Since the identification of Nitrogen mustard as a non-specific DNA alkylating agent and its clinical assessment by Gilman and Goodman at Yale University in 1942 for Lymphoma treatment [1], the field of medical oncology has rapidly expanded. It relies on understanding and exploring numerous factors affecting cancer patient's outcome, including:

1) Familiarity with the natural history of various malignancies [2] and understanding the significant of patients' well-being as predictor to survive and tolerate therapy, scored as the "Performance Status" [3,4];

2) Basic biology events of normal cycling cells [5], and principals involved in cancer pathogenesis named "hallmarks of cancer," including sustaining proliferative signaling, evading growth suppressors, resisting cell death, enabling replicative immortality, angiogenesis, activating invasion and metastasis, and recently expanded to include also reprogramming of energy metabolism and evading immune destruction [6,7];

3) Identification of tumor features or biomarkers for tailored treatment approach including: 'targeted' therapy (i.e., estrogen receptor expression for hormonal manipulations in breast cancer [8]; Her2 overexpression for anti-Her2 agents in breast and gastric tumors [9]), mutations affecting up/down stream biological events, and identifying patients who will not benefit from blocking single protein in a chain process (i.e., RAS mutation signaling lack of benefit from anti-EGFR directed monoclonal therapy in colorectal tumors [10]), and those of unique molecular/genetic conditions signaling tumors prone to specific therapy (i.e., BRCA1/2 mutations and benefit of PARP inhibitors [11]);

4) The role of tumor stroma and its interaction with cancer and anti-cancer drugs [12,13];

5) Pharmacological factors such as those affecting drug absorption and delivery, metabolic production and elimination of active agents from their pro-drugs (i.e., SN-38 production from Irinotecan infusion [14]), and the effect of genetic heterogeneity in metabolizing enzymes, concomitant medications use [15], and liver and kidney function on active drug bio-availability;

6) And above all, the proof of efficacy and safety of tested drugs in well conducted clinical trials with specified clinical outcome [16].

While understanding of each of those factors contributes to the emerging progress in medical oncology, many research and clinical publications focus on only few factors, mostly on those related to a specific new drug or a matched biologic process. Indeed, in real life and post-marketing world, many additional drug/tumor/co-morbidity/ethnic and geographic factors/social status/economic status/insurance coverage and educational factors, contribute to patient ability to access and benefit from anti-cancer care. Pooling information from multiple resources may better uncover some of those 'ingredients' but requires highly accurate and consistent mathematical and algorithmic methods and reliable wide database information at the patient ID level. In an effort to initiate such a process we developed a computational tool for real life data mining in HMO patients with metastatic colorectal cancer treated with first line bevacizumab-based chemotherapy regimens [17]. Patients were identified from Pharmacy records and were followed till death or last follow up if alive. For each patient, information from multiple HMO sources was obtained, including co-morbidity registries, information on purchasing and using other oncological and non-oncological drugs, laboratory tests, hospital claims for chemotherapy infusion, and surgical bills. For each patient, a longitudinal time-line life event was assembled allowing assessment of anti-cancer treatment lines and numerous oncology and non-oncology parameters. Tough retrospective in nature, information allowed data assessment in subsequently operated (#116) and non-operated (#637) patients, alone with focusing on basic differences between groups both demographics and by treatment approach. In the non-operated patients, median survival of only 18.7 months was reported despite HMO coverage for bevacizumab (highly expensive drug) and its use in the first line regiment for 10-11 months (before switching to 2nd line regimen). Numeric longer survival was observed in patients receiving highly active chemotherapies (with the biological agents) with both Irinotecan and Oxaliplatin, (as opposed to those treated with only one of those agents) and in those treated with anti-EGFR antibodies (Cetuximab or Panitumumab) for RAS-wild type tumors. Earlier failure of first line bevacizumab-based regimens was observed with specific laboratory 'finger prints' and with regular use of certain medications (i.e., diuretics and proton-pump inhibitors). We propose that in the future such a toll at the ID level could be implemented into the HMO computerized systems, and will accumulate information from on-going patients' inclusion. It will allow assessment of real life treatment approaches and effect of serial medications, as well as the evaluation of benefit from highly expensive or toxic drugs. In addition, it may serve attending clinicians caring for cancer patients who may have several treatment line options, allowing them to query the system for questions such as 'what is the best chemotherapy-biological regimen to use in the first line in a 73 years old male with metastatic colorectal patients, who also has diabetes and coronary heart disease, and uses regularly anti-depressants, diuretics and beta-blockers". While we are not there yet in oncology care, other fields of research and industry are already relied upon data-driven mathematical and algorithmic...
systems, including speech understanding (Apple’s Siri), gaming (Chess, AlphaGo), and neuroscience research [18].

The data-driven systems are typically based on one of two types of Artificial Intelligence (AI) methods: (1) Symbolic Artificial Intelligence, e.g., expert systems or Bayesian networks [19], where each symbol has an assigned meaning such as the age of the patient, whether or not a particular medication is taken, or the probability or reliability of an estimate; or (2) Subsymbolic Artificial Intelligence methods, most prominently neural networks with deep learning or recurrent update rules [20,21], where many parameters describe the connectivity patterns of artificial neurons, and these parameters are trained from databases to create optimal mathematical descriptions to the problems at hands. In the future, data driven AI will likely affect medical care as well, allowing unrevealing of yet unknown factors affecting patients’ outcome, and could assist clinicians to adapt the best treatment algorithms based on accumulated information of already treated patients.

References