ITCR Monthly Meeting

November 4, 2022

2:00 - 3:00 pm ET

Zoom connection information:

<https://nih.zoomgov.com/j/1605449420?pwd=c01lL2VLRGhQVDRzYUdSNG5TOGVuQT09>

Meeting ID: 160 544 9420

Passcode: 074946

One tap mobile

+16468287666,,1605449420#,,,,\*074946# US

Dial by your location

 +1 669 254 5252 US (San Jose)

 +1 646 828 7666 US (New York)

 +1 551 285 1373 US

 +1 669 216 1590 US (San Jose)

**Agenda**

| **2:00 - 2:10** | Program UpdatesITCR Training Network and Training and Outreach WG updatesTOW Website:<https://www.itcrtraining.org/tow> | Juli KlemmCarrie Wright |
| --- | --- | --- |
| **2:10 - 2:25** | New U01 Award - muMS2: An open source R package for analyzing and integrating multi-omics datasets to improve early cancer detection and understanding of colorectal cancer | Patrick SchlossMarcy Balunas |
| **2:25 - 2:40** | U24 Renewal - OpenCRAVAT: Informatics tools for high-throughput analysis of cancer mutations | Rachel Karchin |
| **2:40 - 2:55** | New R21 Award: Predicting actionable cancer vulnerabilities enabled by mutant-directed protein-protein interactions | Andrey Ivanov |
| **3:00** | Adjourn |  |

ITCR Twitter Hashtag: #nciitcr

Website: <https://itcr.cancer.gov>

GoogleDocs for past meetings: <https://drive.google.com/drive/folders/0BzGsmNN8MvZqeWViY0xCMmRMTDQ>

When you join the meeting, please enter your name and institution below:

Brooke Fridley, Moffitt Cancer Center

Rudi Pillich, UC San Diego

Rob Frost, Dartmouth

Joel Saltz, Stony Brook

Pablo Gonzalez Camara, University of Pennsylvania

Harry Hochheiser, University of Pittsburgh

Jacob Oleson, University of Iowa

Melissa Cline, UC Santa Cruz

Guergana Savova, BCH/HMS

Kivanc Kose, Memorial Sloan Kettering Cancer Center

Yen-Yi Ho, University of South Carolina

Greg Caporaso, Northern Arizona University

Rick Bradshaw, University of Utah Biomedical Informatics

Travis S. Johnson, Indiana University School of Medicine

Helen Parkinson, European Bioinformatics Institute

Helga Thorvaldsdottir, Broad Institute

Giuseppe Narzisi, NYGC

Spyridon Bakas, UPenn

Mary Goldman, UC Santa Cruz

John Quackenbush, Harvard T.H. Chan School of Public Health

Ben Raphael, Princeton

Avi Ma’ayan, Icahn School of Medicine at Mount Sinai

Marcy Balunas, University of Michigan

Jin Zhang, Wash U

Jerry Li, NCI

Dave Miller, NCI

Obi Griffith, WashU

Alexej Abyzov, Mayo Clinic

Malachi Griffith, Washington University

Jill Mesirov, UC San DiegoB

Michelle Berny-Lang, NCI

Jeremy Goecks, Oregon Health & Science University

Leili Shahriyari, UMass Amherst

Michael Reich, UCSD

David Hanauer, U of Michigan

Rachel Karchin, JHU

Zeynep Kosaloglu Yalcin, LJI

Ken Chen, MD Anderson

Kooresh Shoghi, WUSM

Fuyong Xing, University of Colorado Anschutz Medical Campus

Tahsin Kurc, Stony Brook University

Michael Schatz, Johns Hopkins University

Itai Yanai, NYU Langone Health

Daniel Rubin, Stanford University

James Eddy, Sage Bionetworks

Meghan McGrady, Cincinnati Children’s Hospital Medical Center

Veronica Rotemberg, Memorial Sloan Kettering Cancer Center

Prateek Prasanna, Stony Brook University

Martin Morgan, Roswell Park

John Weinstein, MD Anderson Cancer Center

Jiang Bian, University of Florida

Josh Campbell, BU

Philip Montgomery, Broad Institute

Feifan Liu, UMass Chan Medical School

Brian Haas, Broad Inst

Cliff Meyer, DFCI HSPH

Li-Xuan Qin, MSKCC

Christina Yung, Indoc Research

VinceCarey, Harvard

Mauricio Menegatti Rigo, Rice University

Candace Savonen, Fred Hutch Cancer Center

Levi Waldron, CUNY

Malgorzata Marjanska, University of Minnesota

Elana Fertig, JHU

Carrie Wright, Fred Hutch

Jan Schuemann, MGH

Andrey Ivanov, Emory

Info about tools (email cwright2@fredhutch.org with questions)

* What kind of data is used by your tool (at what point in the processing of the data) and for what purpose (pre-processing, normalization, visualization, etc.)?
* Where would one go to get started with your tool?
* What makes your tool different than comparable tools?

We will incorporate your tool in our courses and give you credit for your contribution! This may help users find your tool!

**Tools Description for OPEN**

**All RNA-seq and ChIP-seq sample and signature search (ARCHS4)** (<https://maayanlab.cloud/archs4/>) is a resource that provides access to gene and transcript counts uniformly processed from all human and mouse RNA-seq experiments from GEO and SRA. The ARCHS4 website provides the uniformly processed data for download and programmatic access in H5 format, and as a 3-dimensional interactive viewer and search engine. Users can search and browse the data by metadata enhanced annotations, and can submit their own gene sets for search. Subsets of selected samples can be downloaded as a tab delimited text file that is ready for loading into the R programming environment. To generate the ARCHS4 resource, the kallisto aligner is applied in an efficient parallelized cloud infrastructure. Human and mouse samples are aligned against the most recent Ensembl annotation (Ensembl 107).

The **N**etwork **D**ata **Ex**change (**NDEx**) project provides an open-source framework where scientists and organizations can store, share and publish biological network knowledge. A distinctive feature of NDEx is that it serves as a home for models that are currently available only as figures, tables, or supplementary information, such as networks produced via systematic mining and integration of large-scale molecular data.

NDEx includes features to support data distribution and access according to FAIR principles. Its full integration with [Cytoscape](https://cytoscape.org/), the popular desktop application for network analysis and visualization, provides the cloud back-end component for data I/O; so, if a network file format can be opened in Cytoscape, it can also be stored in (and retrieved from) NDEx.

NDEx can be accessed via its web user interface or programmatically, via REST API and client libraries in Python, R, Java. Web applications can interface with NDEx via JavaScript: MSigDB, CRAVAT, cBioPortal and IQuery, are all examples of web applications integrated with NDEx.

For more information, please review the [About NDEx](https://home.ndexbio.org/about-ndex/) page.

To get started, visit the [NDEx public server](https://www.ndexbio.org/): there, you can review the [NDEx FAQ](https://home.ndexbio.org/faq/), access documentation, contact us, and search or browse thousands of biological network models.

**WebMeV** (https://webmev.tm4.org) is an online tool that facilitates analysis of large-scale RNA-seq and other multi-omic datasets by providing intuitive access to advanced analytical methods and high-performance computing for a wide range of basic, clinical, and translational researchers. Although WebMeV provides support for “bulk” RNA-seq data, single-cell RNA-seq, and other types of -omic data and provides easy access to public data resources such as The Cancer Genome Atlas (TCGA) and the Genotype-Tissue Expression project (GTEx)—as well as user-provided data. WebMeV uniquely provides a user-friendly, intuitive, interactive interface to processed analytical data uses cloud-computing elasticity for computationally intensive analyses that are increasingly required for genomic data analysis. WebMeV’s design places an emphasis on *user-driven* data analysis by providing users the ability to visualize, interact with, and dissect genomic data at each step in the analysis with a “*point-and-click*” interactive data environment. Although the primary input is normalized “count matrices,” WebMeV does include tools for data normalization and quality control and uses Dropbox and Google Drive as means of easily uploading data. Analytical methods include statistical tests for comparing cohorts, for identifying gene seats, for doing functional enrichment analysis on gene sets (GSEA), and for inferring gene regulatory network models and comparing these networks between phenotypes to understand the drivers of disease. WebMeV also provides a platform to support reproducible research and makes code for the entire system and its component methods available as open-source software code.

**MultiAssayExperiment** (<https://bioconductor.org/packages/MultiAssayExperiment/>) is an R/Bioconductor package that harmonizes data management, manipulation, and subsetting of multiple experimental assays performed on an overlapping set of specimens. It supports on-disk and remote data storage, and provides reshaping tools for adaptability to arbitrary downstream analysis.

MultiAssayExperiment is distinct from alternative approaches in its focus on multi’omic data management and manipulation and in its integration with the Bioconductor ecosystem: it is used by more than 50 other Bioconductor packages, it provides a familiar Bioconductor user experience by extending concepts from SummarizedExperiment while supporting an open-ended mix of data classes for individual assays, and it allows subsetting by genomic ranges, row names, phenotypic data, and assays.

You can get started with the MultiAssayExperiment Bioconductor package documentation, or start with prebuilt MultiAssayExperiments objects from curatedTCGAData (https://bioconductor.org/packages/curatedTCGAData/), cBioPortalData (https://bioconductor.org/packages/cBioPortalData/), or SingleCellMultiModal (https://bioconductor.org/packages/SingleCellMultiModal/).

**Genetic Cancer Risk Detector (GARDE)** screens and identifies patients who meet National Comprehensive Cancer Network (NCCN) criteria for genetic evaluation of familial cancer risk based on their family history in the EHR using both structured data and natural language processing of free-text data. Patients identified by GARDE are imported into an EHR's population health management dashboard (e.g., Epic's Healthy Planet module) where genetic counseling staff review individual cases, select, and send bulk outreach messages to patients via chatbot and/or through the patient portal.

GARDE is a population clinical decision support (CDS) platform based on Fast Healthcare Interoperability Resources (FHIR) and CDS Hooks standards to support interoperability and logic sharing beyond single vendor solutions.

Learn more - https://reimagineehr.utah.edu/innovations/garde/

**Patient Derived Cancer Models Finder** ([www.cancermodels.org](http://www.cancermodels.org)) is a cancer research platform that aggregates clinical, genomic and functional data from patient-derived xenografts, organoids and cell lines. The PDCM Finder standardises, harmonises and integrates the complex and diverse data associated with PDCMs for cancer community.

Data types used are model meta data, related clinical metadata from the sample for which the model was derived, e.g. molecular and treatment-based. Data are preprocessed, consistently semantically annotated, harmonised and FAIR.

PDCM Finder contains >6200 models across 13 cancer types, including rare pediatric models (17%) and models from minority ethnic backgrounds (33%), making it the largest free to consumer and open access resource of this kind.

Get started at [www.cancermodels.org](http://www.cancermodels.org) to browse and query models by cancer type

**CIViC (**[**www.civicdb.org**](http://www.civicdb.org)**)** is a knowledgebase and curation interface for the clinical interpretation of variants in cancer. Evidence is curated from published literature describing the diagnostic, prognostic, predictive, predisposing, oncogenic, or functional role of variants in specific cancer types. Evidence submitted by community curators is revised and moderated by expert editors. Individual evidence is synthesized into gene summaries, variant summaries and variant-disease assertions of specific clinical relevance. Anyone can make use of CIViC knowledge through the open web interface or API. Information on how to use or contribute to CIViC is available in our help docs (docs.civicdb.org). The main distinguishing feature of CIViC, compared to similar resources, is its total commitment to open data sharing. All data are available in the Public Domain (CC0) through our website or open, documented API. This has allowed CIViC to be integrated very broadly into both the academic and industry ecosystems for cancer variant interpretation. The code is available for any use under an MIT license.

The **Trinity Cancer Transcriptome Analysis Toolkit** (CTAT, <https://github.com/NCIP/Trinity_CTAT/wiki> ) provides a diverse collection of tools to gain insights into the biology of cancer through the lens of the transcriptome. Using RNA-seq as input, CTAT modules enable detection of mutations, fusion transcripts, copy number aberrations, cancer-specific splicing aberrations, and oncogenic viruses including insertions into the human genome. CTAT uses both read mapping and de novo assembly methods to analyze RNA-seq, leveraging tumor bulk and single cell transcriptomes. CTAT modules provide interactive visualizations as outputs, are easily installed for local execution or run via cloud computing (eg. Terra), have detailed user guides and tutorials, and are well-supported through user forums.

**GenePattern**, www.genepattern.org, is an open software environment providing access to hundreds of tools for the analysis and visualization of genomic data. Analyses include general machine learning methods, the gene set enrichment analysis suite, ‘omics-specific tools for bulk and single-cell gene expression, proteomics, flow cytometry, variant annotation, sequence variation and others, as well as cancer-specific analyses. Also included are data preprocessing and utility tools. A web-based interface provides easy, non-programmatic access to these tools and allows the creation of multi-step analysis pipelines that enable reproducible *in silico* research.

The GenePattern Notebook interface, notebook.genepattern.org, extends the Jupyter Notebook system to allow users to combine GenePattern analyses with text, graphics, and code to create complete research narratives. It includes many additional features to make notebooks accessible to non-programmers. The online GenePattern Notebook Workspace allows investigators to create, run, and collaborate on notebooks using only a web browser. A library of GenePattern Notebooks implementing common scientific workflows is available for investigators to use as templates and adapt to their own requirements.

To get started with GenePattern you can go through the [GenePattern Quick Start Tutorial](https://genepattern.org/quick-start), view the [GenePattern User Guide](https://genepattern.org/user-guide), or the videos on our [YouTube channel](https://www.youtube.com/c/GenePattern). To learn more about GenePattern Notebook, view the [GenePattern Notebook Quick Start](https://notebook.genepattern.org/quickstart/), [GenePattern Notebook documentation](https://notebook.genepattern.org/docs/), run through the [tutorial notebooks](https://notebook.genepattern.org/library/) (click the Tutorial button), or view the videos on the [GenePattern Notebooks YouTube channel](https://www.youtube.com/channel/UCzUaocvV5z_-HTi_uYCNhkw).