# Informatics, Data Science, and Artificial Intelligence : Building the Educational Pipeline for the Future of Radiation Oncology

THE UNIVERSITY OF TEXAS
MD Anderson ~~Cancer~~ Center
Making Cancer History®

Dave Fuller, MD, PhD
cdfuller@mdanderson.org
@cd_fuller

# 2016-19 Funders:

The Andrew Sabin Family Fellowship Program, through an endowment established by the Andrew Sabin Family Foundation

A direct gift from the Beach Family of Phoenix, AZ.

NIH Big Data to Knowledge (BD2K) Program of the National Cancer Institute (NCI) Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Award (1R01CA214825-01)

National Science Foundation, Division of Mathematical Sciences, Quantitative Approaches to Biomedical Big Data (QuBBD)/Big Data to Knowledge (BD2K) Program (NSF1557559;

National Cancer Institute Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science (1 R01 CA214825-01)

•National Institute of Dental and Craniofacial Research (NR56/R01 DE025248-01)

•National Cancer Institute Grant MD Anderson Head and Neck Specialized Programs of Research Excellence (SPORE) Development Award (P50CA097007-10)

•Elekta AB/MD Anderson MRI-LinAc Consortium Seed Grant*

•Elekta AB Travel support &Honoraria*

#-Philantropic individuals/agencies/societies

*-Corporate/industry funders

+-Federal or state funding agencies

# 2014 Statistical and Applied Mathematical Sciences Institute (SAMSI) Innovation Lab
## "Interdisciplinary Approaches to Biomedical Data Science Challenges" Team

**Liz Marai**
PhD
Computer Science
U. Illinois-Chicago

**David Vock**
PhD
Biostatistics
U. Minnesota

**Guadalupe Canahuate**
PhD
Computer Science
U. Iowa

**Dave Fuller**
MD, PhD
Radiation Oncology
UT MD Anderson

-2016 Joint NIH/NSF Division of Mathematical Sciences Initiative on Quantitative Approaches to Biomedical Big Data (QuBBD) Grant, *"Spatial-nonspatial Multidimensional Adaptive Radiotherapy Treatment"* (NSF 1557679)

- 2017-2020 Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Grant, *"SMART-ACT: Spatial Methodologic Approaches for Risk Assessment and Therapeutic Adaptation in Cancer Treatment"* (R01 CA214825-01)

National Science Foundation
WHERE DISCOVERIES BEGIN

NIH NATIONAL CANCER INSTITUTE

# Informatics, Data Science, and Artificial Intelligence

Figure 2. Traditional Research Model and Data-Driven Models in Artificial Intelligence

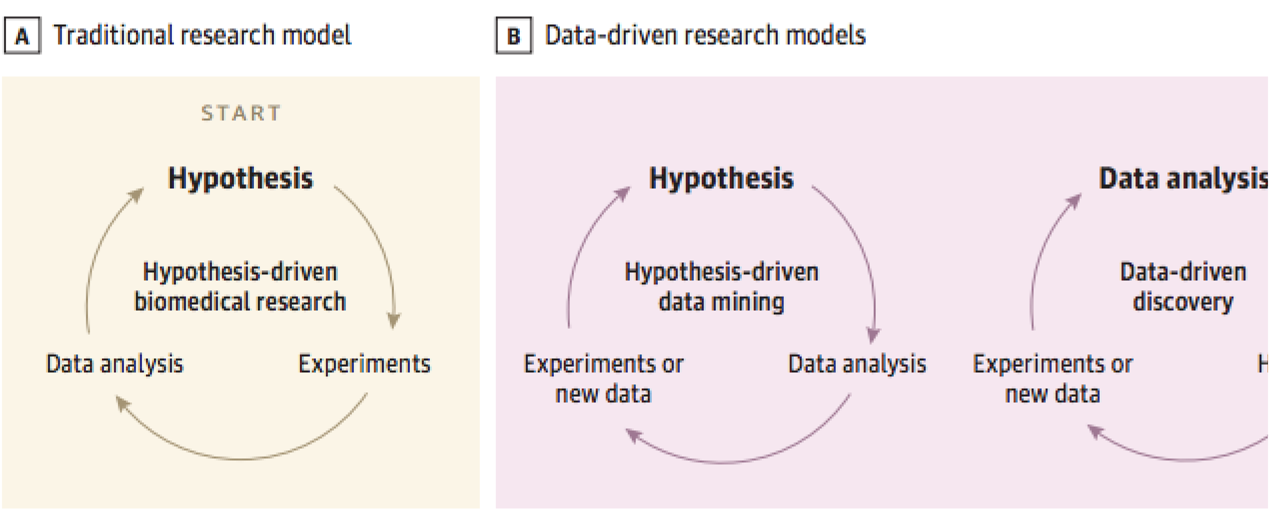
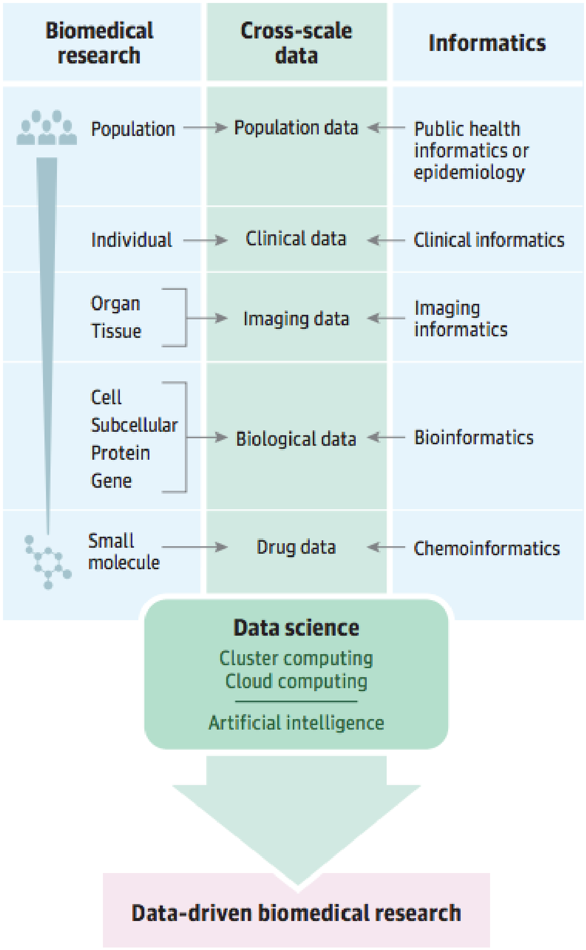





Figure 1. Biomedical Research and Informatics Approaches in Artificial Intelligence



MD Anderson

COMMENTARY

# The Future of Artificial Intelligence in Radiation Oncology

Reid F. Thompson, MD, PhD,*,† Gilmer Valdes, PhD,‡
Clifton David Fuller, MD,§ Colin M. Carpenter,‖ Olivier Morin, PhD,‡
Sanjay Aneja, MD,¶ William D. Lindsay,# Hugo J.W.L. Aerts, PhD,**
Barbara Agrimson, CMD,* Curtiland Deville, MD,††
Seth A. Rosenthal, MD,‡‡ James B. Yu, MD,¶ and
Charles R. Thomas, Jr, MD*

Int J Radiation Oncol Biol Phys, Vol. 102, No. 2, pp. 247–248, 2018

As a specialty, radiation oncology has a long track record of deploying novel technology for the benefit of our patients. AI will be a part of our future. Even though this may seem to be on the distant horizon, the opportunity cost of ignoring AI at this juncture is steep.

# Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation?

Reid F. Thompson [a,b,*], Gilmer Valdes [c], Clifton D. Fuller [d], Colin M. Carpenter [e], Olivier Morin [c], Sanjay Aneja [f], William D. Lindsay [g], Hugo J.W.L. Aerts [h,i], Barbara Agrimson [a], Curtiland Deville Jr. [j], Seth A. Rosenthal [k], James B. Yu [f], Charles R. Thomas Jr [a]

## From vision to reality: How do we get from here to there?

Recognition of the magnitude of the task, the need for collaboration with other disciplines and specialties, and the potential revolutionary, paradigm-changing nature of the changes to come in the coming years should all serve as a "call to arms" for Radiation Oncology to actively engage in harnessing the potential advantages of AI for our specialty and for our patients. The question remains "How?"

# A few definitions…

**Informatics** [ The Discipline]

- "Informatics is the science of processing data for storage and retrieval; information science as a field."

**Artificial intelligence** [The Research domain]

- "The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."

**"Big Data"** [The Input]

- "*Big Data* is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."- Gartner

**Machine learning ("statistical learning")** [The Methodology]

- "Machine learning explores the study and construction of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs."

**"The informatics lecturer at the AI conference gave a talk on what machine learning approaches to use for Big Data."**

# What is informatics as a discipline?

**“ Medical informatics is the rapidly developing scientific field that deals with resources, devices and formalized methods for optimizing the storage, retrieval and management of biomedical information for problem solving and decision making”**

**Edward Shortliffe, MD, PhD**

**1995**

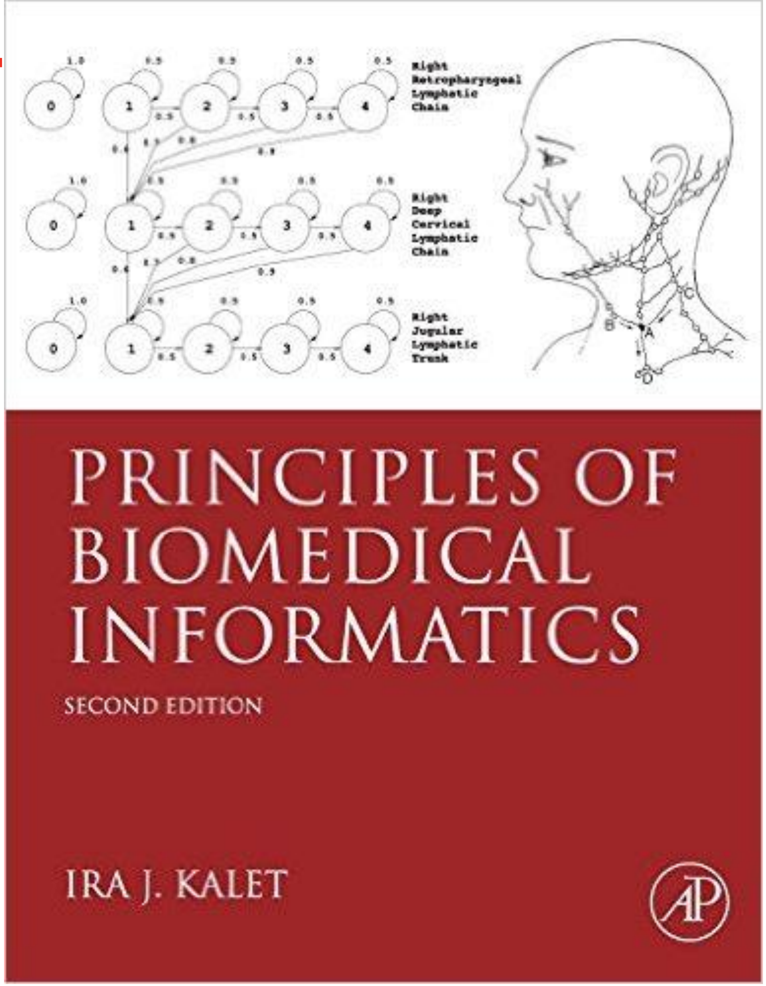
# What is biomedical informatics?

Elmer V. Bernstam [a,b,*], Jack W. Smith [a], Todd R. Johnson [a]

Journal of Biomedical Informatics 43 (2010) 104–110

...**informatics is** the science of information, where information is defined as **data + meaning**."

| | Ackoff's hierarchy | Philosophy of information |
|---|---|---|
| **Data** | Symbols | Lack of uniformity |
| **Information** | Data that have been processed to be useful (answers to: who, what, when or where questions) | Data + Meaning |
| **Knowledge** | Application of data and information to answer "how" questions | Justified, true belief |
| **Wisdom** | Understanding and appreciation of "why" questions | |




Ira Kalet, PhD

## Remembering Ira Kalet, 1944-2015

Retired CSE adjunct professor Ira Kalet passed away last night after a long battle with cancer.

Ira joined the University of Washington in 1978 in the then newly formed Department of Radiation Oncology. Subsequently he held adjunct appointments in Computer Science & Engineering, Bioengineering, and Biological Structure, and a joint appointment in Medical Education (now the Department of Biomedical Informatics and Medical Education).

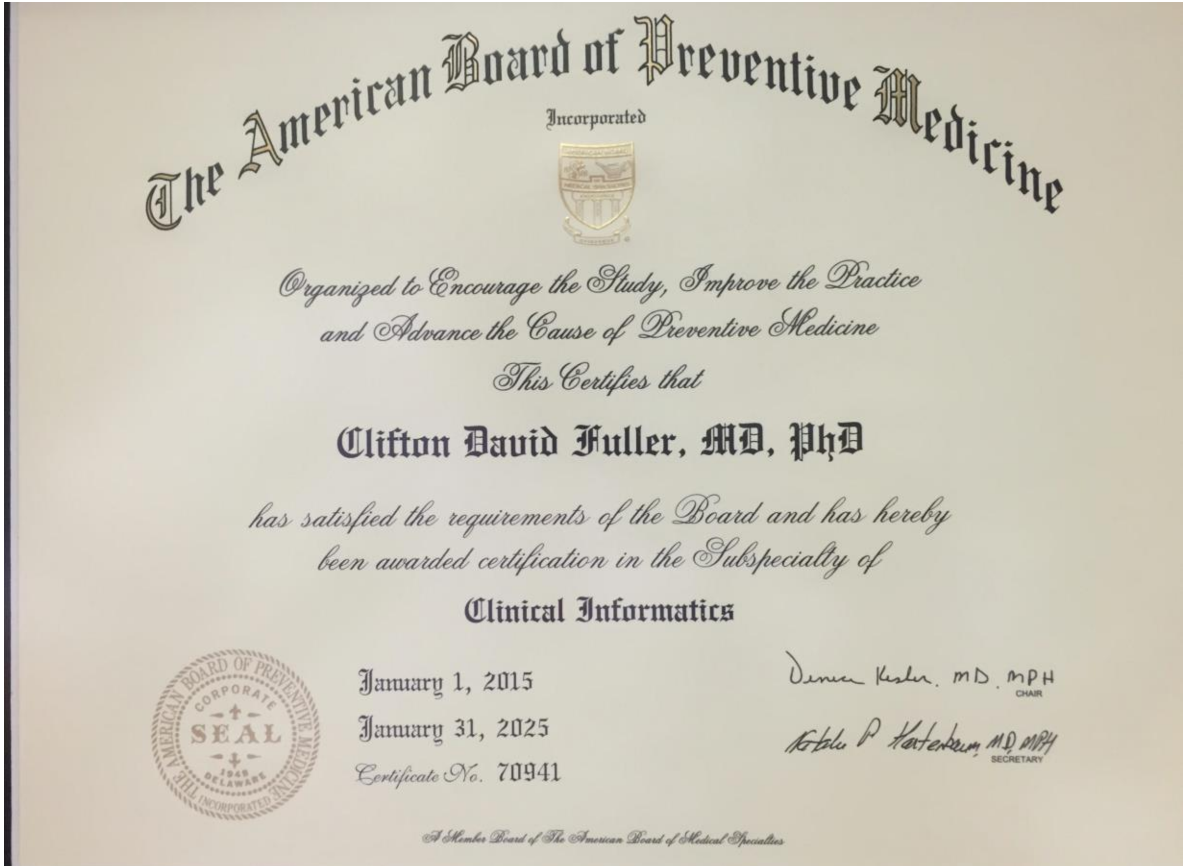
# Clinical Informatics Becomes a Board-certified Medical Subspecialty Following ABMS Vote

**Thursday, September 22, 2011**

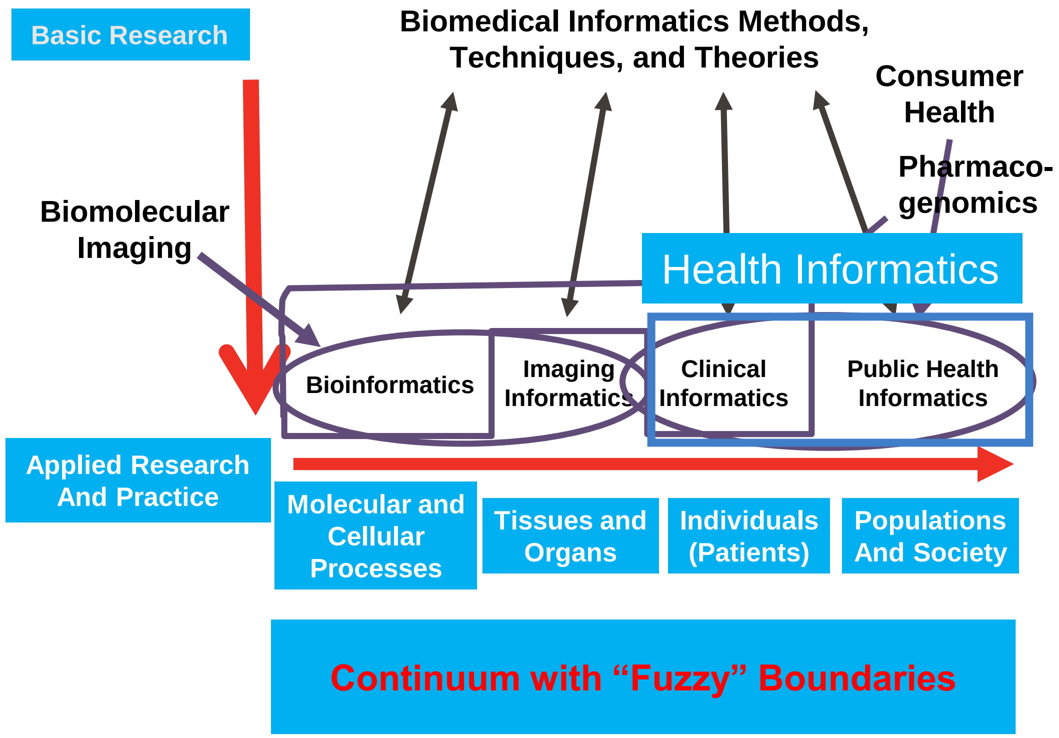
***AMIA to offer prep courses for clinicians who sit for Board Exam***

Washington, DC—Today, AMIA—the association for informatics professionals—announces the success of a multi-year initiative to elevate clinical informatics to an American Board of Medical Specialties (ABMS) subspecialty certified by an examination administered by the American Board of Preventive Medicine and available to physicians who have primary specialty certification through the American Board of Medical Specialties. Joining such subspecialties as pediatric anesthesiology, medical toxicology, sports medicine, geriatrics medicine, and cardiovascular disease, clinical informatics (CI) certification will be based on a rigorous set of core competencies, heavily influenced by publications on the subject

# ABMS Informatics Sub-Specialty- Very Cool!!



# Biomedical Informatics in Perspective



MD Anderso

SPECIAL ARTICLE

# Technology for Innovation in Radiation Oncology

Indrin J. Chetty, PhD,* Mary K. Martel, PhD,† David A. Jaffray, PhD,‡

1. Integrating radiation oncology databases across the [illegible] care (45). The creation of a Virtual Clinical Trials Group that [illegible] for conducting cooperative research is under consideration. Sharing practices and outcomes will permit high [illegible] and [illegible] quality (46).

2. Tools need to be created and made available for patients and physicians to discuss treatment options, as recom- mended [illegible] Institution. Such an approach will drive the development of [illegible] in which one pre- scribes [illegible] 95% local control rate at 3 years with 5% grade 3 or more [illegible] beyond radiation oncology.

3. Expertise in the informatics domain among radiation oncology professionals needs to be developed (6). The most suitable candidates with the appropriate skill sets and multidisciplinary knowledge to succeed in this space are likely medical physicists or physicians with strong computational backgrounds. Training grants for devel- oping programs for oncology informatics will provide these individuals with the knowledge needed to support informatics research initiatives.

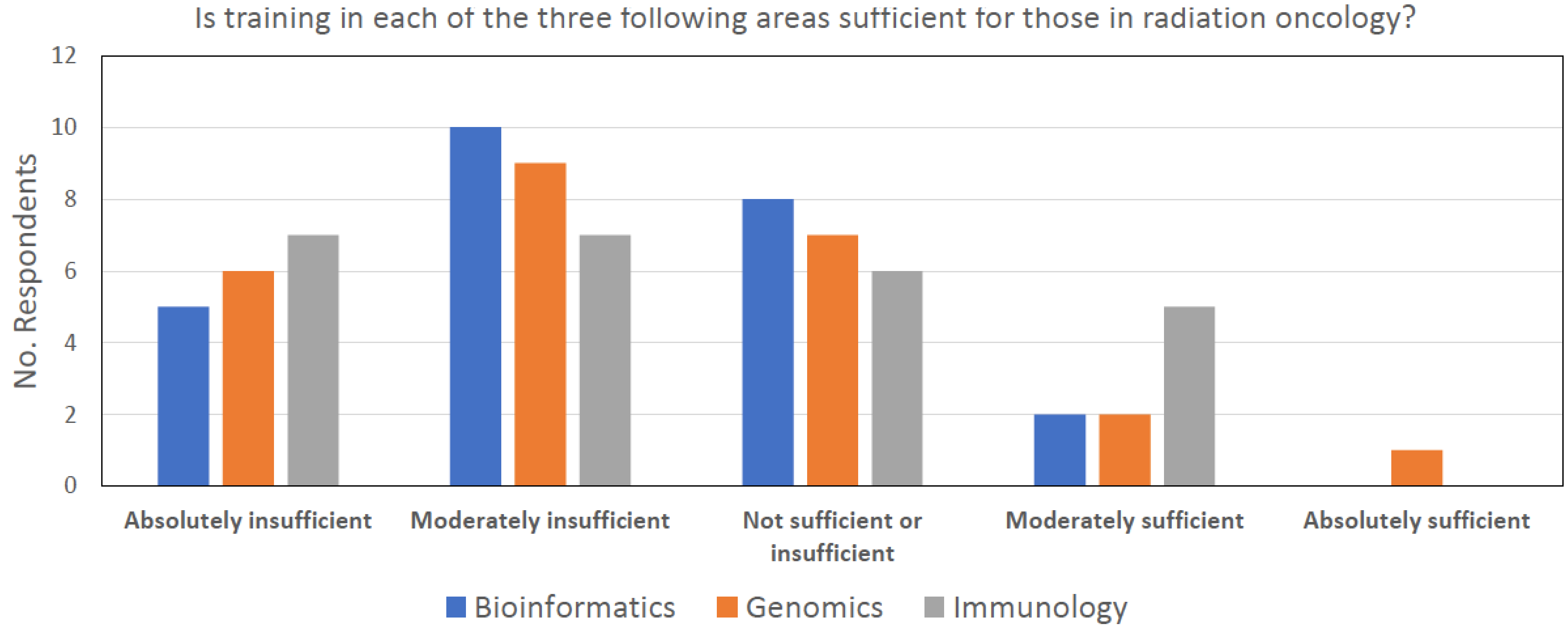
3. Expertise in the informatics domain among radiation oncology professionals needs to be developed (6). The most suitable candidates with the appropriate skill sets and multidisciplinary knowledge to succeed in this space are likely medical physicists or physicians with strong

computational backgrounds. Training grants for devel- [illegible] these individuals with the knowledge needed to support [illegible] monitoring of the quality of oncology care at the point(s) of delivery (45). [illegible] are emerging in other domains and will also benefit the field of [illegible] that 5% differences in dose result in significant changes in [illegible] bilities will be reinforced or challenged through col- [illegible]

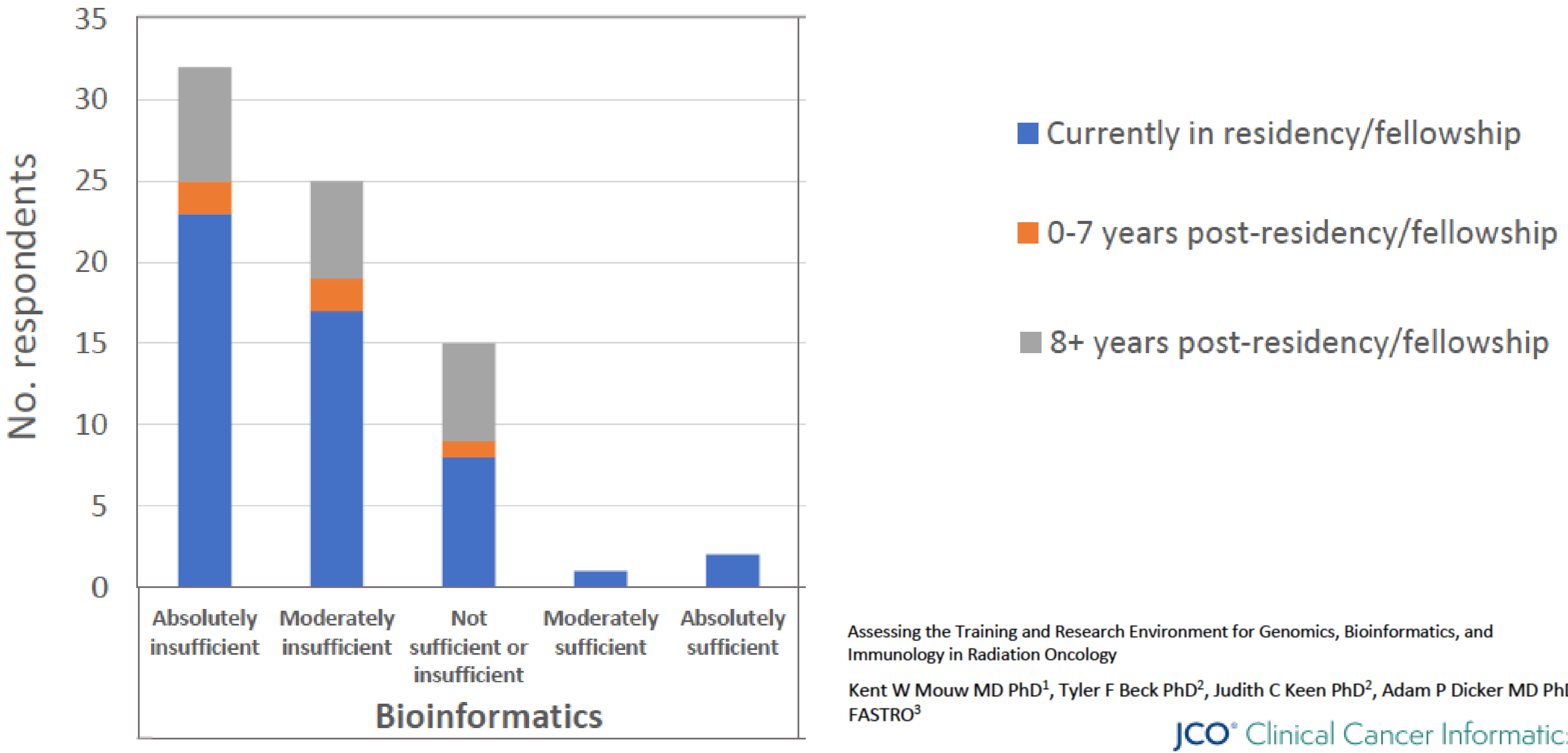
# Assessing the Training and Research Environment for Genomics, Bioinformatics, and Immunology in Radiation Oncology

A.

Kent W Mouw MD PhD[1], Tyler F Beck PhD[2], Judith C Keen PhD[2], Adam P Dicker MD PhD, FASTRO[3]



# Do you think that training in the following areas would allow you to advance your research career?

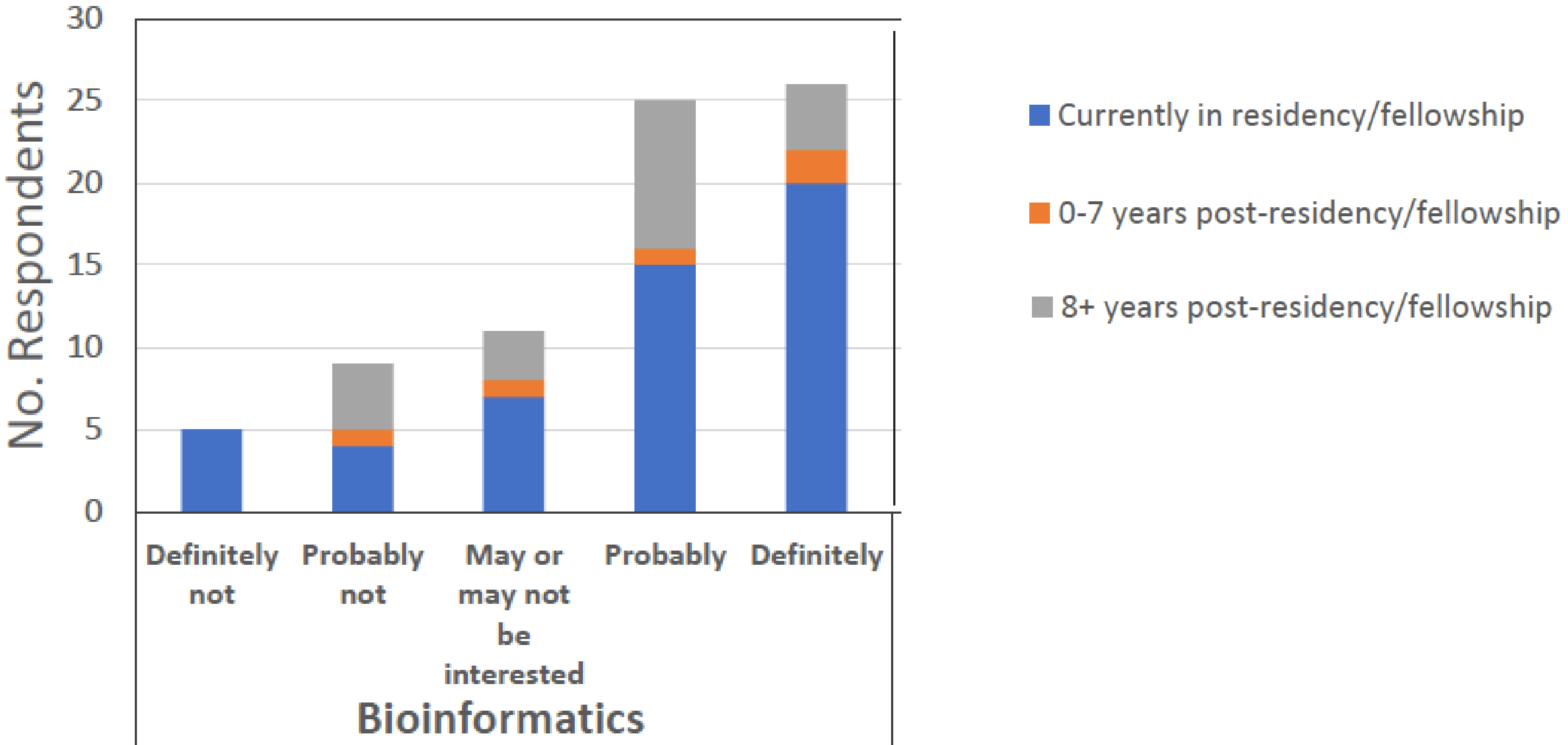



Assessing the Training and Research Environment for Genomics, Bioinformatics, and Immunology in Radiation Oncology

Kent W Mouw MD PhD[1], Tyler F Beck PhD[2], Judith C Keen PhD[2], Adam P Dicker MD PhD, FASTRO[3]

JCO Clinical Cancer Informatics

## Would you be interested in taking formal training courses in the following subject areas?



# Example: Many #RadOnc residents have limited "Big Data" exposure/background

- We informally polled trainees during the MD Anderson Annual RadBio/Physics course during their informatics lecture
  - In 2018, 0/21 participants had heard of "FAIR Guiding Principles"
  - 4/21 could correctly map concepts of a "standard", "ontology", or "syntax"
  - 4/21 had heard of SNO-MED
  - 3/21 could correctly identify HL7 as a standards framework
  - 20/21 could identifty DICOM, but most considered a filetype, not as an informatics satndard

# The Future is FAIR

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

## Box 2 | The FAIR Guiding Principles

**To be Findable:**
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

# Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

## 3. FAIR FOR MACHINES AS WELL A PEOPLE

In eScience, two clearly separated substrates for knowledge discovery can be distinguished.

1. The actual data, which is as a rule beyond human intellectual capacity to analyse and
2. The 'Explicitome' (everything we already made explicit in text, databases and any other format to date).

- Data should be **Findable**
- Data should be **Accessible**
- Data should be **Interoperable**
- Data should be **Re-usable.**

SCIENTIFIC DATA

SCIENTIFIC DATA

OPEN

# Data Descriptor: Dynamic contrast-enhanced magnetic resonance imaging for head and neck cancers

Joint Head and Neck Radiotherapy-MRI Development Cooperative*



SCIENTIFIC DATA | 5:180008 | DOI: 10.1038/sdata.2018.8

| Data record | Description |
|---|---|
| Clinical meta-data for DCE-MRI data set.csv | Clinical meta-data for the 15 oropharyngeal cancer patients, representing our cohort. Patients, disease and treatment identifiers are detailed in this *.csv file. |
| ReadMe_Clinical meta-data.csv | Supplemental information about the headings of the columns in the clinical data file. |
| Arterial Input Function.csv | Arterial input function (AIF) extracted from DCE-MRI and required for pharmacokinetic modelling of tumors are provided. |

**Table 4. Description of data records uploaded to the figshare repository of the HPV and local recurrence prediction challenges.**

# NCI ANNOUNCES ONCOLOGY DATA SCIENCE FELLOWSHIP

3/13/2017

In response to the need to train oncology professionals capable of harnessing the power of big data and emerging health information technologies, the National Cancer Institute (NCI) has developed an innovative oncology fellowship program for qualified radiation oncologists.

The one-year fellowship will provide the fellow with the opportunity to develop and apply cutting-edge oncology data science knowledge to mentored research opportunities at both NCI and the Food and Drug Administration (FDA) Information Exchange and Data Transformation (INFORMED) program. Fellows will also have the opportunity to deliver clinical care at the NCI.

Mentors:
Kevin Camphausen, MD (NCI) [Fellowship Director]
Sean Khozin, MD, MPH (FDA INFORMED)
Anand Shah, MD, MPH (FDA INFORMED)

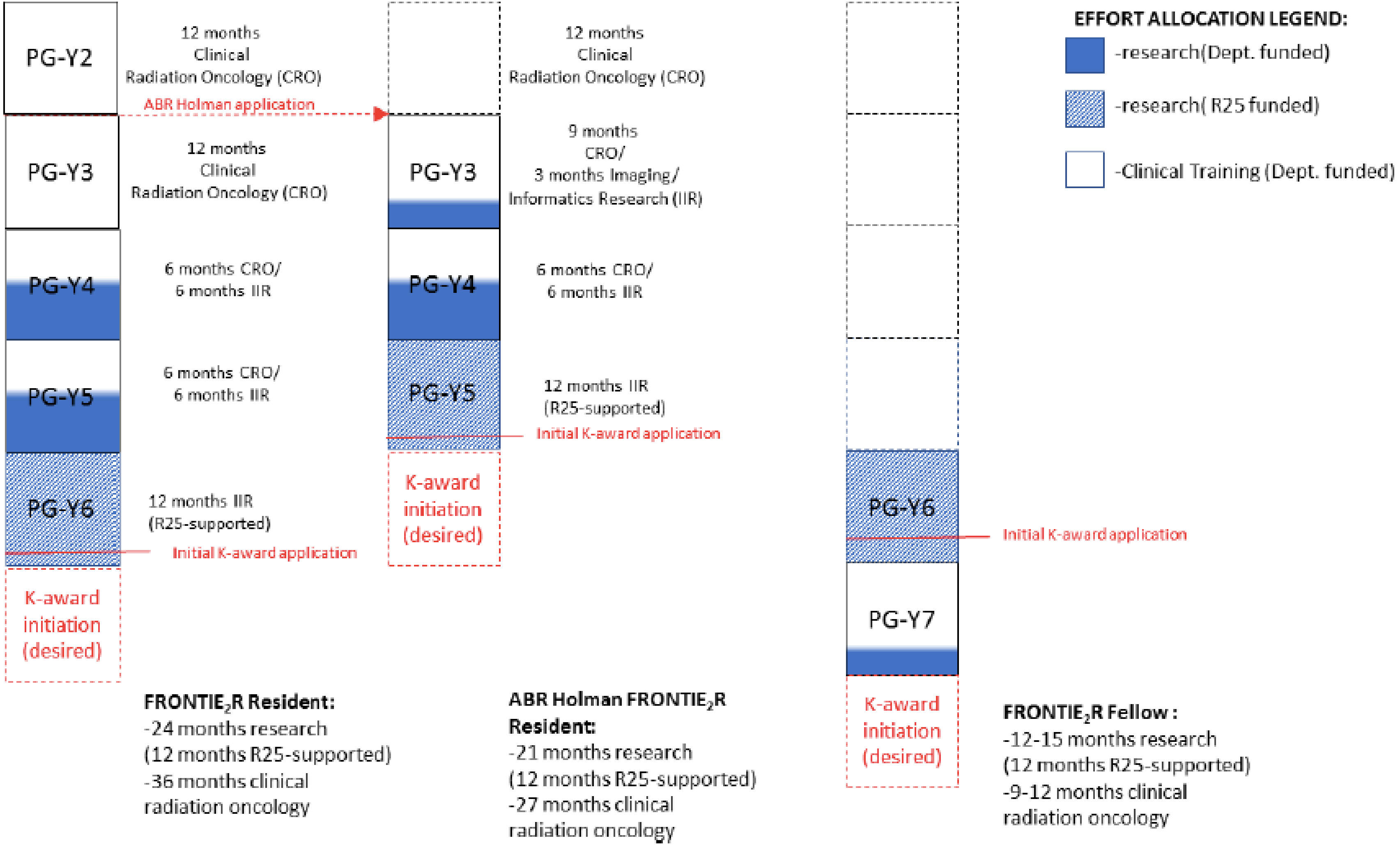
Who should apply:
• Radiation oncologists looking to develop in-depth expertise in data science, health technology, entrepreneurship and big data analytics

# Fellow and Resident Radiation Oncology iNtensive Training in Imaging and Informatics to Empower Research Careers ($FRONTI_2ER$)

- 2 Residents/Fellows admitted annually NIBIB.
- Trainees receive individualized training program.
- Trainees receive Applicants up to $10,000 for research-related expenses,
- All Trainees must commit to submit a K-level application.

**Prajnan Das, M.D., M.S., M.P.H. Professor, & Residency Director co-PI**

**Dave Fuller, MD, PhD Ass. Professor, co-PI**


PG-Y2
12 months
Clinical
Radiation Oncology (CRO)
ABR Holman application
PG-Y3
12 months
Clinical
Radiation Oncology (CRO)
PG-Y4
6 months CRO/
6 months IIR
PG-Y5
6 months CRO/
6 months IIR
PG-Y6
12 months IIR
(R25-supported)
Initial K-award application
K-award initiation (desired)
FRONTIE$_2$R Resident:
-24 months research
(12 months R25-supported)
-36 months clinical radiation oncology
12 months
Clinical
Radiation Oncology (CRO)
PG-Y3
9 months
CRO/
3 months Imaging/
Informatics Research (IIR)
PG-Y4
6 months CRO/
6 months IIR
PG-Y5
12 months IIR
(R25-supported)
Initial K-award application
K-award initiation (desired)
ABR Holman FRONTIE$_2$R Resident:
-21 months research
(12 months R25-supported)
-27 months clinical radiation oncology
PG-Y6
Initial K-award application
PG-Y7
K-award initiation (desired)
FRONTIE$_2$R Fellow :
-12-15 months research
(12 months R25-supported)
-9-12 months clinical radiation oncology
EFFORT ALLOCATION LEGEND:
-research(Dept. funded)
-research( R25 funded)
-Clinical Training (Dept. funded)

# Awarded FY2019...so far, in Year 1:

- 2 trainees awarded cycle 1 (both are Clinical Fellows).
- One trainee has imaging focus, the other informatics.
- Resident applicants will begin matriculating year 2.
- Expansion of the program with potential for increased non-NIH-funded trainees pending.

Karine
Al Feghali
MD

Baher
Elgohari
MD

# Longitudinal Engagement of Pathology Residents

## A Proposed Approach for Informatics Training

*Luigi K. F. Rao, MD, and John R. Gilbertson, MD*

**Table 1**
**Longitudinal Engagement: FIRE Model for Resident Informatics Training**

| Foundation | Immersion | Refinement | Expertise |
|---|---|---|---|
| Didactics and online-based learning | Administrative group involvement | Real-time case studies, directed readings, and active membership | Experiential opportunities with authentic management skills application |
| Skills and project alignment | Exposure to interaction with vendors and professional societies | Continued project stewardship | Project summation (acquisition of systems, implementation and deployment) |
| | | Regional/national meeting exposure | Standards committees (national/international/special interest groups) participation |
| Bioimage informatics example: lectures covering fundamentals of histology, optics, microscopy, digital imaging, and image analysis | Data quality example: hospital clinical decision support and computerized provider order entry group activities | Bioimage informatics example: Digital Imaging and Communications in Medicine Working Group membership | |

FIRE, foundation, immersion, refinement, expertise.

MD An

editorial

# Career Enrichment Opportunities at the Scientific Frontier in Radiation Oncology

Reid F. Thompson, MD, PhD[1,2]; Clifton D. Fuller, MD, PhD[2,3]; Abigail T. Berman, MD[4]; Sanjay Aneja, MD[5]; and Charles R. Thomas Jr, MD[2]

**TABLE 1.** FIRE (foundation, immersion, refinement, and expertise through experiential learning) Longitudinal Engagement Concept

| | Foundation | Immersion | Refinement | Expertise |
|---|---|---|---|---|
| Genomics/bioinformatics example | Core ACGME curriculum; departmental didactic offerings; and scalable online curricula such as MOOCs | Observational shadowing exposure to laboratory or bioinformatics at home facility | Hands-on experiences with core facilities and researchers at home institute | |
| | | | Dedicated research project development and publication (eg, via Holman Pathway) | |
| Clinical informatics example | Core ACGME curriculum; departmental didactic offerings; and scalable online curricula such as MOOCs | Engagement with departmental or institutional electronic medical record, picture archiving and communication system, information technology, or other informatics specialists | Development of a quality assessment/quality improvement project with informatics components | |
| | | | Medical physics or imaging analysis projects that intersect with informatics processes | |
| | | | Electronic health record–based health services research | |

JCO® Clinical Cancer Informatics

# What about post-Rad Onc Training?

## Fellowship in Clinical Informatics: Radiation Oncology Track

Clinical informatics is the subspecialty of all medical specialties that transforms health care by analyzing, designing, implementing, and evaluating information and communication systems to improve patient care, enhance access to care, advance individual and population health outcomes, and strengthen the clinician-patient relationship.

Charles R. Thomas, Jr., MD, Chair,

OHSU

School of Medicine

Radiation Medicine

# Unmet educational needs

- Limited time for devoted machine learning/informatics training and lack of distributed expertise will necessitate online resources (MOOCS, modular courses) as a ROECSG-supported effort
- Potential integration into current rad onc statistics education my be aided by modular exercises Greater support for specific "computational physician-scientist phenotype" by NCI/NLM/NIBIB through targeted training awards
- Development of a core leaership cadre of "first wave" junior faculty physicians through avenues such as U. Michigan Practical Big Data Workshop, ACR Data Science Institute, NCI initiatives focused on physician-scientists.

# Moving forward

- How do we develop the pipeline of **computational scientists who are also radiation oncologists?**
  - Traditional MSTP programs train basic/translational "wet lab"scientists
  - Most Big Data expertise is *ad hoc*
- **Need earlier integration into training for maximum efficacy**
  *(even if at a limited number of training programs)*

# Revitalizing the Radiology Research Enterprise[1]

The Radiological Society of North America has embarked on a program to assist academic radiology departments in becoming more successful in their research mission. As part of this program, the Society is offering a course for chairpersons and other members of departments involved in stimulating and managing research efforts. Acquiring the understanding and the skills necessary to develop, administer, and nurture research programs of various maturity will be the focus of the course.

## Combined Radiology Residency/ PhD Program for Education of Academic Radiologists:

A Response to Revitalizing the Radiology Research Enterprise[1]

Radiology 2007; 245:14–20

- Can we develop Holman-pathway compatible AI-focused educational efforts leading to MS/PhD as part of residency for select trainees?
- Could this be accomplished as a multi-site program ?
- Can this be done through existing structures (e.g. Moffitt PSOC, CTSA programs, NLM training programs) or is NCI support vital?

**Figure 1**

**Key Aspects of the Residency-PhD Program**

1. Efficient curriculum combines radiology residency and PhD training, condensing a typical 1-year internship, 4-year residency, and 3-5-year PhD into 6 postgraduate years
2. Medical school coursework counts for one year of credit toward graduate degree
3. B. Leonard Holman Pathway enables reduction of clinical radiology to 27 months
4. Curriculum provides integrated clinical and research training early in professional career
5. Integrated training prevents reorientation necessitated by transitions in MD-PhD programs
6. Papers submitted for publication in peer-reviewed journals assemble to form dissertation
7. Publications and grant applications provide head start on academic career development

**Figure 1:** Key aspects of combined residency/ PhD program.

# Next steps

- Development of a Rad Onc –focused "core curriculum" in informatics and machine learning (John Kang/Erin Gillespie) with possible ROECSG support.
- Develop curriculum into a MOOC/modular course
- Encourage AAPM TG particiaption by junior faculty recruitment to Big Data subcommittee (Chuck Mayo)
- Identify additional stakeholders (ASTRO, NCI, AAPM, RSNA) to define coherent vision/pathway for training needs across programs.

- IDEAS/SUGGESTIONS APPRECIATED!!

# The future for enhanced performance looks good for advances in informatics, AI, and Big Data in #RadOnc, if we can build the necessary educational tools!


Please email/visit soon!

cdfuller@mdanderson.org

Caroline Chung, MD
Rad Onc MR Program Lead.