The Pathology Informatics Roadmap for Implementing Cancer Personalized Medicine

Michael J. Becich, MD PhD - becich@pitt.edu
Chairman, Department of Biomedical Informatics
http://www.dbmi.pitt.edu
University of Pittsburgh School of Medicine

† Input from Personalized Medicine and Next Generation Sequencing Task Forces at University of Pittsburgh Cancer Institute (UPCI)
Disclosures of COI for 2012 for MJB

- **Corporate Support for API, Strategic Summit and Pathology Informatics 2012**

- **Corporate Sponsored Research – ZERO (1st time in 15 years!!!)**

- **Startup/Public Companies (Consulting, Royalties/Licensing, Stock - MJB):**
  - De-ID Data Corp – de-identification software (licensing agreement) [http://www.de-idata.com/]
  - Empire Genomics - Scientific Advisory Board [http://www.empiregenomics.com]
  - Omnyx – Joint Venture with UPMC and GE [http://www.omnyx.com]
  - iKaryos Diagnostics – Scientific Advisory Board [http://www.ikaryos.com/]
  - NinePoint Medical – Scientific Advisory Board [www.ninepointmedical.com]

- **Consultancy (honorarium)**
  - Cancer Center Consulting – MD Anderson, Karmanos Cancer Center, Moffitt Cancer Center, NFGC, Penn State CC, Roswell, U Buffalo, UMDNJ, U Colorado, VCU
  - CTSA Consulting – Duke, Emory, MCW, Northwestern, UAK, UC Davis, UCLA, U Chicago, U Cincinnati, U IN, U KY, UC Davis, UMich, UMn, UNC, UNM, UWI and Wash U
  - Pathology – GoPath DX (honorarium)
Personalized Medicine Task Force

University of Pittsburgh Cancer Institute effort by MJB and:

• Adrian Lee, PhD – Genomics of Breast Cancer
• Scot Stevens – CIO Cancer Center
• Mike Davis – Senior Programmer

• Next Generation Sequencing (NGS) Task Force
  – Rama Gullapalli, MD PhD – Computational Pathology Fellow
  – Jeff Kant, MD PhD – Director of Molecular Diagnostics Lab
  – Ketaki Desai, PhD – Research Associate, DBMI
  – Kicked Off with “Cancer Genomics” Symposium March 2011
Speakers from 2011 UPCI Symposium

Personalized Genomic Medicine
Technologies, Training & Clinical Applications

Mark S. Boguski, M.D., Ph.D., F.C.A.P.
Center for Biomedical Informatics, Harvard Medical School
Department of Pathology, BIDMC
March 10, 2011

Personalized Genomic Analyses of Human Cancer

Victor Velculescu, M.D., Ph.D.
Ludwig Center for Cancer Genetics and Therapeutics
Johns Hopkins Kimmel Cancer Center

UPCI Symposium
March 10, 2011

Recent Advances in the Analysis of Cancer Genomes

Richard K. Wilson, Ph.D.
Washington University
School of Medicine

Structural Rearrangements in Breast Cancer Genomes

Adrian V. Lee, Ph.D.
Visiting Professor,
Dept. Pharmacology and Chemical Biology
Director, Women's Cancer Research Center
Goals for today’s discussion

- Identifying role of Path Info and the LIS in implementing cancer personalized medicine
- Implementing an Enterprise Analytic Data Warehouse and the Role of Pathology Informatics
- Next Generation Sequencing and impact on cancer personalized medicine and Pathology
- Making the Case for “Computational Pathologists” – a Partnership Between Pathology and Biomedical Informatics
Introduction

• How will Pathology Informatics and the LIS impact personalized cancer care?

• Key areas of evolution needed in LIS:
  – Personalized Medicine and the LIS – Will Require the Next Generation of Warehousing of Data, Tissue/Blood and DNA
  – LIS role in “Real” Decision Support
    • No effective EMRs for Cancer Care – “Actionable” Care
    • No integration of disparate data types (outcomes/CER)
  – Molecular Path–Role of Next Generation Sequencing
Two Interdependent Worlds…

Pathology Informatics

Bioinformatics & Comp/Sys Bio

This is the realm (and the need for) of “computational pathology”

Disease World

LIS

Diagnosis
Histological Subtypes
Grading
Staging
LN Involvement
Margins
Localization and Size
Molecular Markers
Sequence
SNPs
Methylation Status
Biospecimens Inventory
Genomic DNA Store
Disease DNA Store

OMIM
Clinical Synopsis

Genome
Proteome
Interactome
Metabolome
Physiome
Pathome

Variome

Pharmacogenom

With Some Data Exchange…

Modified from Anil Jegga
Division of Biomedical Informatics,
Cincinnati Children’s Hospital Medical Center
Department of Pediatrics, U of Cincinnati
http://anil.cchmc.org
Personalized Medicine and Theranostics

- Theranostics is applying the power of pathology to predict therapeutic response
- Determining whether a treatment is working
- Monitoring healthy people to detect early signs of disease
- Producing safer drugs by predicting potential for adverse effects earlier
- Targeting groups of people most likely to benefit from a drug, while keeping its use from those who may be harmed by it
- Producing better medical products
- Ready access to information
- Decreasing health care costs

Modified from: From Schwartz, CAP, 2009 Emerging Developments and Your Future in Pathology
Personalized Medicine Requires Biomedical Informatics Science

- **Development of a Analytical Data Warehouse is key!!!**
  - This will require biomedical informatics expertise in several areas:
    - Expertise in multidimensional **database design and query** (computer science collaboration?)
    - **Natural Language Processing** of text (H&P, Consults, Discharge, etc..)
    - Structured capture of key medical data will require controlled **vocabularies** and implementation of a cancer **ontology** (EVS of NCI?)
    - **De-identification** of text for sharing with researchers (De-ID Data Corp)
    - Implementation of **decision support** algorithms to make genomic data “actionable” (Bayesian methods/artificial intelligence)
    - **Human-computer interaction** (usability) studies to ensure user friendly design and query for reduction in practice for clinicians
  - This is only for the ADW, for analyzing NGS data you will need significant **bioinformatics** expertise (faculty, staff and trainees).

*(Hood, 2011–1/3 of 120 staff at the Institute for Systems Biology– see www.systemsbio.org)*
Proposed Cancer Personalized Medicine Information Management System

1. Clinical Data Extraction
   - Ontology
   - Harmonization
   - Synoptic reporting
   - Dx/Staging/Response input

2. Enterprise Analytic Data Warehouse
   - Omics Data Bank (ODB)
   - Integration Analysis
   - Clinical Data Center (CDC)
   - Governance\Security\Tracking

3. Clinical Decision Making
   - Comparative Effectiveness Analysis
     - Delivery via:
       1. Pathways
       2. Oncology EMR
       3. CRMS

Vendor Partnership

Cancer Center HIT

Cancer Center HIT
Activities Underway for EADW:

- **Harmonization** of data elements from disparate systems into a unified data model & ontology
- **Extraction of standardized clinical data** into a common data warehouse
- **Extraction of standardized “omics” data** into the common data warehouse
- **Integration of clinical and omics data** linking key data elements
- **Comparative effectiveness analysis of genetic testing** as predictors of key clinical outcomes as well as cost effectiveness of possible therapeutic options
- **Ongoing outcomes analysis** to use data from the integrated data warehouse to confirm the predictions made by the clinical decision making or to define more targeted therapies or new studies
Drugs/Genes Linked in Oncology

- **Drugs:** Arsenic Trioxide, Brentuximab, Vedotin, Busulfan, Capecitabine, Cetuximab (2), Crizotinib, Dasatinib, Erlotinib, Fulvestrant, Gefitinib (2), Imatinib (4), Irinotecan, Lapatinib, Mercaptopurine, Nilotinib (2), Panitumumab (2), Rasburicase, Tamoxifen, Thioguanine, Tositumomab, Trastuzumab, Vemurafenib
  - Early Inventory = 30 FDA Approved Drugs*

- **Genes:** PML/RARa, CD30, Ph Chromosome, DPD, EGFR, KRAS, ALK, ER receptor, CYP2D6, C-Kit, PDGFR, FIP1L1-PDGFRa, UGT1A1, Her2/neu, TPMT, UGT1A1, KRAS, G6PD, ER receptor, TPMT, CD20 antigen, Her2/neu, BRAF
  - Early Inventory = 24 genes

From Gullapalli et al., 2011 (accepted JPI)

* numbers in parentheses are numbers of genetic polymorphisms affecting the drug

Bioinformatics Team, Molecular Pathology and Librarian in our Biomedical Informatics PhD Program Curating Drug Candidates for Public Sharing
New Technologies and Their Impact

• Next Generation Sequencing (NGS)
  – Massively parallel sequencing is here and 2\textsuperscript{nd} generation machines (ABI, Illumina, etc…) are being replaced by 3\textsuperscript{rd} generation technology (Ion Torrent, PacBio, etc…)
  – Implications are that an entire human genome can be done in a laboratory in about a day to a month for about $10K (depending on application)
  – Goal with new technology is hours to days and about $1K!!!
Cancer genomes contain a lot of genetic damage.

- Many of the mutations in cancer are incidental.
- Initial mutation disrupts the normal DNA repair/replication processes.
- Corruption spreads through the rest of the genome.

Today: Find the “driver” mutations amongst the thousands of “passengers.”

- Identifying the driver mutations will give us new targets for therapies.

Tomorrow: Analyse the cancer genome of every patient in the clinic (now underway at Wash U).

- Variations in a patient and cancer genetic makeup play a major role in how effective a particular drugs will be.
- Clinicians will use this information to tailor therapies.
Personalized Medicine Will Rely Heavily on Pathology

- **Tumor diagnostics** are becoming increasingly important and complex
  - In our aging population patients survive with multiple cancer occurrences & die with their tumors not from them
- **Pathology itself is changing from primarily diagnostics to increasingly complex theranostics (personalized medicine)**
  - NGS will have a major impact on this area!!!
- **Pathology Informatics, particularly the LIS and it communication with bioinformatics tools/databases will be key to the implementation of personalized medicine**
  - Need to couple genotype with phenotype (clinical information) and outcomes (Comparative Effectiveness Research or CER).
As Medicine evolves from the treatment of illness to aggressive promotion of wellness – genomics/informatics will be key!!!!

DATA & SYSTEMS INTEGRATION
EVOLUTIONARY PRACTICES
DISTRIBUTED HIGH-THROUGHPUT ANALYTICS
REVOLUTIONARY TECHNOLOGIES

Automated systems
Information Correlation
1st generation diagnostics

HEALTHCARE TODAY

PERSONALIZED HEALTHCARE

Lifetime Treatment
Pre-symptomatic Treatment
CA-diagnosis
Molecular Medicine
Genetic Predisposition Testing
Clinical Genomics
Digital Imaging
Electronic Health Record

TRANSITIONAL MEDICINE

Episodic Treatment
Nonspecific (treat symptoms)
Organized (error-reduction)

Artificial Expert System
Personalized (disease prevention)

Computational Pathology will be the key enabler!!!

SOURCE: IBM LIFE SCIENCES SOLUTIONS

Modified from: From Schwartz, CAP, 2009
Emerging Developments and Your Future in Pathology
Barriers to Implementing NGS in Clinical Practice

• **Superiority of NGS to current Molecular Diagnostic Tools (Quantitative PCR, etc...) must be proven**
  – Collaboration with industry for testing, validating and implementing “stable and version controlled” hardware will be key

• **Bioinformatics analysis of NGS data must be “reduced” to practice**
  – Mostly done by research groups today, collaboration is key

• **Clinically certified laboratories with the proper technical infrastructure is also a critical barrier**
  – Infrastructure for storage of specimens, data mass storage, high speed network connectivity and supercomputing are all critically needed

• **R&D is Key - New U01 NHGRI grant to clinically implement sequencing to clinical practice – see**
The demise of expert-based practice is inevitable

- Functional Genetics: Gene expression profiles
- Structural Genetics: e.g. SNPs, haplotypes
- Proteomics and other effector molecules

Multiple Issues including:

- Reference Genomes – {huge problem}
- Tertiary Analysis (cancer specific)
- How to do QC on the analysis? (no standards)
Recommendation – Pathology Informatics & Personalized Medicine

• **Phase One**
  – Build Multidisciplinary (Dedicated) Team
    • Oncologist, Pathologists, Translational Researchers, Tissue Bankers and Pathology Informatics
    • Build Data Warehouse and NGS Strategy
  – Focus on Developmental Informatics Agenda
    • Secure Intramural (Health System) & Extramural Funding

• **Phase Two**
  – Implement Pilot Programs which involve:
    • Report Integration, Data Warehousing, Next Generation Sequencing and Decision Support
ClusterFASTQ: A Method for the Identification of Translocations in Clinical Next Generation Sequencing Data

Eric J. Duncavage¹, Haley Abel²

¹Department of Anatomic and Molecular Pathology, ²Division of Statistical Genetics, Washington University College of Medicine, St. Louis, MO. E-mail: eduncavage@path.wustl.edu

Progress to date:
- 250 Oncology Patient Tumors
- Reporting Workflow Established
- Not using a commercial solution
- Real opportunity for Pathology Informatics Units

Clinical Genomicist Workstation
Rakesh Nagarajan¹, Mukesh K. Sharma¹, Joshua Phillips², Saurabh Agarwal², Wesley S. Wiggins², Savita Shrivastava¹, Sunita B. Koul¹, Madhurima Bhattacharjee¹, Caerie D. Houchins¹, Raghavendra R. Kalakota¹, Bijoy George¹, Rekha R. Meyer¹, David H. Spencer¹, Christina M. Lockwood¹, TuDung T. Nguyen¹, Eric J. Duncavage¹, Hussam Al-Kateb¹, Catherine E. Cottrell¹, Suhasini Godala², Ravi T. Lokineni², Sameer M. Sawant², Vasudev Chatti², Suresh Surampudi², Raja Rao Sunkishala², Ramakanth Darbha², Sharath Macharla², Jeffrey D. Milbrandt³, Herbert W. Virgin¹, Robi D. Mitra³, Richard D. Head³, Shashikant Kulkarni¹, Andrew Bredemeyer¹, John D. Pfeifer¹, Karen Seibert¹

¹Department of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO, ²SemanticBits LLC, Herndon, VA, ³Department of Genetics, Washington University in St. Louis, St. Louis, MO. E-mail: rakesh@wustl.edu
Vanderbilt University – A Leader in NGS in Practice

The Diagnostic Management Team: How it Works and its Clinical and Financial Implications

Michael Laposata, MD, PhD
Edward and Nancy Fody Professor of Pathology
Vanderbilt University School of Medicine
Pathologist in Chief, Vanderbilt University Hospital

Progress to date:
- Implemented Cardiology Drugs and Genomic Phenotyping – clopidogrel, warfarin and statins
- Reporting Workflow Established
- Not Using a Commercial Solution
- U Pitt, Geisinger and Northwestern have an R18 implementation grant pending with Epic and Cerner help
- Real opportunity for Pathology Informatics

On Path Info website – Wednesday Plenary Session – Personalized Medicine
http://pathinfoarchives.dbmi.pitt.edu/apiii_archives/2012/Wednesday/Plenary/Laposata/Diagnostic_Management_Team/Diagnostic_Management_Team.html
Recommendation – Expansion Strategies

• Phase Three
  – Partner with other Specialties
    • Pharmacology, Biostatistics and Health Care Economists
  – Scale out pilot programs
    • Expansion in Cancer and Non-Cancer Programs
  – Leverage infrastructure to build out research funding
    • Molecular Pathology Informatics
    • Personalized Medicine for Cancer
    • Pathology Informatics Science for Honest Brokering of Phenotype Data and Genomic Data for Translational Research
Computational Pathology Fellowships

• Why this “new” fellowship is key!
  – Pathology Informatics is now established as a “division” or “subspecialty” in many practices
    • This is perceived as a service component to Pathology Practice – the Information Technology component
  – Academic Pathology Informatics is emerging
    • Struggles from the lack of defined research focus
    • PathoBioinformatics (coined by Friedman, 2007) is the domain of Computational Pathology
    • This is the true “Informatics” component of Path Info
  – These fellows will help build the research leaders!!!
Summary

• Pathology Informatics is critical to the development of a Cancer Personalized Medicine strategy and Enterprise Analytical Data Warehouse in partnership with Health System Health Information Technology (HIT)

• Aggressively expand research opportunity and recruit trainees in Computational Pathology, Pathology Informatics and Biomedical Informatics

• Influence leadership to grow the solutions in this space (begin with high quality fellows)!!!
End of Talk – e-mail me at becich@pitt.edu if you have questions/clarifications not covered in the discussion.

NOTE: E-mail me if you want PDFs of articles or presentation.

Thank you for attending Pathology Informatics 2012 (our 17th Annual meeting!!!)
1996-1999 Anatomic Pathology, Imaging & Internet
2000-2003 Anatomic and Clinical Pathology
2004-2007 Oncology & Bioinformatics
2008-11 Imaging Informatics – Radiology and Pathology
2012- Lab InfoTech Summit & APIII present Pathology Informatics 2012 Chicago, IL

http://pathinformatics.pitt.edu
Association for Pathology Informatics (API)

http://www.pathologyinformatics.org

“…to advance the field of pathology informatics as an academic and a clinical subspecialty of pathology…”

Slide 30
Please support JPI, API and Pathology Informatics as the Home for Digital Pathology - Great Academic and Strategic Partnership with Multiple Benefits!!!
Informatics Or Computer Science Jobs

126 of 19,526 jobs found

Search:  |  Advanced Search

| Keywords: informatics or computer science |
| Location: e.g. "Chicago, IL" or "60601" |
| Posted within: Last 30 Days |

Narrow Search

| Category: Information Technology (13455) Engineering (5973) Management (3887) Design (2475) Health Care (1874) |
| Company: ManTech International... (404) Randstad Technologies (372) Robert Half Technology (334) Kforce (245) TEKayastha, Inc (205) |

| City: Chicago (600) Atlanta (454) New York (399) Houston (384) Dallas (206) |
| State: California (2219) Texas (1625) Illinois (1382) Virginia (934) Florida (894) |

Show me: Closest jobs first | Relevant jobs first | Newest jobs first

Search Results

<table>
<thead>
<tr>
<th>Job Title / Description</th>
<th>Company</th>
<th>Location</th>
<th>Posted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomedical Informatics Software Engineer</td>
<td>University of Kansas Medical Center</td>
<td>KS - Kansas City</td>
<td>1 Week Ago</td>
</tr>
<tr>
<td>Clinical Informatics Nurse Specialist</td>
<td>Marvel Consultants, Inc.</td>
<td>MD - Maryland</td>
<td>2 Weeks Ago</td>
</tr>
</tbody>
</table>
CoSBI and “The Pipeline”

• Provide High School Students with an introduction to Computational/Systems Biology and Biomedical (Pathology) Informatics as a research career (CoSSBI Summer Academy)
• Prepare them for what college courses the would need to be competitive in computation and informatics via an immersion in our science.
• Give them a “real” research experience unlike other summers at Brown, Duke, Hopkins, Stanford, etc...
• Encourage them to undergraduate degree programs in bioinformatics (like ours at Pitt)
• Recruit them to Pathology (and Biomedical) Informatics
Pitt’s Bioinformatics Undergraduate Degree

http://www.cs.pitt.edu/undergrad/bioinformatics/

ABOUT THE DEGREE

Bioinformatics is the theory, application and development of computing tools to solve problems and create hypotheses in all areas of biological sciences. Biology in the post-genome world has been and continues to be transformed from a largely laboratory-based science to one that integrates experimental and information science. Bioinformatics has contributed to advances in biology by providing tools that handle datasets too large and/or complex for manual analysis. Examples of some of these tools include assembling the DNA sequence of entire genomes, gene finding algorithms, microarray expression analysis, molecular systems modeling, and biomarker discovery from mass spectra. Computational tools are central to the organization, analysis and harvesting of biological data at the level of macromolecules, cells and systems. Consequently, there is a growing need for trained professionals who understand the languages of biology and computer science. Biologists trained in more traditional programs may not have a working knowledge of statistics and algorithms, whereas computer scientists trained in more traditional programs may not have a working knowledge of the chemistry and biology required in the field.

The Undergraduate Bioinformatics Degree at the University of Pittsburgh, which is operated jointly by the Departments of Biological Sciences and Computer Science, program offers training that builds a solid foundation in chemistry, biology, computer science, mathematics and statistics. This training will enable students to communicate fluently with experts across these disciplines, and to have the skills necessary to apply computing tools to address contemporary problems in biology and medicine. The training will enhance the professional opportunities for undergraduates to pursue careers in pure or applied research in academia, government, pharmaceutical, medical, or biotechnology sectors.
U Pitt’s Bioinformatics Course Curriculum

- Curriculum Loaded with Biology and Computer Science
- Focused learning in:
  - Chemistry
  - Math
  - Statistics
- Includes a “real” research experience with “Capstone” mentored research project.
- 2 yrs electives (CoSB & BMI)
- Work your summers in a research laboratory with the Department of Biomedical (Pathology) Informatics
  - Now a job guarantee!!!
  - i-STEM (Informatics for Science Technology Engineering and Math)
  - 501c3 (not for profit) now proposed to fund summer research for high school and college undergrads

**BIOINFORMATIC SAMPLE SCHEDULE**

**SAMPLE SCHEDULE PLAN**
(see a Bioinformatics advisor for a customized schedule plan)

<table>
<thead>
<tr>
<th></th>
<th>Fall 1</th>
<th>Fall 2</th>
<th>Fall 3</th>
<th>Fall 4</th>
<th>Spring 1</th>
<th>Spring 2</th>
<th>Spring 3</th>
<th>Spring 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course</td>
<td>BIOSC 0150(3 credits)</td>
<td>CS 0445(3 credits)</td>
<td>BIOSC 1810(3 credits)</td>
<td>CAPSTONE(3 credits)</td>
<td>BIOSC 0160(3 credits)</td>
<td>BIOSC 0350(3 credits)</td>
<td>BIOSC 1903(2 credits)</td>
<td>2 GEN ED(6 credits)</td>
</tr>
<tr>
<td></td>
<td>CHEM 0110(4 credits)</td>
<td>CS 0441(3 credits)</td>
<td>CS 1501(3 credits)</td>
<td>CHEM 0120(4 credits)</td>
<td>CHEM 0120(4 credits)</td>
<td>CHEM 0320(3 credits)</td>
<td>BIOSC 1540(3 credits)</td>
<td>MAJ ELECT(3 credits)</td>
</tr>
<tr>
<td></td>
<td>CS 0007(3 credits)</td>
<td>CHEM 0310(3 credits)</td>
<td>MAJ ELECT(3 credits)</td>
<td>CS 0401(4 credits)</td>
<td>CS 0401(4 credits)</td>
<td>STAT 1000(4 credits)</td>
<td>MAJ ELECT(3 credits)</td>
<td>2 GEN ED(6 credits)</td>
</tr>
<tr>
<td></td>
<td>2 GEN ED(6 credits)</td>
<td>MATH 0220(4 credits)</td>
<td>2 GEN ELECT(3 credits)</td>
<td>1 GEN ED(3 credits)</td>
<td>1 GEN ED(3 credits)</td>
<td>2 GEN ED(6 credits)</td>
<td>SEIMINAR(1 credit)</td>
<td>SEMINAR(1 credit)</td>
</tr>
<tr>
<td></td>
<td>Total 16</td>
<td>Total 16</td>
<td>Total 16</td>
<td>Total 16</td>
<td>Total 16</td>
<td>Total 16</td>
<td>Total 16</td>
<td>Total 16</td>
</tr>
</tbody>
</table>