The term big data has become a routine word across many disciplines. The big data in medical terms generally encompasses Next Generation Sequencing (NGS) of the genome from individual patients, mRNA expression landscape of normal and diseased tissues, biobank tissue-derived information, clinical trials, drug efficacy and toxicology data and electronic medical records linked to medical imaging and insurance claims data. During his State of the Union address (January 12, 2016), President Barack Obama announced the establishment of a Cancer Moonshot initiative to accelerate cancer research. This initiative, led by Vice President Joe Biden, aims to make therapies available to a large number of cancer patients and is projected to improve cancer prevention and detection at an early stage. Recently (May 2016), the White House released The Federal Big Data Research and Development Strategic Plan, which provide guidance for developing or expanding Federal Big Data research and development (R&D) plans.

The Accelerating Medicines Partnership (AMP), a new venture involving the US National Institutes of Health (NIH), 10 biopharmaceutical companies, and several nonprofit organizations, has an initial fund of $230 Million. The overall goals are to transform the current approaches for diagnostics and treatments to a new dimension using big data analytics by jointly identifying and validating promising biological targets of disease. The initial therapeutic areas include Alzheimer’s disease, Type 2 diabetes and two autoimmune disorders, rheumatoid arthritis and systemic lupus erythematosus (lupus). The European drug research consortium projects that they will invest more than $5 billion in the next several years to apply big data techniques termed “Big Data for Better Outcomes,” to speed up clinical drug trials while developing a sustainable healthcare delivery system. In the UK, the National Institute for Health Research (NIHR) has put in place a series of initiatives to help exploit the nation’s strengths in technology, medical research and healthcare data. The Genomics England Project is expected to generate a vast amount of genetic information from 100,000 patients with an initial focus on cancer, rare diseases and infectious diseases.

Among numerous therapeutic areas, cancer research area has accumulated huge amounts of big data. This includes datasets from thousands of patients encompassing gene expression, mutations, deletions and amplifications and proteogenomics data. Increasingly, the basic research in cancer is integrated to translational medicine in an attempt to move the discoveries closer to the clinic.

The catalogue of somatic mutations in cancer (COSMIC): The COSMIC database is the world’s largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer. The latest release (v70; Aug 2014), describes 2,002,811 coding point mutations in over one million tumor samples and across most human genes.

The integrated cancer knowledgebase (canSAR): The canSAR database applies machine-learning approaches to provide drug-discovery predictions. The growing database now holds the 3D structures of almost three million cavities on the surface of nearly 110,000 molecules and

The national cancer institute’s clinical proteomic technologies for cancer initiative: This database leverages proteogenomics analysis through the development of the Clinical Proteomic Tumor Analysis Consortium. This consortium is composed of Proteome Characterization Centers, Data Center, and Resources Center, to produce a unique continuum that defines the proteins translated from cancer genomes. This integrative approach provides the broad scientific community with knowledge that links genotype to proteotype and ultimately phenotype. The data sets, analytically validated assays, as well as high quality reagents are publicly accessible. These efforts together with other NCI programs; e.g., the NCI’s Cancer Therapy Evaluation Program (CTEP), the Early Detection Research Network (EDRN), the Cooperative Groups have broadened the scope of cancer research from the bench to bedside.

Other cancer-related metadata includes the Oncomine® Gene Browser (ThermoFisher Scientific) dataset which harbors comprehensive gene profiles across thousands of cancer patient

**Keywords:** big data, biobank, cloud computing, cancer, electronic medical records, genomics, proteogenomics
Big data analytics and cancer

These issues notwithstanding, one can anticipate that the big data infrastructure should help the oncologists and cancer patients around the globe in decades to come. The big data cancer analytics with data encompassing clinical trials to real-world patients and practices can provide answers to effectiveness of treatment and long-term outcome.

Acknowledgements
None.

Conflict of interest
The author declares no conflict of interest.

References
13. Boonstra A, Broekhuis M. Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions. BMC Health Serv Res. 2010;10:231.

Big data analytics and cancer


44. Kasprzyk A. BioMart: driving a paradigm change in biological data management. Database. 2011;2011:bar049.


