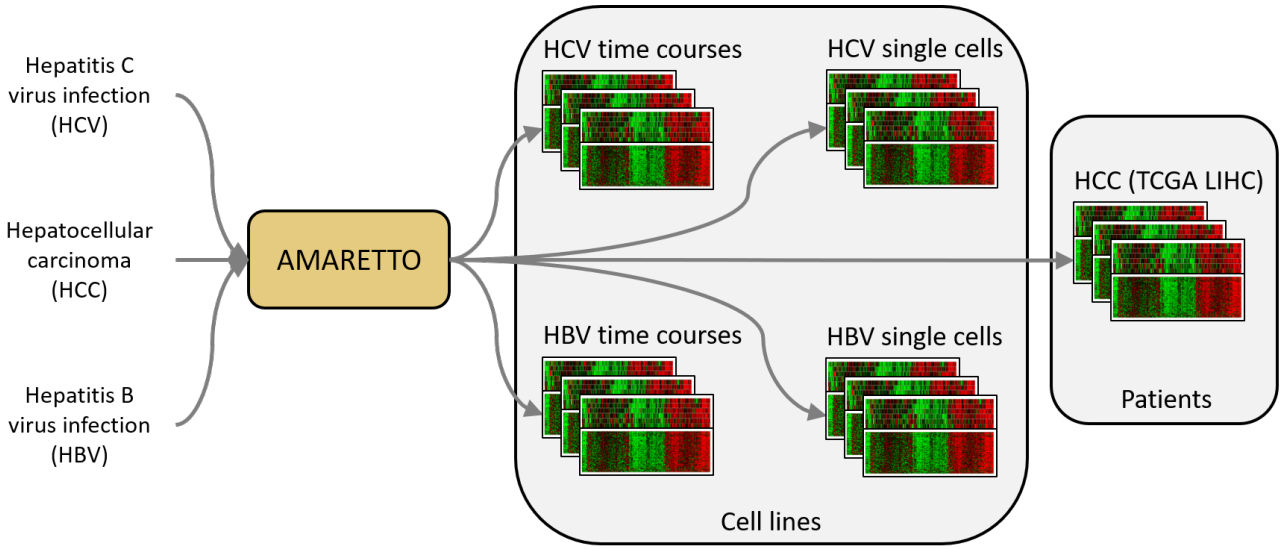
(4) Imaging-AMARETTO

Drug discovery:

Chemoprevention of hepatitis C and B virus-induced hepatocellular carcinoma

Drug discovery:

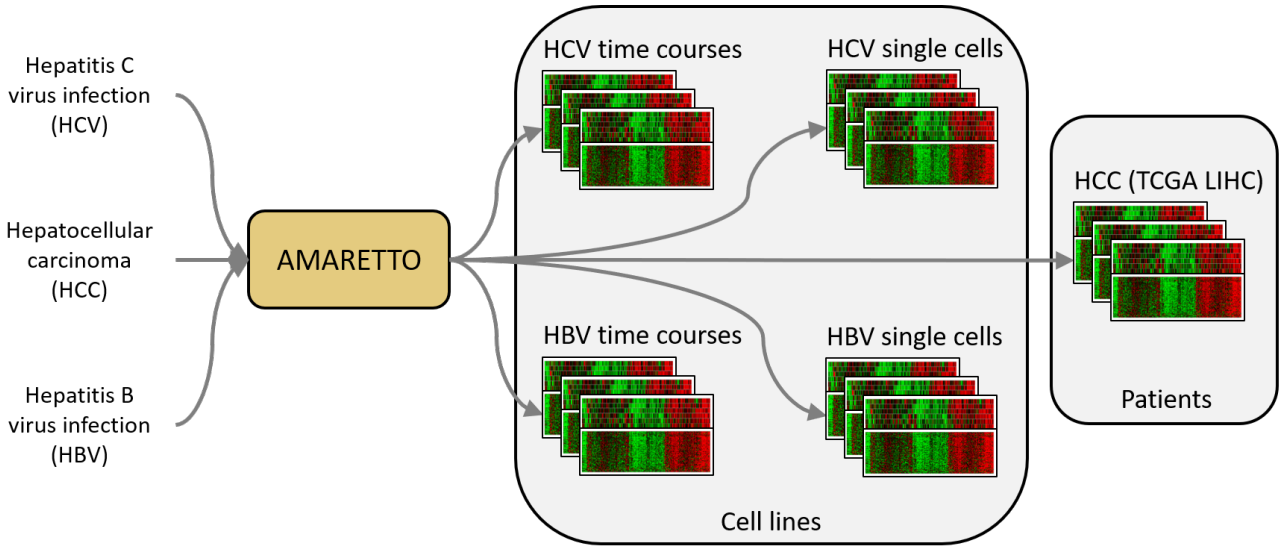
Chemoprevention of hepatitis C and B virus-induced hepatocellular carcinoma



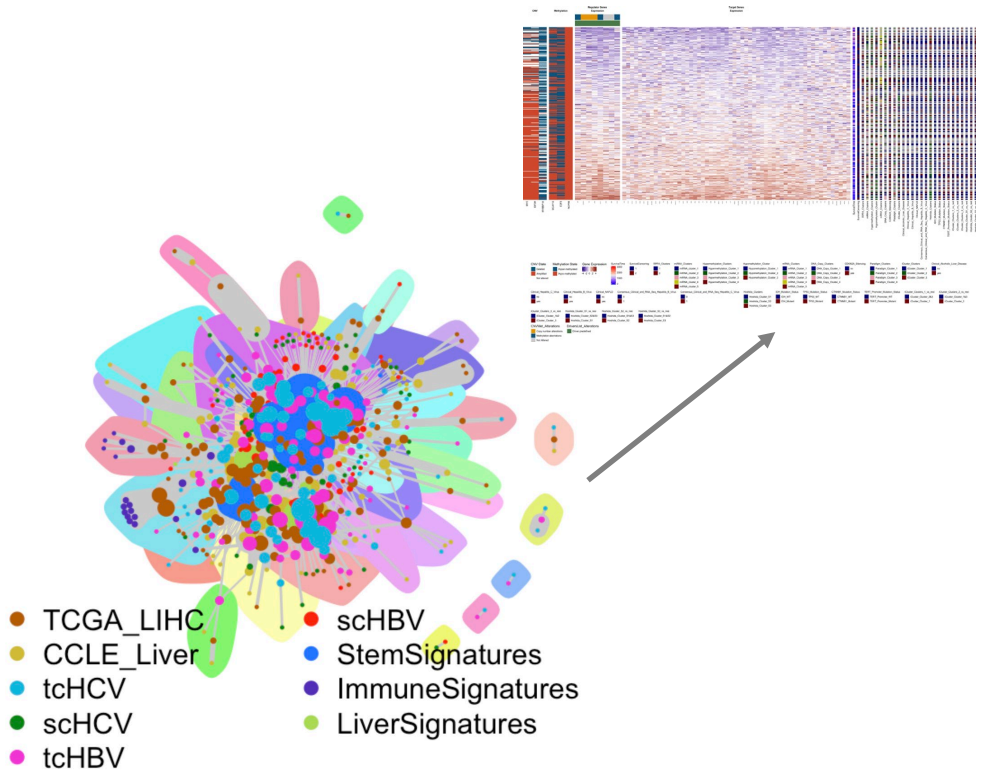
Driver and drug discovery
for hepatitis C (HCV) and hepatitis B (HBV)
virus-induced hepatocellular carcinoma (HCC)

Drug discovery:

Chemoprevention of hepatitis C and B virus-induced hepatocellular carcinoma



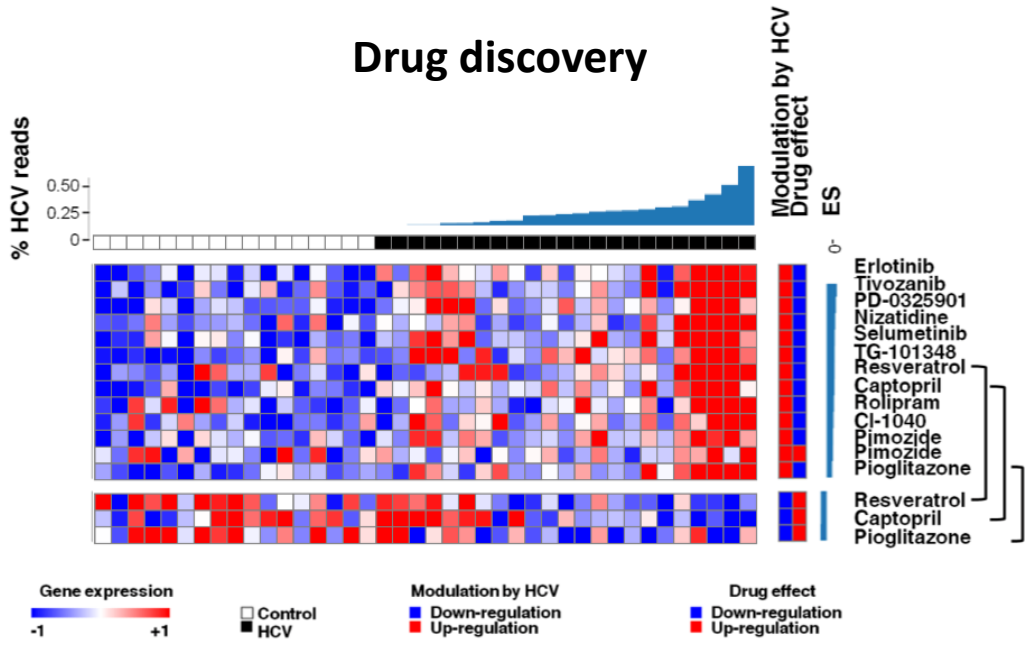
Driver and drug discovery
for hepatitis C (HCV) and hepatitis B (HBV)
virus-induced hepatocellular carcinoma (HCC)



Pan-etiology of liver cancer
networks and subnetworks
of regulatory circuits

Drug discovery:

Chemoprevention of hepatitis C and B virus-induced hepatocellular carcinoma



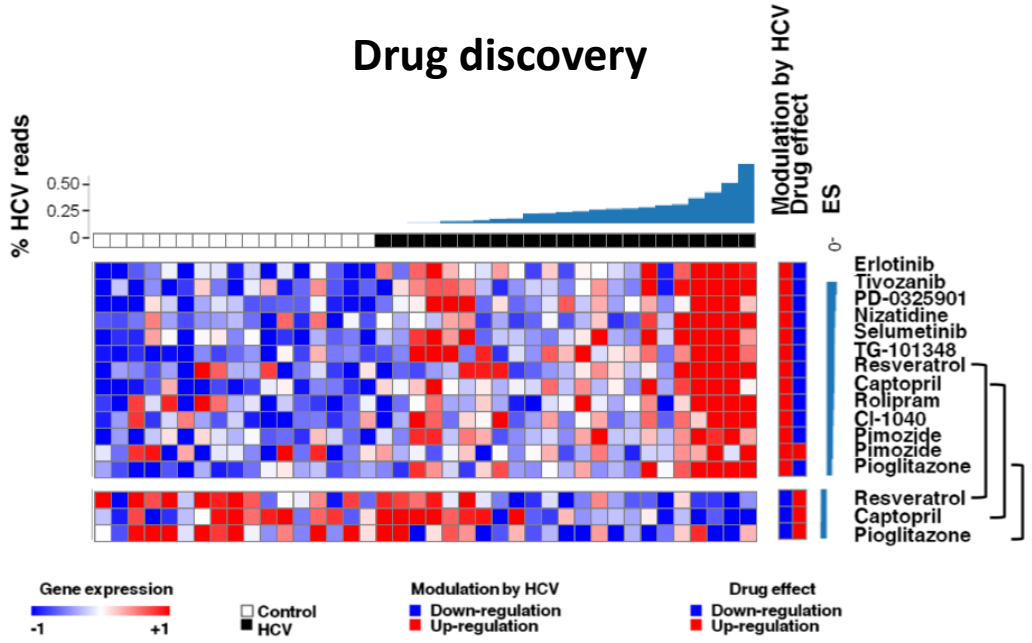
Chemical perturbations in cell lines

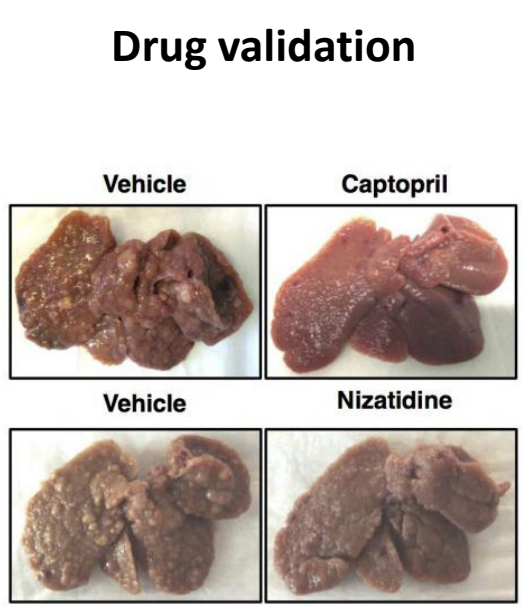
Predict which drugs can reverse disease-associated circuits

Alternative treatments with less severe adverse effects

Drug discovery:

Chemoprevention of hepatitis C and B virus-induced hepatocellular carcinoma





Chemical perturbations in cell lines
Predict which drugs can reverse disease-associated circuits
Alternative treatments with less severe adverse effects

Experimental validation of drugs in rat models
⇒ Two novel compounds attenuate HCC development
⇒ Safe and low-cost approach for chemoprevention of HCC?

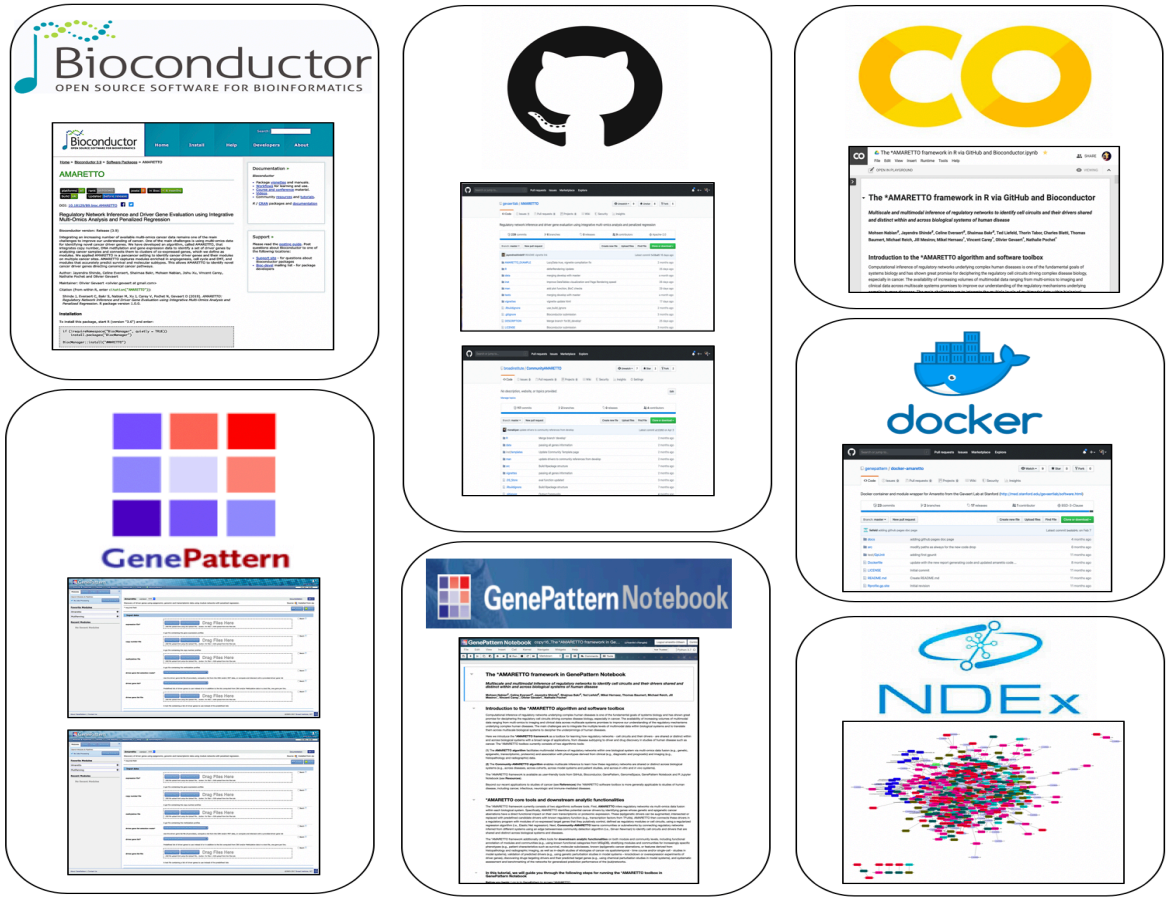
*AMARETTO source code, tools & notebooks

*AMARETTO is available via:

- Bioconductor: https://bioconductor.org/packages/release/bioc/html/AMARETTO.html
- GitHub
- Jupyter Notebook: https://colab.research.google.com/drive/1JfnRoNgTVX_7VEGAAmjGjwP_yX2tdDxs
- GenePattern
- GenomeSpace
- GenePattern Notebook
- …

Automated reports with results:

- Html reports: tables, heatmaps, graphs
- Neo4j embedded Shiny reports: interactive querying of graphs and tables



Tools and resources: http://portals.broadinstitute.org/pochetlab/amaretto.html

pochetlab.shinyapps.io/AMARETTO-Hub_TCGA-GBM/

AMARETTO-Hub: TCGA-GBM

Generate

Driver and Target Genes:

THBS1 MAP2

Show all Driver Genes

Regulatory Modules:

Functional Characterization:

Search: Hallmark

Select:

Show: None

P-value 0.05

FDR Q-value 0.05

Clinical Characterization:

Clinical or Molecular Phenotypes:

Radiography or Histopathology Imaging:

Show: Other Clinical Char.

P-value 0.05

FDR Q-value 0.05

Transcription Factor Discovery:

- show none
- show all transcription factors validated
- show all genes validated

P-value 0.05

FDR Q-value 0.05

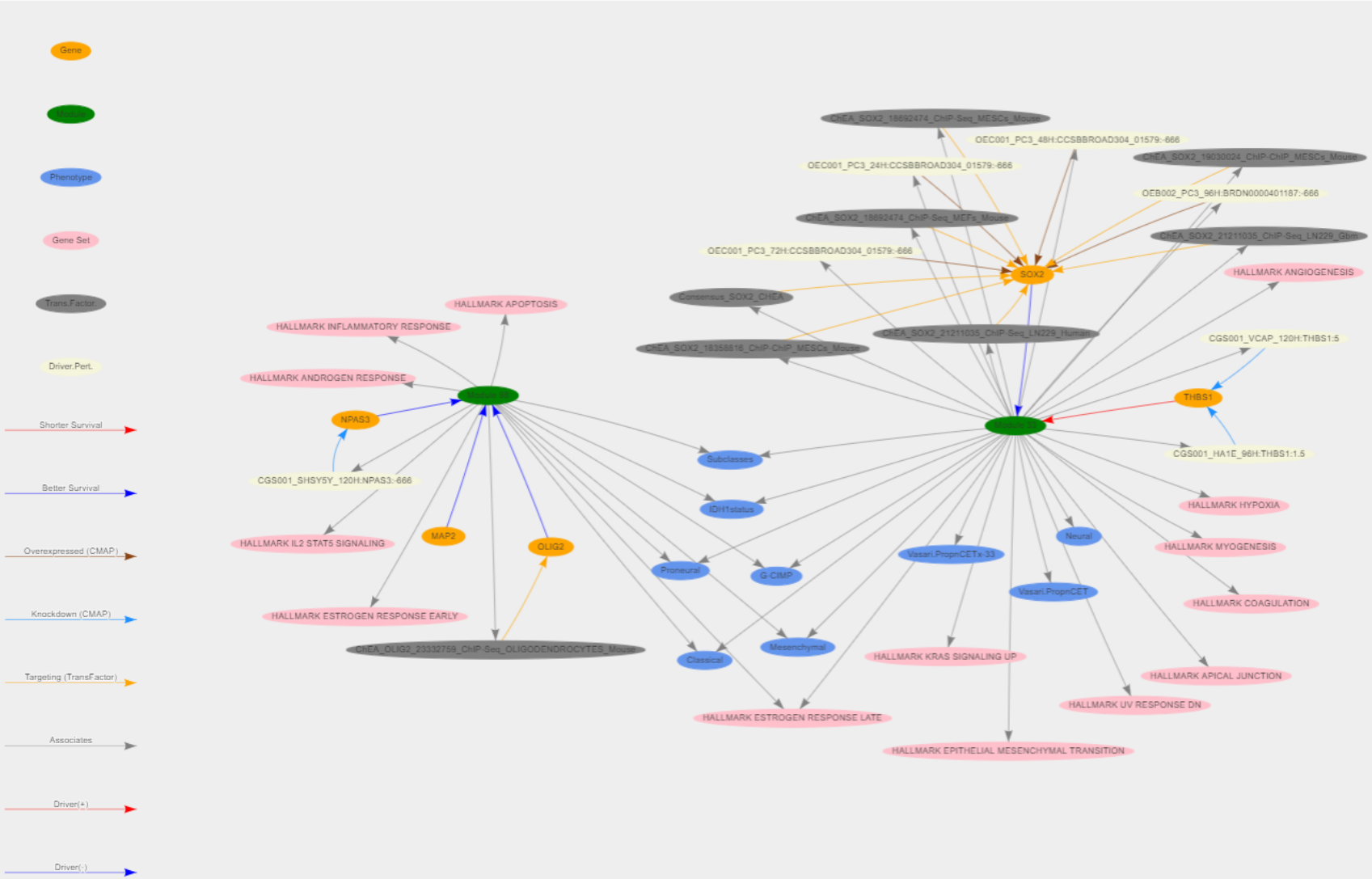
Driver Discovery:

- show none
- show all drivers validated
- show all genes validated

Validation Status

ES+Pval+Padj

Network | Genes | Modules | Phenotypes (Clinical Characterization) | Gene Sets (Functional Characterization) | Transcription Factor Discovery | Driver (Perturbation) Discovery



https://pochetlab.shinyapps.io/AMARETTO-Hub_TCGA-GBM/

Use Cases: integrating multi-omics, clinical, imaging, and driver and drug perturbation data across model systems and patient studies of cancer

Use Cases: integrating multi-omics, clinical, imaging, and driver and drug perturbation data across model systems and patient studies of cancer

1. A study of hepatitis C and B virus-induced hepatocellular carcinoma (LIHC) with driver and drug discovery for chemoprevention across pan-etiologies of hepatocellular carcinoma, experimentally validated in rat models

HTML Reports: http://portals.broadinstitute.org/pochetlab/demo/cAMARETTO_Liver_6DS/index.html

Jupyter Notebook: https://colab.research.google.com/drive/17GieTfYriTVlbKchl-OEb5nI_NA2vvjQ

Use Cases: integrating multi-omics, clinical, imaging, and driver and drug perturbation data across model systems and patient studies of cancer

1. A study of hepatitis C and B virus-induced hepatocellular carcinoma (LIHC) with driver and drug discovery for chemoprevention across pan-etiologies of hepatocellular carcinoma, experimentally validated in rat models

2. A study of glioblastoma multiforme (GBM) and low-grade glioma (LGG) with driver discovery for diagnostic and prognostic molecular subclasses associated with radiography and histopathology imaging-derived features for imaging diagnostics

HTML Reports: http://portals.broadinstitute.org/pochetlab/demo/IcAMARETTO_Brain_3DS/index.html

Jupyter Notebook: https://colab.research.google.com/drive/14u1KZJ3Gf-9qjDycyBKzBiN5VzzOa2xU and https://colab.research.google.com/drive/11Q0GH6YHCoTZrDfHrUavixAi_f_qGBau

Neo4j Graph: https://pochetlab.shinyapps.io/AMARETTO-Hub_TCGA-GBM/

Use Cases: integrating multi-omics, clinical, imaging, and driver and drug perturbation data across model systems and patient studies of cancer

1. A study of hepatitis C and B virus-induced hepatocellular carcinoma (LIHC) with driver and drug discovery for chemoprevention across pan-etiologies of hepatocellular carcinoma, experimentally validated in rat models

2. A study of glioblastoma multiforme (GBM) and low-grade glioma (LGG) with driver discovery for diagnostic and prognostic molecular subclasses associated with radiography and histopathology imaging-derived features for imaging diagnostics

3. A pan-cancer study across twelve cancer sites with driver discovery of pan-cancer drivers of smoking-induced and 'antiviral' interferon-modulated innate immune response cancer

Champion *et al*., *EBioMedicine* 2018

Use Cases: integrating multi-omics, clinical, imaging, and driver and drug perturbation data across model systems and patient studies of cancer

1. A study of hepatitis C and B virus-induced hepatocellular carcinoma (LIHC) with driver and drug discovery for chemoprevention across pan-etiologies of hepatocellular carcinoma, experimentally validated in rat models

2. A study of glioblastoma multiforme (GBM) and low-grade glioma (LGG) with driver discovery for diagnostic and prognostic molecular subclasses associated with radiography and histopathology imaging-derived features for imaging diagnostics

3. A pan-cancer study across twelve cancer sites with driver discovery of pan-cancer drivers of smoking-induced and 'antiviral' interferon-modulated innate immune response cancer

3. A pan-cancer study of squamous cell carcinoma (SCC) across five SCC cancer sites, in particular, lung (LUSC), head and neck (HNSC), esophageal (ESCA), cervical (CESC) and bladder (BLCA)

HTML Reports: http://portals.broadinstitute.org/pochetlab/demo/cAMARETTO_PanCancer_5DS/index.html
Jupyter Notebook: https://colab.research.google.com/drive/17RwBxwWWnXJMRI_VZl-X-hztJTwvlFjl

Team: Lab & Collaborators

Pochet Lab (BWH/HMS/Broad)

Mohsen Nabian

Artur Manukyan

Celine Everaert

Rileen Sinha

Tom Croonenborghs

Carey Lab (BWH/HMS/Broad)

Vincent Carey

Regev Lab (MIT/Broad)

Aviv Regev

Brian Haas

Gevaert Lab (Stanford/Broad)

Olivier Gevaert

Jayendra Ravindra Shinde

Shaimaa Hesham Bakr

Andrew Gentles

Kevin Brennan

Magali Champion

Hernaez Lab (Illinois)

Mikel Hernaez

Mesirov Lab (UCSD/Broad)

Jill Mesirov

Michael Reich

Ted Liefeld

Thorin Tabor

Baumert Lab (Strasbourg)

Thomas Baumert

Joachim Lupberger

Eloi Verrier

Neurology (BWH/HMS)

Anna Krichevsky

Erik Uhlmann

Francisco Quintana

Jishu Xu

Nikolaos Patsopoulos

Vijay Kuchroo

Howard Weiner

NIH NCI CBIIT ITCR Cancer Data Science Pulse Blog

NATIONAL CANCER INSTITUTE
Center for Biomedical Informatics & Information Technology

News & Events

Cancer Data Science Pulse

The Cancer Data Science Pulse blog provides insights on trends, policies, initiatives, and innovation in the data science and cancer research communities from professionals dedicated to building a national cancer data ecosystem that enables new discoveries and reduces the burden of cancer.

Informatics Technology for Cancer Research Program Drives and Fosters Community of Cancer Informatics Researchers: An *AMARETTO Tool Success Story

October 18, 2019

Dr. Nathalie Pochet highlights the Informatics Technology for Cancer Research Program and the support it provides for informatics tools development, including the *AMARETTO framework that is being leveraged to identify novel mechanisms of viral carcinogenesis.

SUBSCRIBE TO UPDATES

CATEGORIES

Data Sharing (17)
Genomics (17)
Precision Medicine (10)
Seminar Series (10)
Data Standards (7)

See all categories +

NATIONAL CANCER INSTITUTE
Center for Biomedical Informatics & Information Technology

News & Events / Cancer Data Science Pulse Blog

Cancer Data Science Pulse

Informatics Technology for Cancer Research Program Drives and Fosters Community of Cancer Informatics Researchers: An *AMARETTO Tool Success Story

Informatics Tools

October 18, 2019

Researchers in cancer informatics are challenged by a profusion of data resources, literature, and computational tools. Mechanisms for breaking down silos, increasing communication, and fostering collaboration are difficult to build and sustain, even in individual institutions. Owing to its impact on my own research on *AMARETTO, I view the Informatics Technology for Cancer Research (ITCR) Program as a major step forward in successfully solving these challenges.

ITCR is a trans-NCI program that supports investigator-initiated, research-driven informatics technology development spanning all aspects of cancer research. Unlike typical research approaches, ITCR places an emphasis on engaging users and investigators in the cancer research community in a variety of methods, including:

- Monthly conference calls and annual meetings of investigators working in a variety of domains spanning tumor genetics, genomics, and imaging, which has exposed me to projects and their initiators in ways that don't arise in typical approaches to research.
- Funding opportunities foster various stages of the informatics technology development lifecycle, including algorithm development, prototyping and hardening, enhancement and dissemination, and sustainment, with administrative effort specifically directed at identifying interconnections between independent projects and set-aside programs to fund specific collaborative proposals.
- Monthly ITCR Working Groups focus on training and outreach, technical, and sustainability and industrial partnership aspects.
- Webinars, demos, and workshops of its informatics technologies at cancer research conferences.

To inspire the community of cancer informatics researchers, here I highlight how the ITCR Program has successfully impacted the development, dissemination, and general applicability of *AMARETTO, and has led to broadly catalyzing and accelerating new discoveries in cancer.

The Development of *AMARETTO

The *AMARETTO framework provides software tools for network biology and medicine, towards a data-driven platform for diagnostic, prognostic, and therapeutic decision-making in cancer. Specifically, *AMARETTO offers modular and complementary solutions to multimodal and multiscale aspects of network graph-based fusion of multi-omics, clinical, imaging, and driver and drug perturbation data across studies of patients, etiologies and model systems of cancer.

The ITCR Program supports the development of *AMARETTO to identify novel mechanisms of viral carcinogenesis and uncover new therapeutic targets for chemoprevention of hepatocellular carcinoma. Through a collaborative set-aside with Vincent Carey, ITCR also supports Bioconductor-embedded developments of *AMARETTO as an imaging genomics tool for diagnostics and therapeutics in hepatocellular carcinoma and glioblastoma multiforme.

The conceptualization of *AMARETTO is co-led by Olivier Gevaert, Mikel Hernaez and myself, and has grown towards a multidisciplinary investigational team of cancer informaticians, biologists, and clinicians, ultimately leading to a continuously expanding network of informatics technology-driven collaborative initiatives to accelerate biomedical research and healthcare delivery, for better diagnosis and therapy of human disease. Originally formulated for studies of cancer, we are reformulating these tools for applications to other complex diseases; for example, for neurological and immune-mediated diseases in collaboration with my colleagues Jishu Xu, Nikolaos Patsopoulos, Anna Krichevsky, Erik Uhlmann, Francisco Quintana, Vijay Kuchroo and Howard Weiner.

*AMARETTO Supports New Cancer Discoveries

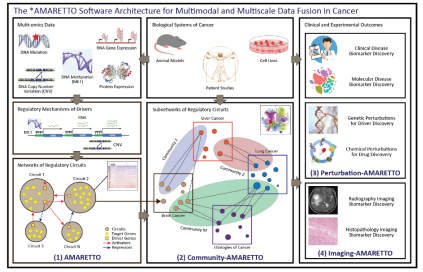
To demonstrate its utility, working with my collaborator Thomas Baumert, we leveraged *AMARETTO to discover novel therapeutic treatments for chemoprevention of hepatitis C and B virus-induced hepatocellular carcinoma. Several predicted drug compounds were successfully validated in vivo as demonstrated by their ability to inhibit liver fibrosis and inflammation with prevention of cancer development in animal models. Thus, they potentially represent a safe and low-cost approach for chemoprevention of hepatocellular carcinoma across viral and other etiologies. For these types of cancer informatics applications, the ITCR Program has recently launched the Working Group on Sustainability and Industrial Partnership to help guide its investigators in successful transitioning of informatics technologies to applications in clinical practice and within industrial partnerships. A startup Alentis Therapeutics has been created to develop novel molecules for treatment of advanced liver disease and cancer.

ITCR Catalyzes the Dissemination and Hardening of *AMARETTO

The ITCR Program already identified and sustained GenePattern's team led by Jill Mesirov and Michael Reich to support Notebook-oriented interfaces for dissemination, and Bioconductor's team led by Vincent Carey, Martin Morgan, and Levi Waldron as a vehicle for dealing with testing, performance evaluation, and continuous integration of the underlying code. These ITCR-initiated collaborations have catalyzed the dissemination of *AMARETTO as user-friendly tools via GenePattern and Bioconductor, and GenePattern and Jupyter Notebooks for several case studies of cancer, including virus-induced hepatocellular carcinoma, glioma and glioblastoma, and pan-cancer studies (EBioMedicine 2018).

Future *AMARETTO Impact Through ITCR

To broaden its impact, our team continues to establish connections to other informatics tools supported by ITCR, including interrogating novel genetic and epigenetic heterogeneity discovered from cancer transcriptomes assembled from individuals or single cells by Trinity CTAT (led by Aviv Regev and Brian Haas) in *AMARETTO networks, disseminating networks via the interactive network database NDEx (led by Trey Ideker and Dexter Pratt) and dynamic interactive heatmaps for interpretation of networks with NG-CHM (led by Bradley Broom). In parallel, the *AMARETTO team will continue to follow the ITCR's guidelines for transitioning *AMARETTO to applications in clinical practice and within industrial partnerships, such as applications of drug discovery focused on validation through clinical trial studies.



The *AMARETTO software architecture. The *AMARETTO framework provides software tools for network biology and medicine, towards a data-driven platform for diagnostic, prognostic and therapeutic decision-making in cancer. The *AMARETTO platform offers modular and complementary solutions to multimodal and multiscale aspects of network graph-based fusion of multi-omics, clinical, imaging, and driver and drug perturbation data across studies of patients, etiologies and model systems of cancer. Specifically: (1) The AMARETTO algorithm learns networks of regulatory circuits - circuits of drivers and target genes - from functional genomics or multi-omics data and associates these circuits to clinical, molecular and imaging-derived phenotypes within each biological system (e.g., model systems or patients); (2) The Community-AMARETTO algorithm learns subnetworks of regulatory circuits that are shared or distinct across networks derived from multiple biological systems (e.g., model systems and patients, cohorts and individuals, diseases and etiologies, *in vitro* and *in vivo* systems); (3) The Perturbation-AMARETTO algorithm maps genetic and chemical perturbations in model systems onto patient-derived networks for driver and drug discovery, respectively, and prioritizes lead drivers, targets and drugs for follow-up with experimental validation; and (4) The Imaging-AMARETTO algorithm maps radiography and histopathology imaging data onto the patient-derived multi-omics networks for non-invasive radiography and histopathology imaging diagnostics. *Credits to *AMARETTO team members: Mohsen Nabian, Artur Manukyan, Celine Everaert, Shaimaa Bakr and Jayendra Shinde.*

Nathalie Pochet, Ph.D.
Assistant Professor, Department of Neurology, Harvard Medical School; Associate Member, Cell Circuits Program, Broad Institute of MIT and Harvard; Associate Scientist, Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital

SUBSCRIBE TO UPDATES

CATEGORIES

Data Sharing (17)
Genomics (17)
Precision Medicine (10)
Seminar Series (10)
Data Standards (7)

See all categories +

ARCHIVE

2019 (9)
2018 (14)
2017 (5)
2016 (14)
2014 (6)

https://datascience.cancer.gov/news-events/blog/informatics-technology-cancer-research-program-drives-and-fosters-community-cancer