Towards a Pathology Imaging Data Commons for Discovery and Precision Oncology

AACR SY32: Integrative Data Science for the Precision Medicine Era
April 17th 2018

Michael J. Becich, MD PhD, FACMI
Chair, Department of Biomedical Informatics
University of Pittsburgh School of Medicine
University Distinguished Professor,
Associate Vice Chancellor for Informatics
Associate Director for Cancer Informatics, Hillman Cancer Center
NCI Board of Scientific Advisors
Becich Conflicts of Interest (& Disclaimer)

- **Spatial Diagnostics, Inc. or SPDx** (founder and stock) – computational pathology newco
- **Nexi** – Newco by Rebecca Jacobson and Text Information Extraction System (TIES) Cancer Research Network (TCRN) team (licensing revenues to my Department)
- Cancer Center Consultancies and EABs – Baylor, Cancer Institute of New Jersey, Indiana University, Northwestern Lurie Cancer Center, University of Colorado, UCLA, University of Michigan, University of New Mexico and Wake Forest
- CTSA Consultancies and EABs – numerous (not a conflict for this talk except possibly for U of Chicago Institute for Translational Medicine)
- **Funding** – CDC/NIOSH, NCATS, NCI, NHGRI, NHLBI & NLM

**Disclaimer** – I am a member of NCI’s Board of Scientific Advisors and Frederick National Laboratories Advisory Committee, Technical Working Group
Outline

• Introduction to Text Information System (TIES) Cancer Research Network (TCRN)
• Introduction to Whole Slide Imaging (WSI) and Computational Pathology (Comp Path)
• Introduction to Pathology Imaging Data Commons Enabled by TCRN

Key Impact Areas – Comp Path Predictive Analytics for Precision Oncology leads to better health and discovery science!

http://cancerdatanetwork.org/
The Text Information Extraction System or TIES
NCI ITCR funded effort (U24 CA180921)

• A natural language processing (NLP) and Information Retrieval system for de-identifying, annotating, storing & retrieving pathology (and radiology) reports

• A system for indexing research resources (clinical data, biospecimens & images) with document annotations

• An GUI for querying large repository of annotated documents and obtaining resources locally, using an honest broker model

• A platform to support phenotype, images and biospecimen sharing among networks of cancer centers and other institutions

http://ties.dbmi.pitt.edu/
TIES Cancer Research Network (TCRN)

Figure 3. TCRN technical architecture enabling intra- and inter-communication of components using wrapping with microservices.

Identified Data Behind HIPAA Firewall

- Clinical Data Sources: Discrete, Text, Images
- ETL Process
- HL7 Service
- Structured Data Service
- Custom Import Service
- Image Import Service

De-identified Data in DMZ

- Indexing Service
- NLP Service
- Text Annotation Service
- De-id Service
- Indexes
- Annotations
- Research ("open") Data Stores
- Feature Extraction Pipelines

TCRN Network + TCIA

- Federated Authentication and Data Movement
- Annotation Service
- Centralized Authorization Service
- Publish/Subscribe Services
- Users
- Public Data

TCRN Network + TCIA

- Pitt
- Augusta
- Penn
- RPCI
- Stony Brook
- Jefferson
- TCIA

http://cancerdatanetwork.org/
TIES Cancer Research Network (TCRN)

- UPMC Hillman Cancer Center (lead)
- Augusta University Cancer Center
- Abramson Cancer Center (Penn)
- Roswell Park Cancer Institute
- Stonybrook University (new partner)
- Sidney Kimmel Cancer Center (TJU)
- 12 Cancer Centers interested in joining

Network Trust Agreements
- IRBs agree that use of data for investigators is Not Human Subjects Research, no need for an additional IRB protocol even to access record level de-identified data
- Governance
- Agreement to abide by SOPs
- Instrument of Adherence

Soliciting new WSI “ready” partners!

http://cancerdatanetwork.org/
Cancer Research

A Federated Network for Translational Cancer Research Using Clinical Data and Biospecimens

Rebecca S. Jacobson, Michael J. Becich, Roni J. Bollag, Girish Chavan, Julia Corrigan, Rajiv Dhir, Michael D. Feldman, Carmelo Gaudioso, Elizabeth Legowski, Nita J. Maihle, Kevin Mitchell, Monica Murphy, Mayurapiyan Sakthivel, Eugene Tseytlin, and JoEllen Weaver

Abstract

Advances in cancer research and personalized medicine will require significant new bridging infrastructures, including more robust bio-repositories that link human tissue to clinical phenotypes and outcomes. In order to meet that challenge, four cancer centers formed the Text Information Extraction System (TIES) policies, and procedures, enable regulatory compliance. The TIES Cancer Research Network now provides integrated access to investigators at all member institutions, where multiple investigator-driven pilot projects are underway. Examples of federated search across the network illustrate the potential impact on...
Table 2. TCRN case statistics for numbers of patients and cases (A) and the number of cases of rare tumors (B) and common cancer categories (C) based on final diagnosis.

<table>
<thead>
<tr>
<th></th>
<th>GRU</th>
<th>RPCI</th>
<th>ACC</th>
<th>UPCI</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Case statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>76,404</td>
<td>72,376</td>
<td>465,717</td>
<td>1,840,156</td>
<td>2,454,653</td>
</tr>
<tr>
<td>Pathology cases</td>
<td>157,316</td>
<td>156,555</td>
<td>857,681</td>
<td>4,588,017</td>
<td>5,759,569</td>
</tr>
<tr>
<td><strong>B. Rare tumors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenoid cystic carcinoma</td>
<td>41</td>
<td>88</td>
<td>404</td>
<td>509</td>
<td>1,042</td>
</tr>
<tr>
<td>Adrenocortical carcinoma</td>
<td>5</td>
<td>20</td>
<td>59</td>
<td>63</td>
<td>147</td>
</tr>
<tr>
<td>Alveolar soft part sarcoma</td>
<td>3</td>
<td>15</td>
<td>10</td>
<td>25</td>
<td>53</td>
</tr>
<tr>
<td>Angioimmunoblastic lymphadenopathy</td>
<td>12</td>
<td>35</td>
<td>58</td>
<td>84</td>
<td>189</td>
</tr>
<tr>
<td>Chordoma</td>
<td>5</td>
<td>14</td>
<td>124</td>
<td>245</td>
<td>388</td>
</tr>
<tr>
<td>Follicular dendritic cell sarcoma</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>Merkel cell carcinoma</td>
<td>9</td>
<td>72</td>
<td>165</td>
<td>196</td>
<td>442</td>
</tr>
<tr>
<td>Ovarian granulosa cell tumor</td>
<td>4</td>
<td>10</td>
<td>23</td>
<td>48</td>
<td>71</td>
</tr>
<tr>
<td>Phaeochromocytoma</td>
<td>15</td>
<td>38</td>
<td>272</td>
<td>164</td>
<td>489</td>
</tr>
<tr>
<td>Pleomorphic xanthoastrocytoma</td>
<td>2</td>
<td>5</td>
<td>12</td>
<td>12</td>
<td>27</td>
</tr>
<tr>
<td>Pseudomyxoma peritonei</td>
<td>6</td>
<td>36</td>
<td>46</td>
<td>129</td>
<td>217</td>
</tr>
<tr>
<td>Rhabdomyosarcoma</td>
<td>34</td>
<td>70</td>
<td>86</td>
<td>270</td>
<td>460</td>
</tr>
<tr>
<td>Sebaceous adenocarcinoma</td>
<td>13</td>
<td>33</td>
<td>26</td>
<td>94</td>
<td>166</td>
</tr>
<tr>
<td>Sinonasal undifferentiated carcinoma</td>
<td>2</td>
<td>6</td>
<td>31</td>
<td>27</td>
<td>66</td>
</tr>
<tr>
<td>Thymoma</td>
<td>13</td>
<td>45</td>
<td>433</td>
<td>210</td>
<td>701</td>
</tr>
<tr>
<td><strong>C. Common cancer categories</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bladder carcinoma</td>
<td>345</td>
<td>1,618</td>
<td>3,873</td>
<td>6,711</td>
<td>12,547</td>
</tr>
<tr>
<td>Breast carcinoma</td>
<td>1,143</td>
<td>9,605</td>
<td>28,262</td>
<td>37,691</td>
<td>76,701</td>
</tr>
<tr>
<td>Colorectal carcinoma</td>
<td>465</td>
<td>2,530</td>
<td>6,898</td>
<td>11,608</td>
<td>21,501</td>
</tr>
<tr>
<td>Endometrial carcinoma</td>
<td>394</td>
<td>1,815</td>
<td>3,707</td>
<td>7,706</td>
<td>13,622</td>
</tr>
<tr>
<td>Esophageal carcinoma</td>
<td>63</td>
<td>1,477</td>
<td>2,452</td>
<td>3,514</td>
<td>7,506</td>
</tr>
<tr>
<td>Hepatic carcinoma</td>
<td>153</td>
<td>633</td>
<td>2,912</td>
<td>5,720</td>
<td>9,418</td>
</tr>
<tr>
<td>Lung carcinoma</td>
<td>820</td>
<td>4,264</td>
<td>10,208</td>
<td>17,955</td>
<td>33,247</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>1,387</td>
<td>6,795</td>
<td>10,605</td>
<td>15,689</td>
<td>34,476</td>
</tr>
<tr>
<td>Malignant glioblastoma</td>
<td>242</td>
<td>292</td>
<td>2,198</td>
<td>4,943</td>
<td>7,675</td>
</tr>
<tr>
<td>Malignant melanoma</td>
<td>335</td>
<td>2,675</td>
<td>5,180</td>
<td>7,068</td>
<td>15,258</td>
</tr>
<tr>
<td>Ovarian carcinoma</td>
<td>503</td>
<td>2,872</td>
<td>4,659</td>
<td>6,446</td>
<td>14,480</td>
</tr>
<tr>
<td>Pancreatic carcinoma</td>
<td>162</td>
<td>740</td>
<td>1,866</td>
<td>3,622</td>
<td>6,390</td>
</tr>
<tr>
<td>Prostate carcinoma</td>
<td>903</td>
<td>3,612</td>
<td>18,867</td>
<td>19,445</td>
<td>42,827</td>
</tr>
<tr>
<td>Renal cell carcinoma</td>
<td>364</td>
<td>1,319</td>
<td>3,183</td>
<td>10,950</td>
<td>15,816</td>
</tr>
<tr>
<td>Thyroid carcinoma</td>
<td>474</td>
<td>1,236</td>
<td>7,681</td>
<td>12,387</td>
<td>21,778</td>
</tr>
</tbody>
</table>
Merkel Cell Carcinoma: A Virus-Induced Human Cancer

Yuan Chang1,3 and Patrick S. Moore2,3
Yuan Chang: yc70@pitt.edu; Patrick S. Moore: psm8@pitt.edu
1Department of Pathology, University of Pittsburgh, Pittsburgh, Pennsylvania 15213
2Department of Microbiology and Molecular Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15213
3Cancer Virology Program, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania 15213

Abstract

Merkel cell polyomavirus (MCV) is the first polyomavirus directly linked to human cancer, and its recent discovery helps to explain many of the enigmatic features of Merkel cell carcinoma (MCC). MCV is clonally integrated into MCC tumor cells, which then require continued MCV oncprotein expression to survive. The integrated viral genomes have a tumor-specific pattern of tumor antigen gene mutation that incapacitates viral DNA replication. This human cancer virus provides a new model in which a common, mostly harmless member of the human viral flora can initiate cancer if it acquires a precise set of mutations in a host with specific susceptibility factors, such as age and immune suppression. Identification of this tumor virus has led to new opportunities for early diagnosis and targeted treatment of MCC.

Survey for human polyomaviruses in cancer

Tuna Toptan,1 Samuel A. Yousem,2 Jonhan Ho,1 Yuki Matsushima,4 Laura P. Stabile,5 Maria-Teresa Fernández-Figuera,6 Rohit Bhargava,7 Akihide Ryo,8 Patrick S. Moore,7 and Yuan Chang1

1Cancer Virology Program, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, USA.
2Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA.
3Department of Dermatology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA.
4Division of Virology, Kawasaki City Institute for Public Health, Kanagawa, Japan.
5Department of Pharmacology and Chemical Biology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA.
6Department of Pathology, Hospital Universitari Germans Triás i Pujol, Universitat Autònoma de Barcelona, Barcelona, Spain.
7Mage-Wiemens Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA.
8Department of Microbiology, Yokohama City University School of Medicine, Kanagawa, Japan.

Over the past 8 years, the discovery of 11 new human polyomaviruses (HPyVs) has revived interest in this DNA tumor virus family. Although HPyV infection is widespread and largely asymptomatic, one of these HPyVs, Merkel cell polyomavirus (MCV), is a bona fide human cancer virus that induces a distinctive form of skin cancer, Merkel cell carcinoma.
Adding Cancer Registry Data to TIES Outcomes Annotation

- Identified as a valued development target by users
- We have secured additional funding from Institute for Precision Medicine in Pittsburgh
- Mike Davis and Jonathan Silverstein lead this effort
- Starting with breast cancer (first publication in review)
- Immediately leveragable by existing TCRN nodes by adding Cancer Registry data to TIES instances

http://cancerdatanetwork.org/
## Cancer Registry Data Elements

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Primary</th>
<th>Treatment</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race</td>
<td>Primary Site</td>
<td>Surgery</td>
<td>Vital Status</td>
</tr>
<tr>
<td>Gender</td>
<td>Histology</td>
<td>Chemotherapy</td>
<td>Cancer Status</td>
</tr>
<tr>
<td>Age @ Diagnosis</td>
<td>Grade</td>
<td>BRM</td>
<td>Recurrence</td>
</tr>
<tr>
<td>Smoking</td>
<td>Path TNM</td>
<td>Hormonal</td>
<td>Cause of Death</td>
</tr>
<tr>
<td>Alcohol</td>
<td>Clinical TNM</td>
<td>Immunotherapy</td>
<td></td>
</tr>
<tr>
<td>Prognostic Factors</td>
<td>BRM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enabling Research on the Cancer Registry
Validation of TIES/TCRN Cancer Registry Integration

- Kaiser team publication – 2017 J. Path Info article:


Applying of Text Information Extraction System for Real-Time Cancer Case Identification in an Integrated Healthcare Organization

Fagen Xie, Janet Lee, Corrine E. Munoz-Plaza, Erin E. Hahn, and Wansu Chen

1Department of Research and Evaluation, Kaiser Permanente Southern California Medical Group, Pasadena, CA, USA

Address for correspondence: Dr. Fagen Xie, Department of Research and Evaluation, Kaiser Permanente Southern California Medical Group, 100 S Los Robles Ave, 2nd Floor, Pasadena CA 91101, USA. E-mail: fagen.xie@kp.org

Received 2017 Jul 28; Accepted 2017 Nov 1.

Copyright: © 2017 Journal of Pathology Informatics

- Result = deeper patient annotation, including outcomes data and enabling study cohort identification

http://cancerdatanetwork.org/
# TIES and the TIES Cancer Research Network

## TIES Team

<table>
<thead>
<tr>
<th>Girish Chavan</th>
<th>Adi Nemlekar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eugene Tseytlin</td>
<td>Yining Zhao</td>
</tr>
<tr>
<td>Kevin Mitchell</td>
<td>Vanessa Benkovich</td>
</tr>
<tr>
<td>Julia Corrigan</td>
<td>Liron Pantanowitz</td>
</tr>
<tr>
<td>Liz Legowski</td>
<td>Rajiv Dhir</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Roswell Park</th>
<th>GRU</th>
<th>Penn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carmelo Gaudioso</td>
<td>Roni Bollag</td>
<td>Michael Feldman</td>
</tr>
<tr>
<td>Monica Murphy</td>
<td>Samir Khleif</td>
<td>Nate DiGiorgio</td>
</tr>
<tr>
<td>Mayurapriyan Sakthivel</td>
<td>Jennifer Carrick</td>
<td>Tara McSherry</td>
</tr>
<tr>
<td>Amanda Rundell</td>
<td>Nita Maihle</td>
<td>Joellen Weaver</td>
</tr>
<tr>
<td></td>
<td>And more....</td>
<td></td>
</tr>
</tbody>
</table>

## Funding

NCI U24 CA180921 Enhanced Development of TIES

Led by
Rebecca Jacobson, MD MSIS

http://ties.dbmi.pitt.edu/tcrn/
Outline

• Introduction to Text Information System (TIES) Cancer Research Network (TCRN)

• Introduction to Whole Slide Imaging (WSI) and Computational Pathology (Comp Path)

• Introduction to Pathology Imaging Data Commons Enabled by TCRN

**Key Impact Areas** — Comp Path Predictive Analytics for Precision Oncology leads to better health and discovery science!
Traditional vs. Digital Pathology
Intro to Whole Slide Imaging (WSI)

From Bertram and Kopfleish, Vet Path 2017
WSI – Imaging Big Data Drives Computational Pathology

Pipeline for Whole Slide Feature Characterization

- $10^{10}$ pixels for each whole slide image: $10^5 \times 10^5$
- 10 whole slide images per patient
- $10^8$ image features per whole slide image
- $10^{15}$ pixels
- $10^{13}$ features

From Saltz, circa 2011, to NLM

FDA News Release

FDA allows marketing of first whole slide imaging system for digital pathology

For Immediate Release April 12, 2017

1000 scanned slides
3.03 Tbyte
444 registrants
WSI Comp Path Adds Spatial Big Data for Tumor Analysis

H&E stained whole slide images from FFPE tumor sample

Multi to hyperplexed fluorescence imaging of whole slide images for higher spatial resolution and tissue context

Thanks to Drs. Joe Ayoob and Chakra Chennubhotla for this illustration, 2017
Outline

• Introduction to Text Information System (TIES) Cancer Research Network (TCRN)
• Introduction to Whole Slide Imaging (WSI) and Computational Pathology (Comp Path)
• Introduction to Pathology Imaging Data Commons Enabled by TCRN

**Key Impact Areas** — Comp Path Predictive Analytics for Precision Oncology leads to better health and discovery science!
Towards Computational Pathology Imaging Commons for Cancer – TCRN Data Types

• Clinical Phenotype – From Pathology Lab Info Sys
  – Structured data
  – Unstructured data via NLP (Text Info Extract Sys – TIES)

• Outcomes Data – From Cancer Registry Systems

• Biobank Data – From Formalin Fixed Paraffin Embedded Tissue

• Imaging – WSI - Spatial Annotation – HTAN relevant

• Integrate Challenge – Clinical, Outcome, Biobank, Imaging Data for analysis with Genomic data

http://cancerdatanetwork.org/
Computationally Annotated WSI Assist in Quantification of Tumor Heterogeneity

Intra-tumoral spatial heterogeneity complicates accurate diagnosis & prognosis

Tumor Heterogeneity Research Interactive Visualization Environment (THRIVE – NCI ITCR U01 @ Pitt)

- Computational Pathology
  - Hyperplex Immunofluorescence (9 to > 50 Ab)
  - Machine Learning + Spatial Statistics
  - Network Systems Biology
  - Partnership with General Electric (GE)

- Iterative experimental–computational tumor micro-environment studies

Courtesy Chakra Chennubhotla, 2017
NCI ITCR funded effort (U01 CA204826)
Platform for Quantitative Evaluation of Spatial Intratumoral Heterogeneity in Multiplexed Fluorescence Images

Daniel M. Spagnolo¹,², Yousef Al-Kofahi³, Peihong Zhu³, Timothy R. Lezon²,⁴, Albert Gough²,⁴, Andrew M. Stern²,⁴, Adrian V. Lee⁵,⁶, Fiona Ginty⁷, Brion Sarachan³, D. Lansing Taylor²,⁴,⁵, and S. Chakra Chennubhotla²

Abstract

We introduce THRIVE (Tumor Heterogeneity Research Interactive Visualization Environment), an open-source tool developed to assist cancer researchers in interactive hypothesis testing. The focus of this tool is to quantify spatial intratumoral heterogeneity (ITH), and the interactions between different cell phenotypes and noncellular constituents. Specifically, we foresee applications in phenotyping cells within tumor microenvironments, recognizing tumor boundaries, identifying degrees of immune infiltration and epithelial/stromal separation, and identification of heterotypic signaling networks underlying microdomains. The THRIVE platform provides an integrated workflow for analyzing whole-slide immunofluorescence images and tissue microarrays, including algorithms for segmentation, quantification, and heterogeneity analysis. THRIVE promotes flexible deployment, a maintainable code base using open-source libraries, and an extensible framework for customizing algorithms with ease. THRIVE was designed with highly multiplexed immunofluorescence images in mind, and, by providing a platform to efficiently analyze high-dimensional immunofluorescence signals, we hope to advance these data toward mainstream adoption in cancer research. Cancer Res. 77(21): e71–74. ©2017 AACR.

Cancer Res. 2017 Nov 1;77(21):e71-e74. doi: 10.1158/0008-5472. PMID: 29092944

NCI ITCR funded effort (U01 CA204826)
TCRN Enabled Data Sharing and Query

Figure 2. TCRN innovation in policy and technology - illustrated by use case examining TILs in ovarian cancer

Are there specific features of the TILs, such as spatial distribution, that influence survival?

1) Request permission from TCRN members to open a new study
2) Contact Biospecimen and Pathology Core Directors of Joint RPCI-Hillman SPORE to request WSI of high-grade serous tumors from 200 patients under protocol at both RPCI and Hillman

TCRN UPLOAD

NEW TCRN DATA COLLECTIONS
- RPCI WSI
- Hillman WSI
- RPCI Clinical Data
- Hillman Clinical Data

TIES DE-IDENTIFICATION SOFTWARE
TIES NLP SOFTWARE

OTHER TCRN MEMBER COLLECTIONS

Images made available for widespread use by the Cancer Research Community

1) Survival relates to positioning of TILs
2) Survival relates to the absolute number of responding tumor nodules
Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images

Authors
Joel Saltz, Rajarsi Gupta, Le Hou, ..., Alexander J. Lazar, Ashish Sharma, Vésteinn Thorsson

Correspondence
joel.saltz@stonybrookmedicine.edu (J.S.), vesteinn.thorsson@systemsbiology.oi (V.T.)

Highlights
- Deep learning based computational stain for staining tumor-infiltrating lymphocytes (TILs)
- TIL patterns generated from 4,759 TCGA subjects (5,202 H&E slides), 13 cancer types
- Computationally stained TILs correlate with pathologist eye and molecular estimates
- TIL patterns linked to tumor and immune molecular features, cancer type, and outcome

Cell Reports, 2018

Tumor-infiltrating lymphocytes (TILs) were identified from standard pathology cancer images by a deep-learning-derived “computational stain” developed by Saltz et al. They processed 5,202 digital images from 13 cancer types. Resulting TIL maps were correlated with TCGA molecular data, relating TIL content to survival, tumor subtypes, and immune profiles.
Conclusions

• The Text Information System (TIES) Cancer Research Network (TCRN) is an ideal platform for coupling clinical/outcomes data to biospecimens/images

• Whole Slide Imaging (WSI) and Computational Pathology will enable NCI’s Moonshot in new ways to couple spatial analysis features to –omic data sets

• A WSI Imaging Data Sharing and Query Toolkit (TCRN) with linkages to NCI ITCR and GDC programs can assist the NCI vision for a Cancer Research Data Commons

Key Impact Areas – Computational Pathology Predictive Analytics for Precision Oncology leads to better health and discovery science!

For Copy of PPT – becich@pitt.edu
Computational Pathology @ Pitt
Led by Chakra Chennubhotla, PhD

- Organized as a Interest Group and Lecture Series
- Currently 109 members from Pitt, CMU, UPMC and regional companies/startups (*need sponsor!!!*)
- Pittsburgh Computational Pathology Interest Group - [http://www.csb.pitt.edu/comppath/](http://www.csb.pitt.edu/comppath/)
- Computational Pathology Lecture Archive - [https://www.youtube.com/channel/UCWfBS3PLWHTIACeccm2Clog](https://www.youtube.com/channel/UCWfBS3PLWHTIACeccm2Clog)
- Supported by Akif Burak Tosun, PhD (post-doc)
Computational Pathology Lecture Archive -
https://www.youtube.com/channel/UCWfBS3PLWHTIAccm2Clog
“...Data commons collocate data, storage, and computing infrastructure with core services and commonly used tools and applications for managing, analyzing, and sharing data to create an interoperable resource for the research community...”
FAIR Data Principles

- Findable
- Accessible
- Interoperable
- Reusable

Data Use Case for Data Commons

— See Data Med v3.0 — http://datamed.org
Six Requirements of a Data Commons

- Permanent digital IDs (data and knowledge)
- Permanent metadata (data describing data)
- API (interface)-based access (interoperability)
- Data portability (standard containers)
- Data Peering (commons 1 can access commons 2)
- Pay for compute (allocate computing/charging)
  - Demand higher than computing resources available

*From Grossman et al 2016 CISE IEEE*
Pitt’s Department of Biomedical Informatics is a Center of Excellence in Big Data to Knowledge
Informatics Fuels Data to Knowledge

DIKW paradigm today

Adding understandable data to the stack

Wisdom

Ability to use knowledge and experience to make good decisions and judgments

Knowledge

Information and understanding that humans have in mind

Information

Facts about a situation, person, event, etc.

Data

Information in the form of text, numbers, or symbols that can be used by a computer

Semantic metadata

Knowledge in the form of data that a computer can autonomously use
The Human Tumor Atlas Network (HTAN)

**Goal:** Pilot-scale, high-priority human tumor atlases that facilitate basic and clinical scientific discovery regarding important transitions during tumorigenesis.

**Components of the HTAN:**
1. **Human Tumor Atlas (HTA) Research Centers** (U2C) focused on construction of dynamic 3D tumor atlases. 
   - RFA-CA-17-034 (closed)
2. **Pre-Cancer Atlas Research Centers** (U2C) focused on characterization of pre-malignant lesions. 
   - RFA-CA-17-035 (closed)
3. **Coordinating Center** (U24) focused on integration of the HTAN through administrative and scientific support. 
   - RFA-CA-17-036 (closed)

**Human Tumor Atlas Network**

**High Priority Tumors HTA Research Centers:**
- *(Non)Responsive to immunotherapy*
- *Highly metastatic*
- *High-risk hereditary*
- *Pediatric*

**Tumor Criteria for PCA Research Centers:**
- *Public Health Impact*
- *Access*
- *Feasibility*
- *Partnerships*

**Transitions:**
- Pre-cancer → Cancer
- Invasive → Metastatic
- Responsive → Resistant
Human Cell Atlas & Human BioMolecular Atlas Program (HuBMAP) Drive Relevance of Comp Path

**BRAIN**
These projects seek to continue the development of technologies, many from the neurobiology field, that address the specific challenges of single-cell analysis of primary human tissues. They are also benchmarking existing tools and methods against each other and across regions of the nervous system.

SEE THE 7 PROJECTS

**TISSUE HANDLING & PROCESSING**
These projects seek to establish best practices for processing heart tissues and other tissues that are notoriously difficult to process. Studies will test multiple protocols and methods on multiple tissues and deposit data from common sources for robust benchmarking.

SEE THE 10 PROJECTS

**IMMUNE**
These projects look at the multi-modal analysis of current methods of the immune system to develop richer single-cell analysis, evaluate inter-individual variation, more reliably detect cell types of known rarity, and compare dissimilar immune cell populations.

SEE THE 7 PROJECTS

**SKIN**
Human skin has a wide range of cell types that exist in a relatively well described spatial distribution. These projects will use this empirical knowledge to generate benchmarks across assays that reveal cell type bias and also inform tissue isolation and preservation techniques.

SEE THE 2 PROJECTS

**GASTROINTESTINAL**
These projects study diverse approaches to preservation and processing of tissues from often sensitive organs that are critical for nutrient uptake and metabolism. Relatively little is known about the diversity of cells that comprise these tissues and differences among their distinct spatial compartments.

SEE THE 6 PROJECTS

**TECHNOLOGY DEVELOPMENT**
These projects will develop technology and computational tools that can increase the scale of single-cell analysis, identify systematic errors among collections, and increase detection signals for sequencing and imaging.

SEE THE 6 PROJECTS

---

**The Human BioMolecular Atlas Program**

**For the Public**
- Public Health Relevance
- In the News
- Strategic Planning

**For Researchers**
- Frequently Asked Questions
- Funding Opportunities
- HuBMAP QTs
- NIH Working Group
- Meetings and Workshops

**Program Snapshot**
In organisms consisting of multiple cell types, diverse cells with different functions and structures develop as we grow and age. The organization and variability of these cells have a profound impact on the function of different tissues, process of aging, and emergence of diseases and conditions. Recently developed technologies are allowing researchers to explore cells on the individual (single cell) level and provide an opportunity to establish a census of cells in human tissues and to study functional interactions among single cells. The Human BioMolecular Atlas Program (HuBMAP) aims to facilitate research on single cells within tissues by supporting data generation and technology development to explore the relationship between cellular organization and function, as well as variability in normal tissue organization at the level of individual cells.
Enabling Research on the Cancer Registry

TIES tranSMART