BEAGLE

Science Highlight

Data Integration and Statistical Modeling for Cancer Genomics

Yitan Zhu, Subhajit Sengupta, Yang Ni (UT Austin), Shengjie Yang, Lin Wei, Tianjian Zhou, Lorenzo Pesce, Ian Foster, Peter Mueller (UT Austin), Yuan Ji

ABSTRACT

Yuan Ji (Northshore University HealthSystem and The University of Chicago) and his group apply Bayesian models to solve problems in cancer research, such as system-wide genetic interaction, tumor heterogeneity, and adaptive designs for precision clinical trials. In all these problems, large-scale computing resources are required for assessing the statistical models and performing statistical inference.

INTRODUCTION

[Zodiac] Back in 2015, the group conducted a genome-wide integration of multiple -omics data (Figure 1) in The Cancer Genome Atlas (TCGA) and spent about 2.5 million CPU hours on Beagle to generate a comprehensive database called Zodiac . The related papers are published in Nature Methods and JNCI. The latter included several researchers as co-authors, including Drs. Ian Foster and Lorenzo Pesce from the Beagle team itself. Both papers are accessed extensively by researchers worldwide.

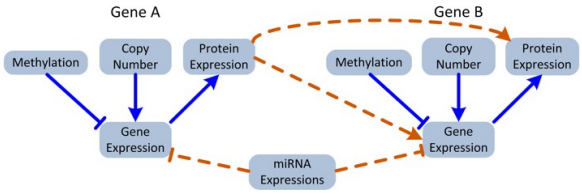


Figure 1. TA multi-omics interaction network between two genes in Zodiac.

METHODS AND RESULTS

[Tumor Heterogeneity] An ongoing large-scale project is related to intra-tumor heterogeneity, in which Bayesian feature allocation models are developed to analyze deep DNA-Sequencing data aiming at reconstructing clonal expansion of tumor cells via somatic mutations. A sequence of papers have been published, including BayClone1/2, LocHap , and TreeClone. The current project is to build a high-performance computational pipeline to analyze TCGA and ICGC whole-genome sequencing data on tens of thousands of tumor samples. The goal is to evaluate the subclonal phylogeny (Figure 2) of tumors in a pan-cancer fashion, and understand the population behavior of tumor cell evolution. Some of the recent press reports can be found in: MAD Bayes for Tumor Heterogeneity—Feature Allocation With Exponential Family Sampling ,
Statistical technique helps cancer researchers understand tumor makeup, personalize care.

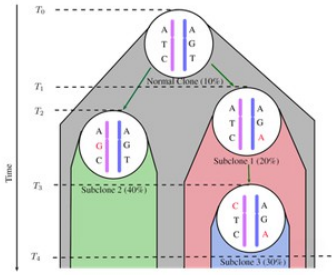


Figure 2. Subclonal evolution based on somatic mutations accumulated over time.

[Adaptive design for precise clinical trials] Recently, Yuan's group conducted large-scale simulations to develop and assess Bayesian adaptive designs for clinical trials. For example, they published a paper [Xu Y, Trippa L, Muller P, Ji Y. Sub-group Based Adaptive (SUBA) Designs for Multi-Arm Biomarker Trial. Statistics in Biosciences, 2016, and the AAA design (Lyu et al., 2017, In revision). These designs are aimed at speeding up drug development by providing better mechanisms for the evaluation process of human clinical trials. In addition, through collaborations, Yuan's group also pioneered the development of designs for two-cycle dose finding trials in oncology, which has drawn a press release:
Statisticians develop new 2-cycle dose-finding method for personalized cancer treatments . Using Beagle2, they extended this method to incorporate ordinal outcomes that differentiate the severity of toxicity. A paper based on this work has been published in JASA: Bayesian Dose-Finding in Two Treatment Cycles Based on the Joint Utility of Efficacy and Toxicity and was broadcast in the press by many news outlets such as Science Daily:
Statisticians develop new 2-cycle dose-finding method for personalized cancer treatments.
[Imaging Genomics] Lastly, in collaboration with Dr. Maryellen Giger at the University of Chicago, Yuan's group has developed novel methods and made new discoveries in the field of radiogenomics
Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma aiming at integrating genomics and radio-mics features for cancer diagnosis and prognosis (Figure 3). All of these works are made possible through the availability of the Beagle1/2 computing power.

Resouces:

Beagle Wiki
Get detailed usage information from the Beagle2 team

Beagle Support
Contact the Beagle experts for help

Globus
For file transfer. Get started moving files to/from Beagle2 using this fast service

Other CI resources
Learn about other computing resources available at the Computation Institute

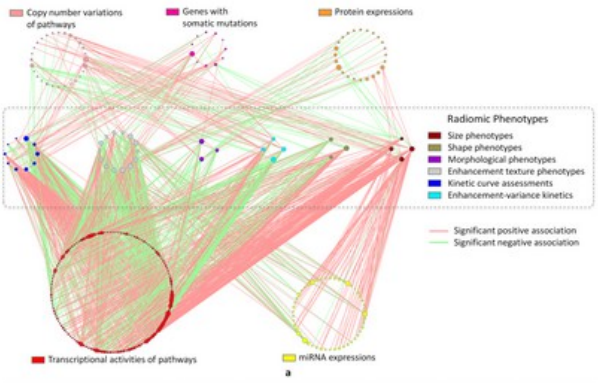
Training:

Statistics and Probability using Python
February 5th, 10AM
Room 240A, at the Computation Institute of the University of Chicago

Topics will include:

- Counting and combinatorics
- Discrete and continuous probability
- Conditional probability and Bayes' Rule
- Random variables
- Expectation, variance, and correlation
- Probabilistic inequalities and concentration
- Moments and limit theorems
- Hypothesis testing
- Sampling and confidence intervals
- PCA and regression
- Entropy and compression

Beagle2 Events To learn more about our trainings.



	Size phenotypes	Shape phenotypes	Morphological phenotypes	Enhancement texture phenotypes	Kinetic curve assessments	Enhancement-variance kinetics
Transcriptional activities of pathways	173	109	49	374	268	130
Copy number variations of pathways	24	7	7	14	15	21
Mutated genes	3	1	1	15	22	3
miRNA expressions	73	0	0	58	1	0
Protein expressions	10	0	9	17	0	0

b

Figure 3. Imaging genomics analysis reveals association of genomcis features and imaging features.

There are some major new initiatives from Yuan's group in 2017. First, Zodiac-2 is currently underway which will require multi-million CPU hours. In Zodiac-2 cancer-type-specific multi-omics pathways will be inferred using Bayesian graphical models, and differential network analyses will be implemented to compare network differences between different phenotypes. Second, tumor heterogeneity analysis will be conducted for TCGA data, which requires large-scale bioinformatics processing of next-generation sequencing data and inference based on Bayesian feature allocation models. Third, many new adaptive designs will be developed to speed up cancer drug development. These designs will be tested using large-scale computer simulations which will be conducted on Beagle 2.
Yuan's group looks forward to continue their successful collaboration with the researchers at the CI.

Additional information about Science on Beagle can be found here: Beagle2 website

Beagle Related Publications

Manching H, Sengupta S, Hopper KR, Polson SW, Ji Y, Wisser RJ.
Phased Genotyping-by-Sequencing Enhances Analysis of Genetic Diversity and Reveals Divergent Copy Number Variants in Maize
G3 (Bethesda). 2017 Jul 5;7(7):2161-2170. doi: 10.1534/g3.117.042036. PMCID:PMC5499125

Wentian Guo, Sue-Jane Wang, Shengjie Yang, Henry Lynn, Yuan Ji
A Bayesian interval dose-finding design addressing Ockham's razor: mTPI-2
Cont. Clinical Trials. July 2017. Vol. 58, Pg. 23–33, DOI: http://dx.doi.org/10.1016/j.cct.2017.04.006

Rangarajan D. Nadadur, Michael T. Broman, Bastiaan Boukens, et al.
Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm
Sci Transl Med. 2016 August 31; 8(354): 354ra115. PMCID: PMC5266594

C.K. Li, P. Tzeferacos, D. Lamb, G. Gregori et al.
Scaled laboratory experiments explain the kink behaviour of the Crab Nebula jet
Nature Communications, DOI: 10.1038/ncomms13081, 7 Oct 2016

Esmael J. Haddadian, Hao Zhang, Karl F. Freed, and Jack F. Douglas
Comparative Study of the Collective Dynamics of Proteins and Inorganic Nanoparticles
Sci Rep. 2017; 7: 41671. doi: 10.1038/srep41671, PMCID: PMC5296861

Searle Chemistry Laboratory
5735 South Ellis Avenue
Chicago, IL 60637

beagle-support@lists.uchicago.edu

beagle.ci.uchicago.edu