ITCR Annual Meeting, May 31st - June 1st, 2017, Santa Cruz

Meeting Notes with Focus on Identifying Challenges, Needs and Collaboration Opportunities

Short URL https://goo.gl/nEnO7A
Twitter #nciitcr

TOPICS FOR FINAL DISCUSSION

- Fill Out 1 Link (Proposed or Existing) From One Resource to Another On The Shared Google Spreadsheet.
- TOPICS:
    - API ecosystems for “serverless” computational pathology
    - Collaborations (and co-funding opportunities) between ITCR resources and expert curation teams (e.g., ClinGen)
    - ITCR/IMAT co-hosted meeting
    - Plans for another ITCR special issue publication
    - Ideas for next year’s F2F meeting
    - New or modified tools to support comparative cancer genomics (mouse and human, for example. But not limited to mouse as a model system)
    - Should research be funded as a function of the analytics of API calls to the resources we develop (just kidding?)
    - Slack channel for ITCR Developers?
    - Add social media links for resources (optional) to ITCR tools page
    - Other working groups?
    - The topic of creating JSON/XML etc. files describing computational components wrapped in Docker containers. I have a feeling we are duplicating work. Can we discuss as a technical WG activity to understand what people do, and how to harmonize? (Andrey Fedorov)
        - See biocontainers.pro, dockstore.org, commonwl.org, boutiques.github.io
        - Several ITCR groups are using Slicer Execution Model
            - QIICR, HistomicsTK
- DISCUSSED
    - List resources that are used, ranked by usage frequency (e.g. Docker). Pointers to help/resources for these.
        - Recommend tools that you might also be interested in (ala Amazon)

- Add interactive version of ITCR resource linkage net to wiki. (Andrey), web service with tsv data published real-time here.
    - Keep internal and include all (include proposed resources)
    - Publish the version with existing resources only
    - NDEx can do this!
    - Add ability to vote for popular/most requested
    - Add target audience
    - Analyze for synergy/combined capabilities for collaborating
        - (e.g. tumor evolution inference w/ relevant imaging data)

Obi Griffith, CiVIC -- Clinical Interpretations of Variants in Cancer
www.civicdb.org - Nature Genetics paper
CiVIC principles: Standards, openness, transparency (community-driven)
Gene > Variant > Evidence
No login to consume content
Behind any comment is a "talk" page, to discuss the content
Considerations for merging conflicts
Data accessible via restful API.
2674 interpretations for 1012 variants, 308 genes, 188 cancer types
External contributions account for about 50% of the content
Coming soon to cBioPortal, along with many other resources that incorporate CiVIC content
Q: How much of community is clinical?
A: No exact numbers of clearly a mix. It is being used in molecular tumor board settings. User type is in the profile.
Q: Link to other information websites, e.g. UniProt?
A: Yes, link out to many variant resources. Also, trying to create a common portal for knowledge bases. Via GA4GH subgroup.
Q: How different from FoundationOne interpretations?
A: They are an inspiration for this. They have an undocumented process, use an external provider. It is part of their product.
Q: Size of the output per variant?
A: No more than one paragraph, though no fixed rules. Aiming for concise.

Terry Meehan -- PDX Integrator
Aims
- Allow clinicians and researchers to find relevant PDX models that are available for distribution
- Allow them to find and reuse data from PDX models

Defining minimal information for PDX models

- Bult, Butte, others

PDX-MI Manuscript: Feedback from over 35 institutes, for special issue of Cancer Research
AACR Satellite workshop >> www.pdxfinder.org << Jackson Labs + EMBL-EBI co-developed resource
Graph database backend - github.com/pdxfinder
Using a semantic middle layer. Zooma webservice: Find the best term, what are the terms most used by other resources
Provide links to related data for the PDX model (e.g., genomics, gene expression, etc).
Already have exchange mechanisms with NCBI
SAB established - first call in July
EGA - Patient identifiable data
Q: Extensible to organoids? A: Yes, keeping this in mind.
Q: Tracking failures? A: Some to fail to engraft and this aspect is covered. Gets tricky after that.
Q: Do you restrict the number of passages?
A: Do track this information, including when QC was done. But not doing any restriction. Most not going over 6.
Q: How much clinical information:
A: Pretty minimal, b/c we don't want to identify the patient

Bradley Broom, NG-Clustered Heat Maps
Dynamic exploration of very large data sets
Biggest HM to date: 20k genes
Link-outs based on rows, commons, or matrix elements.
Toolboxes: Annotation, Biology, Chromosomal, Clinical, Statistical
Builder levels; GUI, R, Specification files
New system architecture
- Standalone server-less mode possible. Allows NG-CHMS to be saved locally and shared.

Integrated into the Galaxy infrastructure
BigQuery Interface
Available: Docker container, Galaxy toolshed, GitHub
Tcga.ngchm.net: new TCGA compendium
Many new improvements for the viewer itself
Developing a tablet-optimized user interface
Plug-in or map-centric applications
Q: What is the API?
A: RESTful

Q: Can you compare a patient sample with the compendium?
A: No, but great idea. Can upload a data vector
Q: Who is your user audience? Can non-bioinformatians use the tool?
A: Bioinformaticians create the HM, then give it to the clinician.

Levi Waldron - MultiAssayExperiment
GitHub repository; Bioconductor package to represent multiple genomic assays from overlapping samples

- Gene expression; Copy number; Variants; …

Design goals: Relate each assay to others and to clinical data. Robust and standardized. Handle missing data. Handle id-based and range-based data. Support on-disk representations.
Example

- 33 cancer types from TCGA all represented and readily accessed
- Integration into 'experiment hub' Bioconductor infrastructure a next step.

Example operations

- 'Upsetter' summary of sample presence / absence across assays
- Integrative analysis, e.g., mutations / mb vs. SCNA (copy number) score; key lines of code (4 total) replicate original analysis in *Science*.
- Integrating **remote data**, e.g., 1k Genomes variant call files, incorporated into MultiAssayExperiment and local analysis

Video.

- Easiest way to access TCGA data for statistical analysis

Q: What do you do about batch effects?
A: That is generally up to the user. Weinstein: We have MBatch, that may be applicable here.
Q: Can you include radiological data? If no, any reason not to?
A: Just need something that is vector-like for each patient. Information after feature-extraction. Interested in looking at the H&E slides.

Rachel Karchin and Mike Ryan - CRAVAT and MuPIT
CRAVAT is a portal for cancer mutation analysis
https://www.cravat.us/CRAVAT/
https://www.cravat.us/MuPIT_Interactive/
Dockerized: https://hub.docker.com/r/karchinlab/cravatmupit/
Niche: Biologist-friendly and interactive
Docker image: 647 pulls. But difficult to track usage stats here. ITCR training and outreach group is working on a white paper to address this issue.
Can integrate as an I-Frame within your own site

To link out to MuPIT, you just need a genomic location
CRAVAT Web service - returns machine-readable JSON structure
Q: Do you have a way to distinguish artifact?
A: We have tracks to identify where there might be sequencing errors. Could reference the paper that was recently published on this topic.
Q: How do you handle mapping between database? UniProt, RefSeq, PDB, etc.
A: We have done it and our data is freely available to you.
Q: How do you make this a clinician-friendly tool? E.g. for CiVIC users
Q: (unasked) with strongly visual tools it seems challenging to perform *reproducible* analysis (user can recreate the analysis tomorrow that they did today) and to retain *provenance* (the data in this derived result is based on data resource versions x, y, and z). It seems like these aspects become increasingly important when transitioning to clinician-oriented tools. So are reproducibility and provenance emphasized enough in these and other visual tools?

Mike Becich -TIES Cancer Research Network (TCRN)
https://cancerdatanetwork.org/
Network Trust Agreements

- IRBs agree that use of data for investigators is NHSR, no need for an additional IRB protocol even to access record-level de-ID data

>2.4M patients represented in the network
Use in Tissue Bank

- Honest Broker is key
- Whole slide images - cited recent FDA approval

Combining rare-cohorts across multiple institutions is a key use case
New development: Adding Cancer Registry Data to TIES
Nexi supports deployment and use of TIES
Q: Do you have corresponding genomics data for these patients
A: Have coupled the TCRN resources with TCGA >> TCGA Expedition (11% of all TCGA samples came from Pittsburgh). Deeper data mining platform.
Q: Any contact with MD Anderson about the network?
A: Some, though the skin SPORE. But, should discuss a partnership…
Q: How many lawyer's hours have this taken? How is IP handled?
A: Pitt working very hard to make sure this is an innovation engine. "Fired our conservative lawyers…"

Xiaodong Wu - PET-CT Image analysis tools
Deep learning approach for response prediction

Tumor co-segmentation in PET-CT. Note that different imaging modalities produce different tumor volume measurements. Provide different information. Tumor contour difference in the two modalities.
Aim 1: Co-Segmentation method
Co-segmentation method as a 3D Slicer module
User Video available on YouTube, GitHub
Aim 2: Prediction of therapeutic response

- Data Collection
- Generate ground truth for training CNN

Q: How many experts for the ground truth?
A: In the second set, there were 3.
Q: input from experts on energy constraints?
A: Just manually trying the value
Q: How to evaluate in the context of FDR?
A: Separate the training and validation data sets.

Lauren O'Donnell - Open source diffusion MRI software for brain cancer research
Measures brain connections in vivo.
Goal: Better open source software for supporting surgery for brain tumors

- Remove as much of the tumor as possible while maintaining quality of life

Slicer dMRI dmri.slicer.org
Measure distances from the fiber tracts to the tumor
Create state of the art dMRI workflows.
3D Slicer as a platform for dissemination - Extension manager

- Separate from core code for faster developments
- Easy selection of extensions from within 3DSlicer

New user forum for slicer: discourse.slicer.org. Includes community forums
Documenting usage on the website, based on downloads
Released DICOM Tractography. First implementation of the new international standard.
Interoperability challenge: RSNA 2017
Published automated tractography
New open source community interactions

- Automated tract ID now in testing at Mayo
- Discussion started to combine DICOM code with community software TractConverter

Community library Dipy: python toolbox for analysis of MR diffusion imaging (dipy.org). Very popular in dMRI community but no UI. Trying to integrate and become compatible with more community libraries.
Q: Considered using WebGL/Cornerstone?
A: This is more for visualization, but not so much about computation components.

Christos Davatzikos - Cancer Imaging Phenomics Toolkit (CaPTk)
Leveraging long-standing collaborations with clinical teams who provide input on development of the algorithms and provide data sets.
Tools and algorithms:

- Breast segmentation (LIBRA)
- ITK-SNAP
- GLISTRboost Segmentation (Imaging Processing Portal: ipp.cbica.upenn.edu)
- Confetti - Visualizing fiber tracts

CaPTk Radiomic Panel - Feature synthesis and Integration via Machine Learning
Examples

- Imaging signatures of molecular characteristics, e.g., EGFR
- Predicting infiltration and recurrence
- Predictors of clinical outcome
- Connectomics signatures (related to O'Donnell tractography)
- Radiomic breast cancer phenotypes

Q: Can you develop a script for batch processing of large numbers of cases?
A: This is an active area for us.
Q: What have you learned about cancer phenotypes? When will we see this in the clinic?
A: Surprised at how well the heterogeneity signal predicts EGFR mutation or other mutations or clinical outcomes.

John Quackenbush - Quantitative Radiomics
Use quantitative image features to make predictions
Goal: Synthesis of multi-scale patient data and correlate with clinical data

Radiomic engineered feature set - ~1600
www.radiomics.io
PyRadiomics package - extracts 2000 engineered features per image. Easily customizable. All parameters in 1 file. Extensive documentation
as well as a package integrated into 3D slicer
Q: Won't AI features be superior over engineered features? Less bias?
A: Yes, you must be skeptical, but if it is a good biomarker we can provide to augment what a radiologist does.
Q: What about integrating with genomics?
A: When you add them together, you get a better predictive outcome.
Q: Moonshot is interested in developing a human tumor atlas. The vision is that combining all this information, you will have enhanced predictive ability. What is your vision?

A; This is what we have but there are more things that we'd like to do.

TOW Lunch Breakout Session http://bit.ly/2rEYkSJ
Should this group provide outreach to the general public? This would be best done in collaboration with OCPL. E.g., connect popular topics such as AI to real activities in cancer informatics
Office of advocacy/outreach to get patient advocates involved. Consider inviting Tony Dickherber to an ITCR meeting to discuss.
Link concepts such as "data sharing is good" to real data analysis tools and activities.

Question: What are the pros/cons of a "phone home" functionality?
A: Not a huge deal, as long as people can opt out. Much of the software we use every day has features like this.
Maybe ok to do "lightweight" (e.g. version) checking
Summary: This topic did not raise any significant concerns

GitHub Topics: Label tools supported in some way through ITCR
Use "nci-itcr" as the topic for GitHub
Topic: nci-itcr
This is really targeted for the technical audience

Vincent Carey - Cloud Scale Bioconductor

1. Autonomous provisioning of hardware and software
2. Onco-nuggets for cloud-based app development
3. Scalable statistical learning with cancer applications

Q: what about people who don't want to send their data to a public cloud?
A: Private clouds need to be addressed as well.

Nathalie Pochet - Virus reprogramming for HCC
AMARETTO: Integrates genetic, epigenetic, proteomic and transcriptomic data
Updates for this grant;

- Incorporate prior knowledge
- Jointly model multiple biological systems and functional genomics levels
- Model known dependencies between genes

Raja Mazumder - Data integration for biomarker evaluation and discovery
Based on technology originally developed through EDRN
Q: How will you plan to do the integration?
A: Depends on data type. Work with curators to develop a common view of information.

Hua Xu -

1. Data extraction through NLP
2. Data harmonization
3. Demonstration projects

Challenge of clinical NLP - Portability (across record types, institutions, ..)

- None can achieve optimal performance for all tasks

Need a solution for users to efficiently develop NLP pipelines for specific use
CLAMP - Clinical Language Annotation, Modeling, and Processing
GUI for building customized clinical NLP pipelines
Enterprise version: uses the pipelines built using the CLAMP GUI to do the actual processing
Added rule specification
CLAMP-Cancer
Future work:

- Facilitate portable NLP solutions
- New interfaces for cancer researchers
- More cancer specific NLP pipelines
- Interoperability with other NLP systems

Clamp.uth.edu
Q: How do you compare with IBM?
A: They do have not real clinical NLP pipeline. They are using MetaMap.
Q: How would it use on radiology notes?
A: Yes, should work!
Q: REST API? Containerized?
A: Desktop version has 4 Java APIs. Can build a web service on this.

Guoqian Jiang - caCDE-QA
ISO 11179 Common Data Element Standard

1. Tools for the validation and harmonization of cancer study CDEs
    a. UMLS-based
    b. Semantic Web-based
2. Apply tools to audit experimental cancer study CDEs
    a. Including FHIR data validation
3. Deploy a web-portal for collaborative CDE review

ShEx is now part of the FHIR specification

D2Refine for Metadata Harmonization and Validation

Q: What backend should be used for FHIR? Triple store?
A: Most implementations are based on relational databases.

Q: Do you have a guide for non-experts?
A: Best if you talk to an ontology expert. E.g., is there a standard vocabulary available for your domain? E.g. check NCBO, OBO Foundary

Hong Yu - Machine Learning Intelligence
Detecting medication and adverse drug events from EHRs
Annotated >1000 EHR notes. Making this corpus freely available.
Duplication is a challenge for data mining. Have shown that 40% of EHR content is duplicated.

- Exact copy and paste
- Approximate copy and paste
- Event repeat

Findings

- 23% events duplicated
- Only 6% ADE's duplicated

RNN - Recurrent neural network
Developed 4 new deep learning models
Q: Would prior knowledge improve performance of deep learning?
A: Yes, and we are working on this.

Guegana Savova - DeepPhe Project
Question: Given the entire EHR of a patient, can you extract the cancer phenotype of that patient. No longer dichotomization for a particular phenotype of interest. All the phenotypes associated with the patient
Driven by translational research as well as surveillance (SEER)
Using Apache cTAKES for entity-level recognition
Major challenge: Temporarily in the summarization module

SEER Dev Subset Evaluation
Good initial results w/o training

Developing a visualization tool
Also output to I2B2 TransMart

Gabor Marth -- Tumor subclonal heterogeneity
SubCloneSeeker method identifies subclones consistent with observed data and eliminates those that are not
Number of "consistent" subclone structures is small
Can further reduce feature space with multiple biopsies from the same patient

Now extending tools to enable analysis for longitudinal tumor evolution >> SuperSeeker >>Looking for SC collaborations - datasets!

Dinler Antunes - Structure-based selection of tumor antigens for T-cell based Immunotherapy

Molecular docking to explore the configuration of the peptide ligand in the binding cleft
Need to reduce dimensionality

- Ad hoc approaches
- Crystal templates
- HLA-specific constraints

DINC - Docking INCrementally
Use an iterative method to address the DoF problem - use a few residues/atoms at a time
DINC 2.0 now available
Using the compute cluster at Rice as well as TACCC

Grace Xiao - Single-Nucleotide Analysis of RNA-seq

Why single-nucleotide analysis

- Expressed SNP/mutations
- RNA editing

RASER: Read aligner for SNPs and Editing sites in RNA
GIREMI: Genome-independent identification of RNA editing my mutual information
deSNP: De novo expressed SNP identification using RNA-Seq alone
**eGRAS,iGRAS: Predict functional SNPs in mRNA processing
**Functional RNA editing sites in 3' UTRs or introns

CLIP-Seq: Protein-RNA interaction profiles
Stat model for allele specific binding (ASB) needs to account for

- Systematic bias due to crosslinking
- Variation between replicated experiments

Q; could any of this variation be due to infidelity of RNA Polymerase?
A: Trying to derive a new method to identify this.
Q: what about artifacts due to de-methylation? Also, have you experimentally evaluated the binding?

Brian Haas - Trinity

De novo transcriptome assembly
~2000 unique users/month
Active open source community
Usage tracked by unique IP address

Transcriptome Analysis Toolkit (CTAT) available through Galaxy via NGAS at IU
Integration with IGV.js
DISCASM - just discordant or unmapped reads
STAR-fusion outperformed other fusion finding tools in benchmarking. Also fast!
(https://github.com/STAR-Fusion/STAR-Fusion/wiki)
(http://biorxiv.org/content/early/2017/03/24/120295)
FusionInspector - In silico validation
Driving project: Fusions in lymphocytic leukemia, collaboration with Cathy Wu at DFCI
Virus detection, in collaboration with Steven Salzberg, JHU HPV detection
InferCNV tool - just released
Moving everything into FireCloud. Creating Docker and WDL workflows

Q: Difference between running Trinity at IU vs on FireCloud?
A: FC is more scalable but not freely available. IU offers Galaxy integration.

Ben Raphael - Algorithms for Evolution and Migration of Cancer Genomes

Andrey Fedorov - QIICR
QIICR Goals: Standardized machine-readable tools, sharing analysis results publicly
DICOM Image

- Pixel data
- Non-pixel data (metadata or header)

Dcmqi: DICOM for Quantitative Imaging library: Convert between research representations and DICOM representations

Software at https://github.com/QIICR/dcmqi, tools at http://qiicr.org/dcmqi/#/home
DICOM + nonDICOM (e.g. segmentation) + JSON
Includes tools for ensuring conformance to the schema
Validation of the approach - QIN as the driver

- PET/CT head and neck
- Multiparametric MRI, prostate cancer
- CT lung cancer - radiomics for lung nodule malignancy grading

Connectathon at RSNA, including tractography
Q: How to provide a link to related research data?
A: Use domain-specific standards, where it makes sense. LInk through the patient identifiers
Q: Do you see instrument makers adopting this standard?
A: Some do already adopt this kind of approach. Working with vendors. FDA should recognize this as a problem.
Q: How do we put CaPTk in this format?

A: Start with documentation, then let's talk.
Q: Medically-sophisticated patients may want to use this?
A: This is one level up - analysis results.

Steve Moore, XNAT, I3CR
Feature-rich platform
- Pipeline processing
- DICOM integration
- Extensibility
- …

Use Cases
- CNDA
- Human Connetome Project
- Dementia Platform UK
- Clinical trials
- TIP

Many collaborations in discussion - Christos, David, Gordon, Anant, Bruce
Aim 1: Data management
Aim 2: Knowledge management
Aim 3: Proof of concept network

I3CR builds on XNAT
Planned integrations: 3D Slicer, OHIF
- DICOM interfaces
- Extensions to the XNAT API

System integrations
- I2B2, RedCap
- PACS, TCIA
- Clinical tx systems

Reproducibility goals: Store containers used to compute initial results and to reuse
Pilot network - in collaboration with the QIN

XNAT Container Service
Docker server >> REST API << XNAT
- Make it easier to install tools
- Make it easier to run tools

Q: REST API?
A: Yes, all functionality is available through the REST API

Anne Martel - PIIP

Goal: Digital path platform to facilitate wider adoption of WSI
Includes a goal of multi-modality support
Built on the Sedeen viewer - WSI viewer, annotation tools, conversion and crop, mulitmodality imaging, support for many file formats
Algorithms and plugins: color normalization, biomarker quant, Auto TMA spot extraction, stain analysis, out of focus detection. Engaged with a human factors engineer
SDK - provides utilities to efficiently access pixel data. Developer is shielded from the multi-tile formats
In the pipeline:

- Mac and Linux
- Matlab support

Curated datasets

- Collect image databases, richly annotated by pathologists, develop ontologies, benchmark datasets

Image repository - Using Pathcore Web
Q: Deconvolution matrix -details?
A: Can do region of interest or whole slide, depending on how the algorithm has been set up.

Kun Huang - LInks between histological features and genomics
Aims: develop software libraries for integrating genomic, histological images and clinical data. Develop a platform for performing these analyses

Ongoing - Grassmanian Integration of multi-modal data types
iGenomicsR - Analyze and visualize multidimensional genomics data (R Shiny tool)
ITCR Collaborations: Han Liang & Lee Cooper
OSUMO - platform for data analysis

Jayashree Kalpathy-Cramer - Tools for optimized imaging biomarkers
Project driven by activities in the QIN, esp. PET-CT group
Goal: Make radiology image data more accessible to non-imaging scientists & support reproducible research

- Platform development
    - Challenge mgmt
    - Dockerized code
    - Execution engine
    - Evaluation & viz
- Algorithm development
    - Segmentation, radiomics
- Host challenges

- QIN, QIBA, Societies

Recently added the ability to upload a Docker container to submit code. Set up a private registry for those who don't want these to be shared publicly.
Dockerized implementations of many types of segmentation and feature extraction

Joel Saltz - Tools to Analyze Morphology & Spatially mapped data

FDA has approved WSI pathology for primary diagnosis

- Collaboration with SEER - Virtual Tissue Repository - Add pathology image data to SEER registry (pilot)
- Collaboration with TCGA Pan Can lymphocyte infiltration. Key tool is visual, deep learning editor
    - Image and molecular characterizations are not identical
- Collaboration with Jayashree's group to do analyses of pathomic and radiomic data. Uses caMicroscope, FeatureDB, FeatureVis

Q: Are the 20K patches publically available?
A: It will be, when the study is published. Will include the TIL maps.

Jim Robinson, IGV
igv.org

MSKCC reviews all actionable mutations in the IGV
Recent updates

- Features to support long reads and synthetic long reads (barcoded short reads)
- CRAM alignment format
- Performance improvements
- CRAVAT/MuPIT integration

Igv.js - Embeddable component. Integrated in ISB CGC, FireCloud, cBioPortal, Trinity GitHub

- 111 forks
- 81 accepted pull requests, 22 distinct contributors

Works with the GA4GH API

Tim Griffin, Pratik Jagtap - Galaxy-P

Extend Galaxy platform for multi-omic data
Aim 2: Metabolomics support, aimed at non-technical users. Reproducible, scalable workflows for MS-based metabolite quant and annotation.
Usegalaxyp.org - Training and testing of workflows using small test data sets
MVP -- Multi-omics visualization platform. Protein-centric. Plans to launch IGV.js
Aim4 - Metaproteomics - large scale characterization of the entire protein complement of environmental microbiota at a given point in time. Implications in cancer susceptibility and progression. Developing tools and workflows through community effort.

Promotion: Workshops and oral presentations
ITCR integrations - integration with IGV.js and with CRAVAT and with JetStream. New IMAT-ITCR collaboration as part of Beau Biden Moonshot grant.
Collaboration with users and developers all across the world.

Josh Stuart, Kyle Ellrott BMEG
DREAM: Computation to science
But, getting stuck at recapitulating the method. Next-Gen DREAM challenges will produce re-runable code. Method only scored and ranked if they can run it.
Workflows are now being released publically - available to be remixed and improved
BMEG: Layered evidence gathering
Inspired by the Google Knowledge Graph >> Scalable graph-based framework
Developed apps that sit on top of BMEG - TumorMap
Use evidence graphs to find “patients like me”
pCHIPs: Network-based selection of targets and target combinations. Portal of patient-specific Hallmarks

Dexter Pratt, Trey Ideker - NDEx
“The Cytoscape Cloud”
Public commons where users can store, share, and publish their networks
- Hallmark cancer pathways via social organization and automation
    - Social pathway to enable the creation of molecular pathways
- Networks in next-gen cancer genome analysis
- Publishing of interactive networks on journal websites
- Scaling the framework for a growing community

“RAS Machine” - from the Sorger Lab
Use NDEx to share pre-publication pathways with reviewers and then reference in the article
ndexr: Access from R / Bioconductor
CyNDEx - used to transfer networks between NDEx and Cytoscape
Network sets: Provide user-driven structure as NDEx grows to very large numbers of networks
Patient similarity networks
Q: What about the community rating networks? What about providing evidence?
A: Rating is in the grant…

Jing Zhu - UCSC Xena
Multiple public hubs: Pancan Atlas, ICGC, TCGA, PCAWG, GA2GH/Toil
Data integration occurs on the client -maintain security of the data
Supports most types of functional genomics data

Usage: 430 laptop hubs “phoned home” in April 2017
Browser: 6915 sessions in the past month; ave of 7 min/session
New whole genome data visualizations, driven by PCAWG project
New integrative views
Create a bookmark of view, to share with collaborators
Created a Xena Python package, enables use of Jupyter Notebooks
Integrations with MuPIT, TumorMap, UCSC Genome Browser
Working on ingesting data from the GDC
New adventure: single cell RNAseq visualization. E.g., 1.3M mouse brain cells
Q: Are the custom query languages going to converge? E.g., cBio has a different one.
A: Unclear…
Q: How do you access data from ICGC, GDC
A: Bulk download

Han Liang -- The Cancer Proteome Atlas
Reverse Phase Protein Arrays
Characterized >100k samples, >10k though TCGA
>8000 users, 90 countries

1. Build all in 1 software for data processing
2. Enhance web platform for patient RPPA data analysis
3. Web platform for cell line RPPA data analysis
4. Interaction with user community

MD Anderson Cell Lines Project - MCLP

Dongjun Chung - Literature Guided ID of Cancer Subtypes
Aim 1: improve pathway knowledge by integrating biomedical literature with multiple pathway databases (Medline abstracts)
Aim 2: Integrate with multiple genomic data sets
Aim 3: Test case - mucinous ovarian cancer subtype identification
BioLitDB: Web interface for literature mining
bayesGO: Bayesian bi-clustering to identify novel pathways using the ontology fingerprint data. https://github.com/dongjunchung/bayesGO
Ontology fingerprint: Association measures between genes and GO terms. Takes into account how often gene and GO term have been studied in the literature
Pathclust: identify patient subgroups and key molecular features. Currently extending to support joint analysis of multiple genomic data types and simultaneous utilization of multiple pathway databases

Alexander Krasnitz, Computational framework for single cell genomics of Tumors

Conventional pathology: 65% probability that any 2 pathologists will disagree by at least 1 Gleason level
Single-Cell Genome Viewer - now released
Docker image for the upstream pipeline is coming soon
Q (slide): Is machine learning from images possible, with genomics as ground truth?

Jingshan Huang - OmniSearch
Semantic integration and search software tool designed for microRNA cancer research
Aim 1: Create miRNA domain ontologies
Aim 2: Use the ontology to annotate existing data sources
Aim 3: Evaluate
121 evaluators, 19 institutions
Google-based survey

Chakra Chennubhotla and Brion Carachan - Tools for Tumor Heterogeneity in Multiplexed Fluorescence Images
Heterogeneity Metrics
- Shannon Index
- Quadratic Entropy
- Pointwise mutual information
- Other

Global heterogeneity vs local heterogeneity
THRIVE: Tumor heterogeneity research interactive visualization environment
- Open software framework
- Compatible with standard microscopes
- Easy to contribute algorithms
- Easy to contribute datasets
- Actively used by the research community

Architecture heavily leveraging Docker
Can run locally or on a cloud
Q: How will you demonstrate correlation with outcomes?
A: Making progress in correlating metrics with outcomes, using a dataset from GE

Trinity Urban - LesionTracker
Some challenges:
- Demand for same day turn around
- More varied and complex tumor metrics criteria

Architecture
- Cornerstone
- Dicomeparser

- MongoDB

53% is funded by the NCI U24

Also:

- Commercial vendors
- Students
- Hospital research groups

Precision Imaging Metrics - Integration with LesionTracker

- Workflow mgmt
- Data management
- Image Management
- Image Analysis

PIM: two integrated applications

1) Workflow and data management
2) Image analysis platform Current: Clear Canvas; Future: LesionTracker

Current PIM usage: DFCI, Yale, FHCRC, Huntsman, <...>

David Gutman - The Digital Slide Archive

http://cancer.digitalslidearchive.net/

After last year, spent several months Dockerizing everything

Every operation in the UI is driven by a RESTful API

Slide management system

HistomicsTK - co-developed with Kitware

Working with TCGA Sarcoma Analysis WG - histomic-genomic analysis of nuclear pleomorphism. And PanCan Heterogeneity WG

Future development

- UI for managing slide sides
- Analyzing imaging analysis outputs
- Documentation