CANDLE ECP - Preliminary Workshop at NIH
April 18-19, 2017
Workshop Summary

The first CANDLE Workshop at NIH, held April 18-19, 2017, proved to be a tremendous success in weaving the cancer research community into the emerging computational architecture being developed through CANDLE.

A clear demand for deep learning was expressed among the research community for a variety of applications, many of which introduce new challenges feeding directly into the expanding CANDLE environment.

Participation in the workshop included roughly 60 attendees spanning 13 NIH institutes, listed by breadth of attendance as follows:

- NCI
- NHLBI
- NLM
- DRD/Warren Grant Magnuson CC
- CIT
- Netrias LLC in collaboration with NCI, NINDS, NIAAA
- NHGRI
- NIA
- NIDCD
- NIDCR
- NIDDK
- NINDS
- IRSB

*Please refer to attached attendance list for full roster.

The agenda of the two-day workshop included a high-level overview of deep learning and the CANDLE project, followed by a hands-on exploration of the environment as applied to emerging cancer challenge areas.

The overview and discussion elements of the workshop were led by Eric Stahlberg of *Frederick National Laboratory for Cancer Research*, and Rick Stevens of *Argonne National Laboratory*.
Technical presentations and exploration of the CANDLE environment were led by the following instructors:

- Fangfang Xia, *Argonne National Laboratory*
- Tom Brettin, *Argonne National Laboratory*
- George Zaki, *Frederick National Laboratory for Cancer Research*
- Brian Van Essen, *Lawrence Livermore National Laboratory*
- Arvind Ramanathan, *Oak Ridge National Laboratory*

*Please refer to attached for full workshop agenda.

Key challenges were identified among participants throughout the workshop, primarily focused on the application of DL to imaging and genomic data.
Discussion points may be characterized as follows:

1. Imaging: pressing interest in extending DL environment to include imaging domain
    a. Image segmentation, phenotype analysis
    b. Supervised/unsupervised learning
    c. Applying DL to biological images
2. Data interpretability
    a. Desire to explore visualization/interrogation methods being applied to DL (e.g. feature analysis, significance plots)
3. Application/feasibility of executing DL in the cloud

4. Computational linguistics, NLP
    a. Desire to improve performance/throughput expectations, and resolving the challenge of spelling inaccuracies
5. Using DL for predicting small molecule in vivo activity against cancer outcomes
6. Using DL/ML to classify data by tumor type
7. Genotype/phenotype analysis: predicting phenotypes from genotypes

Key outcomes of the meeting include:

1. Request to establish a trans-NIH group on the topic of DL. There are DL thought leaders across NIH, but working in relative isolation.
2. Plan for a workshop on DL applied to genomic data.

The event was very well received overall.