Testimony of

Dr. Andy Futreal, Chair of the Department of Genomic Medicine at The University of Texas MD Anderson Cancer Center

Subcommittee on Space, Science, and Competitiveness hearing on "The Dawn of Artificial Intelligence"

November 30, 2016

Subcommittee Chairman Cruz, Ranking Member Peters, and members of this committee, thank you all very much for the opportunity to testify before you today. My name is Andy Futreal and I am Chair of the Department of Genomic Medicine at The University of Texas MD Anderson Cancer Center.

We are now entered into a completely unprecedented time in the history of medicine. We have the ability to investigate the fundamental molecular underpinnings of disease, to leverage technology and computational capabilities with the real prospect of fundamentally altering the natural history of disease. We can now determine each individual's genetic blueprint with relative speed and accuracy at a cost of less than a millionth of the price tag of the first human genome sequenced just a little more than 13 years ago. We are moving into an era of tackling the sequencing of very large groups of individuals and defining the role of common variation, that which is shared by more than 1-5% of the population, in health, risk and disease. The challenge of reducing this watershed of data into practical implementation to improve human health and provide better care for patients is upon us. The opportunities to improve and tailor healthcare delivery – the right drug for the right patient at the right time with the right follow-up – are being driven by exploiting computational approaches and so-called "big data". AI and machine learning approaches have the potential to help drive insights and deliver improved standards of care. Taking oncology as the proving ground where a very great deal of these efforts are currently focused, there are several challenges, opportunities and issues that present themselves.

The clinically meaningful implementation of machine-assisted learning and AI is, of course, crucially dependent on data - lots of it. Herein lies perhaps the biggest challenge. Substantial and varied clinical data is generated on every patient cared for every day. These data are generally held in non-interoperable systems whose principle purpose is to facilitate tracking of activities/services/tests for billing purposes. The richest clinical data is effectively locked in various dictated and transcribed notes detailing patients' clinical course, responses, problems and outcomes from the various treatments/interventions undertaken. We need to further develop capabilities to both get these data from their source systems and standardize their ongoing collection as practically as possible.

As well, a proportion of those under our care take part in research studies, generating research data in both the clinical and more translational/basic science realms. These data, including

increasing amounts of detailed large-scale genomic sequencing information, are not generally available for integration with clinical data on a per-patient or aggregate basis in a way that would facilitate implementation of advanced analytics. The ability to purposefully integrate clinical and research data for analytics, without the need for predetermining and rigidly standardizing all data inputs up front is what is needed.

There are substantial opportunities for AI, again anchoring in oncology by way of example. Perhaps the most concise way of framing where we need to be headed, in my view, is the concept of real-time "patients like mine" analytics. Leveraging clinical, molecular, exposure and lifestyle data of patients that have been treated before to understand and predict what the best choices are for the current patient. But even more so, not just choice of therapeutic but how to improve and intercede as needed in management such that positive outcome chances are maximized. We need to make predictive analytics the norm, learning from every patient to improve the outcome of the next. Importantly, we need to be thinking now about training our best and brightest in the next generation of physicians and medical professionals to drive this progress, as it will take a new wave of computationally savvy individuals to build, train and grow these systems. Further, we need to think carefully about how we promote data sharing, particularly in the clinical arena. Open access is a laudable goal, but one that must be tempered with the relevant privacy and security practices. Facilitated collaboration on specific topics with honest broker mechanisms to demonstrate rapid progress and real value in data sharing early will, I think, be key.

At MD Anderson, we have been exploring the possible utilities of AI and related technologies in collaboration with IBM. We are utilizing the Watson platform for cognitive computing to train an expert system for patient-centric treatment recommendation and management. Currently, we are evaluating performance in the context of lung cancer. Future work reflects the challenges and opportunities that the entire field faces – namely that of what to deploy in the near-term where dissemination of expert knowledge in the context of rule-based approaches could have significant impact on potentially improving standard of care and where to take efforts in the longer term with learning, AI type approaches.

The ability to have data-driven, AI empowered point-of-care analytics holds the promise of improving the standard of care in medically underserved areas, of guaranteeing that every patient –regardless of zip code – can be assured of up-to-date and appropriate care taking into account their own particular data and circumstance. A massive undertaking to be sure, but one that is, I believe, within our collective grasp.

I thank you again for the opportunity to testify before this committee and I would be happy to answer any questions you may have.