Generating and Validating Cancer Mutation Trees

Stanford | ENGINEERING
Electrical Engineering

Christian Choe[1,2], Min Cheol Kim[1]
Departments of Electrical Engineering[1] and Chemical Engineering[2], Stanford University

Stanford | ENGINEERING
Chemical Engineering

Introduction

Single cell exome sequencing of cancer cells can give us information regarding their progression. Specifically, we can infer the order in which specific genes get mutated for normal cells to transform into cancer cells. We present a Bayesian technique to infer the mutation orders and a validation procedure for our algorithm.

Results

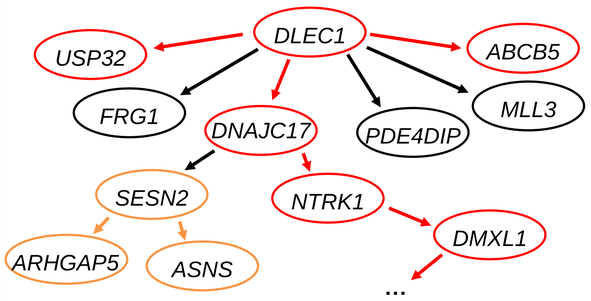


Figure 4. Mutation tree for essential thrombocythemia (ET) tumor. The tree was generated using Hou et al.'s data from a patient. From the data, we chose to examine the eighteen mutations that Hou selected as important in order to build a similar mutation tree as Kim and Simon.

Comparing our ET mutation tree to theirs, the two trees have a similarity score of 0.65. Although the similarity score may seem low, our tree has the same parent node as well as all the correct leaf nodes. The colored nodes and arrow show the substructure in our trees that are also found in Kim and Simon's tree.

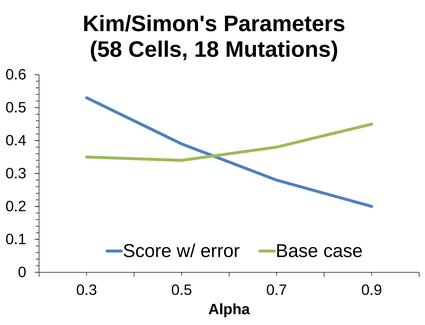
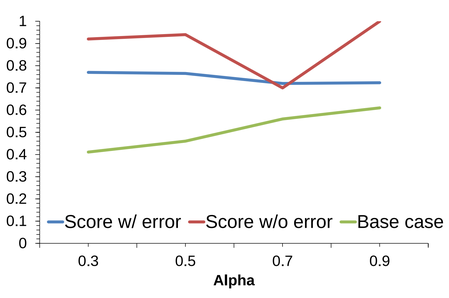
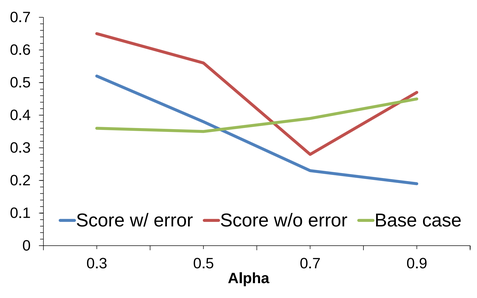


Figure 5 Simulation Results. (Left) We see that with these parameters, the algorithm does not perform significantly different than the baseline. (Bottom, Left) With large cell counts and small number of mutations tracked, we can always stay above the base-line. (Bottom, right) Similar to Kim and Simon's experiment, simulation results with 50 cells and 19 mutations show that algorithm does not perform better than the base-line.





Method

Step 1 - Inferring Pairwise Relations Between Mutation Sites

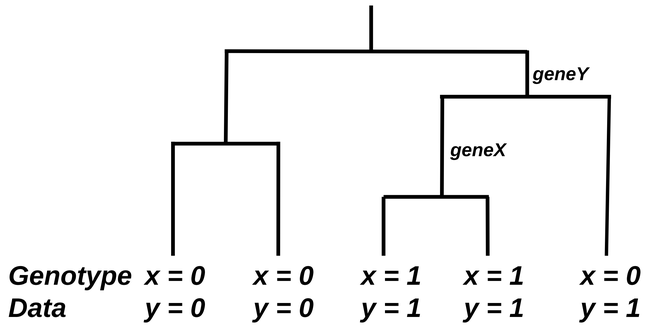
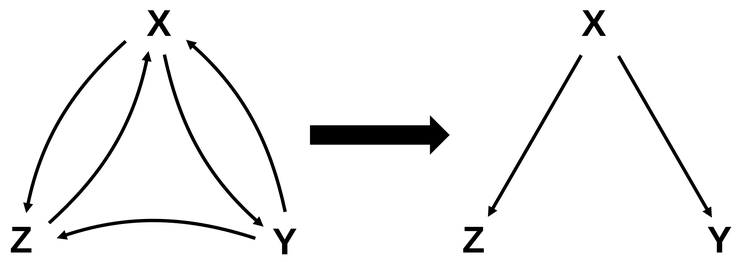


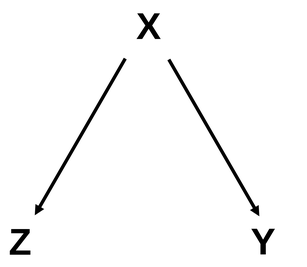
Figure 1. Genealogy tree. This tree represents one possible genealogy that generates the given genotype data. Here, two mutations at gene X and gene Y occurred at marked points on the tree. The five samples with integer genotypes x and y are attached to the leaves of the tree. The root is assumed to be wild-type and have the genotype of ($x = 0$, $y = 0$). The top of the tree represents the cell line before mutagenesis and bottom represents the current state.

Step 2 - Constructing overall mutation relationships

Figure 2. Mutation relationship. The tree on the left shows all the relationships between each mutation X, Y, and Z. And the tree on the right is the minimum spanning tree. The arrows represents the temporal relationship between the mutations. The weights of the edges are equal to $-\log[\Pr(i \to j)]$ so the higher the probability of that relation, the smaller the weight.



Step 3 - Tree Validation and Simulation



Similarity: $\frac{2}{3}$

Figure 3. Tree comparison. When generating trees we compared the predicted relationships to the true relationship to measure their similarities. The similarity ranges from 0 to 1 where 0 represents no similarities and 1 represents an exact match. We measure similarity by looking at two nodes and seeing if those two nodes have the same relationship in the other tree (e.g. nodes are in the same lineage and are in the same temporal order). The two trees above have a similarity of 2/3 because the Z and Y relationships differ. Using this technique, we tested our algorithm with many testing parameters.

Summary and Future Works

We were able to successfully infer a mutation tree similar to that of Kim and Simon, and also devised a framework to test our algorithm in numerous scenarios. In addition, we discovered that our algorithm is effective when the number of sampled cells are much greater than the number of mutation sites examined; based on our analysis, we question the validity of the mutation tree generated by Kim and Simon, as they had relatively small number of cells but tracked many mutations.

In the future, we may want to try a machine learning approach to predicting tree structure. Another future challenge would be inferring the alpha value from the given data, which was not in the scope of this project.