



---

## INTERPRETABLE DEEP LEARNING FOR CANCER RESEARCH

---

SEPTEMBER 20-21, 2018

NATIONAL INSTITUTES OF HEALTH  
BUILDING 31 C-WING 6<sup>TH</sup> FLOOR, ROOM 6  
31 CENTER DRIVE  
BETHESDA, MARYLAND 20892

---

### INTRODUCTION OF THE WORKSHOP

---

Recent developments in convolutional neural networks and related machine learning techniques commonly referred to as deep learning (DL) have made breakthroughs in computer vision, robotic control, machine translation, voice recognition, gaming, and many other areas. While DL has seen rapid adoption by various cancer imaging communities, its use by the broader cancer biology community has been less marked. This can be attributed, at least in part, to a limitation frequently referred to as the “black box”, where DL models can make correct predictions but without association to underlying mechanisms for biological interpretation. This workshop will bring together DL researchers, cancer systems biologists, and computational biologists to discuss challenges and opportunities to develop interpretable DL methods that can be applied to cancer biology investigations for knowledge generation. The workshop aims to enumerate significant hurdles and identify potential pathways to promote innovations in DL method development, with the goal of overcoming existing limitations and making the DL approach most impactful in cancer research.

### DISCUSSION TOPICS OF WORKSHOP SESSIONS:

---

**Session 1. Transferability of *tabula rasa* learning methods to biomedical problems.** Most deep learning models developed in the biomedical domain rely on benchmarked datasets for supervised training. Entirely rule-based learning without human knowledge or data has demonstrated success in competing with traditional supervised learning models in certain areas such as gaming. This session will focus on discussing the transferability of deep learning algorithms that learn *tabula rasa* to solve biological problems.

**Session 2. Transferability of biomedical domain knowledge to deep learning methods development.** The original design of deep learning neural networks was inspired by the structure of the human brain. With the ever-increasing sophistication of deep learning network development, biological knowledge can contribute to the design of new topology of neural networks, new training methods, and new ways of integration with other biologically informed machine learning algorithms. This session focuses on discussing avenues to integrate biological knowledge in designing new deep learning methods.

**Session 3. Challenges and solutions associated with noisy, heterogeneous, and limited amount of data.** Biological data are intrinsically noisy, heterogeneous, and are often distributed and insulated with limited access. This session will focus on discussing strategies to improve model quality with limited amounts of training data and/or data quality considerations.

**Session 4. Interpretability of deep learning models.** This session will focus on discussing strategies to open up the “black box” and make output interpretable for mechanistic insight and knowledge generation.

**Session 5. Model validation, sharing, and comparison.** This session will focus on discussing strategies and methods for model validation, model sharing, and model comparison so that we can learn from each other and improve as a field. Key resource needs will also be discussed at this session.

## ORGANIZING COMMITTEE:

---

Trey Ideker, Ph.D.	University of California, San Diego
Claire Tomlin, Ph.D.	University of California, Berkeley
Jennifer Couch, Ph.D.	National Cancer Institute
Jerry Li, M.D., Ph.D.	National Cancer Institute
David Miller, Ph.D.	National Cancer Institute

## AGENDA: THURSDAY, SEPTEMBER 20, 2018

---

8:15 am Shuttle departs hotel for NIH

---

### 9:00 am – 9:15 am WELCOME AND INTRODUCTION OF THE WORKSHOP

---

Dinah Singer, Ph.D.  
Acting Deputy Director, National Cancer Institute

Trey Ideker, Ph.D.  
Professor of Medicine, University of California, San Diego

Claire Tomlin, Ph.D.  
Professor of Electrical Engineering and Computer Sciences, University of California, Berkeley

---

### 9:15 am – 10:45 am OPENING KEYNOTE PRESENTATIONS

---

9:15 am – 9:55 am Overview of new deep learning methods  
Rick Stevens, Ph.D.  
Associate Laboratory Director for Computing, Environment and Life Sciences,  
Argonne National Laboratory

9:55 am – 10:30 am Interpretable deep learning for biological insights  
Trey Ideker, Ph.D.  
Professor of Medicine, University of California, San Diego

10:30 am – 10:45 am NCI Director's remarks  
Norman Sharpless, M.D.  
Director, National Cancer Institute

---

### 10:45 am – 11:00 am BREAK

---

<b>11:00 am – 12:40 pm</b>	<b>SESSION 1: TRANSFERABILITY OF <i>TABULA RASA</i> LEARNING METHODS TO BIOMEDICAL PROBLEMS</b>								
	<p>Panel members:</p> <table> <tr> <td>Rick Stevens, Ph.D. (Chair)</td> <td>Argonne National Laboratory</td> </tr> <tr> <td>Ziv Bar-Joseph, Ph.D.</td> <td>Carnegie Mellon University</td> </tr> <tr> <td>Anna Goldenberg, Ph.D.</td> <td>University of Toronto</td> </tr> <tr> <td>Shun Miao, Ph.D.</td> <td>NVidia</td> </tr> </table>	Rick Stevens, Ph.D. (Chair)	Argonne National Laboratory	Ziv Bar-Joseph, Ph.D.	Carnegie Mellon University	Anna Goldenberg, Ph.D.	University of Toronto	Shun Miao, Ph.D.	NVidia
Rick Stevens, Ph.D. (Chair)	Argonne National Laboratory								
Ziv Bar-Joseph, Ph.D.	Carnegie Mellon University								
Anna Goldenberg, Ph.D.	University of Toronto								
Shun Miao, Ph.D.	NVidia								
11:00 am – 11:50 am	Short presentations from panel members								
11:50 am – 12:40 pm	Panel discussion								
<b>12:40 pm – 1:40 pm</b>	<b>LUNCH (ON OWN)</b>								
<b>1:40 pm – 3:30 pm</b>	<b>SESSION 2: TRANSFERABILITY OF BIOMEDICAL DOMAIN KNOWLEDGE TO DEEP LEARNING METHODS DEVELOPMENT</b>								
	<p>Panel members:</p> <table> <tr> <td>Chris Sander, Ph.D. (Chair)</td> <td>Dana-Farber Cancer Institute</td> </tr> <tr> <td>Anthony Gitter, Ph.D.</td> <td>University of Wisconsin Madison</td> </tr> <tr> <td>Jinbo Xu, Ph.D.</td> <td>Toyota Technological Institute at Chicago</td> </tr> <tr> <td>Alexander Anderson, Ph.D.</td> <td>Moffitt Cancer Center</td> </tr> </table>	Chris Sander, Ph.D. (Chair)	Dana-Farber Cancer Institute	Anthony Gitter, Ph.D.	University of Wisconsin Madison	Jinbo Xu, Ph.D.	Toyota Technological Institute at Chicago	Alexander Anderson, Ph.D.	Moffitt Cancer Center
Chris Sander, Ph.D. (Chair)	Dana-Farber Cancer Institute								
Anthony Gitter, Ph.D.	University of Wisconsin Madison								
Jinbo Xu, Ph.D.	Toyota Technological Institute at Chicago								
Alexander Anderson, Ph.D.	Moffitt Cancer Center								
1:40 pm – 2:40 pm	Short presentations from panel members								
2:40 pm – 3:30 pm	Panel discussion								
<b>3:30 pm – 3:45 pm</b>	<b>BREAK</b>								
<b>3:45 pm – 5:35 pm</b>	<b>SESSION 3: CHALLENGES AND SOLUTIONS ASSOCIATED WITH NOISY, HETEROGENEOUS, AND LIMITED AMOUNT OF DATA</b>								
	<p>Panel members:</p> <table> <tr> <td>Claire Tomlin, Ph.D. (Chair)</td> <td>University of California, Berkeley</td> </tr> <tr> <td>Anshul Kundaje, Ph.D.</td> <td>Stanford University</td> </tr> <tr> <td>Young Hwan Chang, Ph.D.</td> <td>Oregon Health &amp; Science University</td> </tr> <tr> <td>Kristin Swanson, Ph.D.</td> <td>Mayo Clinic</td> </tr> </table>	Claire Tomlin, Ph.D. (Chair)	University of California, Berkeley	Anshul Kundaje, Ph.D.	Stanford University	Young Hwan Chang, Ph.D.	Oregon Health & Science University	Kristin Swanson, Ph.D.	Mayo Clinic
Claire Tomlin, Ph.D. (Chair)	University of California, Berkeley								
Anshul Kundaje, Ph.D.	Stanford University								
Young Hwan Chang, Ph.D.	Oregon Health & Science University								
Kristin Swanson, Ph.D.	Mayo Clinic								
3:45 pm – 4:45 pm	Short presentations from panel members								
4:45 pm – 5:35 pm	Panel discussion								
<b>5:35 pm</b>	<b>ADJOURN DAY 1</b>								
5:45 pm	Shuttle from NIH back to hotel								
7:00 pm – 9:00 pm	Group dinner at Chef Tony's restaurant 4926 St Elmo Avenue, Bethesda, Md								

---

## AGENDA: FRIDAY SEPTEMBER 21, 2018

---

8:15 am Shuttle departs hotel for NIH

---

### 9:00 am – 10:50 am **SESSION 4: INTERPRETABILITY OF DEEP LEARNING MODELS**

---

Panel members:

Regina Barzilay, Ph.D. (Chair)	Massachusetts Institute of Technology
Elias Bareinboim, Ph.D.	Purdue University
Xinghua Lu, M.D., Ph.D.	University of Pittsburgh
Trey Ideker, Ph.D.	University of California, San Diego

9:00 am – 10:00 am Short presentations from panel members

10:00 am – 10:50 am Panel discussion

---

### 10:50 am – 11:05 am **BREAK**

---

---

### 11:05 am – 12:50 pm **SESSION 5: MODEL VALIDATION, SHARING, AND COMPARISON**

---

Panel members:

Peter Sorger, Ph.D. (Chair)	Harvard Medical School
Emek Demir, Ph.D.	Oregon Health & Science University
Doron Levy, Ph.D.	University of Maryland College Park
Ronald Summers, M.D., Ph.D.	NIH Clinical Center

11:05 am – 12:05 pm Short presentations from panel members

12:05 pm – 12:50 pm Panel discussion

---

### 12:50 pm – 1:50 pm **LUNCH (ON OWN)**

---

---

### 1:50 pm – 3:00 pm **SESSION 6: SUMMARY OF PANEL DISCUSSIONS AND OUTLOOK**

---

Chairs: Trey Ideker and Claire Tomlin

1:50 pm – 2:40 pm Session summaries  
Session Chairs

2:40 pm – 3:00 pm Summary of the workshop and closing remarks  
Trey Ideker and Claire Tomlin

---

### 3:00 pm **ADJOURN**

---