This event was the first in a series of events on justice in an era of big data, one of the Center’s themes for the year. The working group meeting was a conversation between Peter Yu (incoming President of the American Society of Clinical Oncology and Director of Cancer Research at the Palo Alto Medical Foundation) and David Haussler (Director of the UCSC Center for Biomolecular Science and Engineering) about genome data and the future of cancer research. Julie Harris (Assistant Adjunct Professor at UCSF, Institute on Health and Aging; Staff Scientist at Kaiser Permanente Division of Research; and Associate Director of the Center for Translational Genomics and Ethics) provided commentary. Science & Justice Center Director Jenny Reardon moderated the conversation and introduced the panelists.

Reardon’s introduction provided an overview of some of the concerns that the working group hopes to pursue with this series. Genome research is seen as powerful, and cancer research can now studies the genomic changes that occur during the development of cancer. The techniques that were developed in Haussler’s lab to understand the human genome are now being used to think about cancer and evolution. This kind of genomic research would benefit greatly from additional data that could be collected from cancer patients, but doing so raises ethical, epistemological, and infrastructural questions. As a society, we have yet to figure out what to do with big data. At present we mostly collect data of unknown significance, but there is no clear precedent for who governs it, how to store it, or how to make sense of it. Who pays to store it? Who gets to work with it and try to make sense of it? As the first of many working group meetings to discuss this issue, the goal for this meeting was to outline which of these questions needs further discussion.

Peter Yu spoke first, speaking from the perspective of a doctor practicing clinical oncology. He described the efforts of the American Society of Clinical Oncology (ASCO) to accelerate learning and analysis through computerized health care. They envision a rapid learning system model, which would allow clinicians to generate new data and better models while treating patients, by incorporating data from clinical practice instead of just clinical trials. Such a system would require that patients and doctors be willing to share data, and it presents problems for managing data, such as standardizing it and safeguarding it in a centralized repository. The organization is still in the process of funding these efforts, but they are trying to address these ethical and epistemological questions before they arise.

David Haussler followed up by first thanking Yu for his organization’s efforts, which he sees as a tremendous boon to cancer care. Speaking from the point of view of a data scientist, Haussler argues that big data is absolutely necessary for cancer research. Most mutations are insignificant and very few are meaningful, so in order to establish a clear understanding of the drivers of cancer, there needs to be a large number of genomes available to work with. These numbers are unobtainable in the current system of clinical trials and academic research, but they would be accessible if information could be captured from clinical practice. Haussler is hoping that under Yu’s guidance the ASCO will be able to incorporate data collection into medical practice. Yu agreed, saying that he believed the “holy grail” for cancer research would be a healthcare system that engages people in research without sacrificing their rights.
Julie Harris provided her comments at this point, reminding us that the strong division between research and clinical practice was established as a response to the Belmont report. At the time the ability to distinguish between the two was useful, but times have changed. Big data has brought new challenges, but a lack of community involvement in research continues to be a concern. With Kaiser, she has been involved in a project to build a biorepository that members can volunteer to donate samples to. The samples are linked to clinical records and environmental databases so that they may be used to research gene-environment interactions. This project has been successful so far. Harris attributes at least part of the success to a community advisory panel that brings together diverse representatives of the public. According to Harris, many of the participants don’t fully understand the program but trust Kaiser to use the information in a way that may benefit them someday.

Trust was a primary concern during the question and answer session. Some audience members were concerned that storage for big data is not secure; that the information could be accessed by governments or individuals with malicious intent. This is especially problematic when the information could easily be de-identified using phenotypic information. Yu mentioned that there had been a study on establishing and maintaining trust around research samples, and that most people were more concerned with what the information was used for than who was using it. This is troublesome for two reasons, one, because it is difficult to anticipate what the information might be used for in the future, and two, it is not always clear who the “who” is that might eventually use the data, as institutions are increasingly amorphous. Researchers often try to maintain trust by assuring donors that the samples will be used for good, but the notion of good is itself abstract and a part of the question of justice that the working group will continue to explore at future events.