Population dynamics modeling and Bayesian model selection reveal behaviors governing small cell lung cancer tumor composition

Samantha P. Beik[1], Leonard A. Harris[8], Sarah Groves[2], Alissa Weaver[3,4], Vito Quaranta[5,6], Carlos F. Lopez[2,3,4]

Cancer Biology Graduate Program[1], Chemical and Physical Biology Graduate Program[2], Dept of Biochemistry[2], Dept of Cell and Developmental Biology[3], Dept of Pathology, Microbiology, and Immunology[4], Dept of Biochemistry[5], Department of Pharmacology[6], Dept of Biomedical Informatics[7], Vanderbilt University, Nashville, TN, USA; Dept of Biomedical Engineering, University of Arkansas, Fayetteville, AR[8]

Background. Small cell lung cancer (SCLC) has been studied in different model systems and determined to have 4 subtypes, which interact with each other.

- SCLC tumors are heterogeneous, consisting of different cell subpopulations.
- At this time, the field accepts 4 cell subtypes: A, N, P, Y.
- Our group has predicted an additional subtype within the A group, named A2.
  - While all cell populations in the A group are positive for Achaete-Scute Family BLHL Transcription Factor 1 (ASCL1), some of these populations express Hes1+, a Notch pathway activator that typically causes repression of ASCL1 (yet does not in this case).
- We have applied CIBERSORT to RNA-seq data from genetically engineered mouse model tumors provided to us by Dr. Julien Sage at Stanford. In this way w can predict the proportion of each tumor that falls in each subtype A, A2, N, Y.
  - We have not found subtype P in this particular dataset.

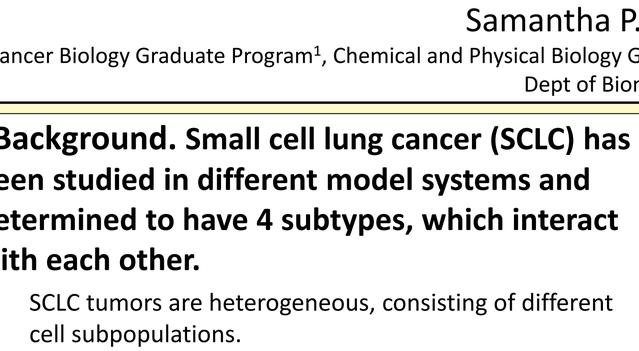

FIG. 1. Four consensus subtypes plus an additional subtype our previous work has predicted. Figure adapted from Rudin CM, et al. (2019) *Nature Reviews Cancer.*

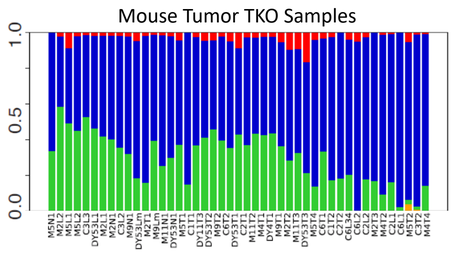

FIG. 2. CIBERSORT deconvolution of triple-knockout (TKO) GEMM tumors. TKO tumors are the result of conditional knockouts (*p53$^{flox/flox}$;Rb$^{flox/flox}$;p130$^{flox/flox}$*). Figure from Wooten, Groves, et al. (2019) *PLOS Comp Biol.*

Construction of the population dynamics model and performance of Bayesian model selection.

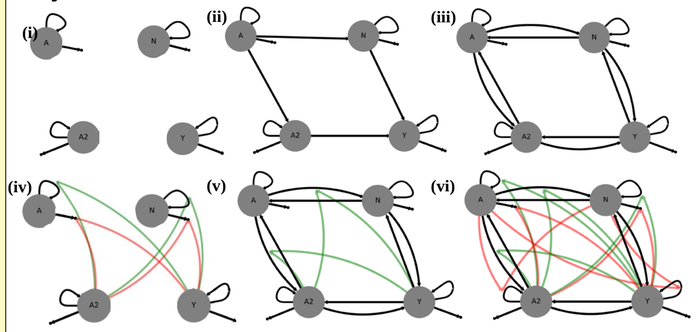

FIG. 3. Features of potential population dynamics models and examples of model possibilities. (i) Each subtype divides and dies. (ii) Each subtype can transition to one or more other subtypes (not all possibilities shown). (iii) Transitions can be bidirectional. (iv) Certain subtypes interact with others by supporting their growth (green lines indicate promotion of cell division, red lines indicate inhibition of cell death). (v) Subtype interactions also include promotion of certain transitions. (iv) Model schematic showing bidirectional transitions, division and death supports, and transition supports.

FIG. 4. Bayesian Model Selection. Compares how well the model fits the data and penalizes a model for overfitting.

- Population dynamics modeling tracks the number of cells over time, and can track which cells are part of which subpopulation during growth simulations.
- We investigate 88 combinations of transitions and inter-subtype effects. Consequently, the models go from simple to much more complicated.
- We use principles of Bayesian model selection to investigate the ability of a model to match the provided data; we thus assess which model features are supported by the data at hand.

Results: a confidence set of more likely models and weighted parameters to indicate more or less likely model actions.

- Model selection evidence scores indicate a subset indistinguishable from the perspective of Bayesian evidence; that is, equivalent in quality given the data at hand.
- **We can conclude from this:**
  - Bidirectional transitions are supported by the data.
  - Inter-subtype effects that increase the speed of transitions are not supported.
  - The data cannot indicate whether the remaining transitions, or inter-subtype effects that support overall cell growth, are likely to exist.

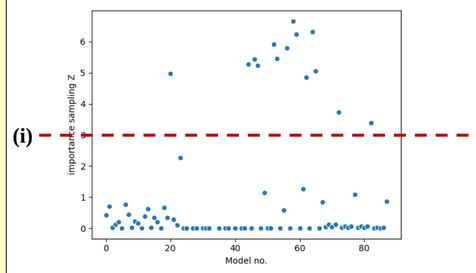

FIG. 5. Results of evidence calculation for each model. (i) A confidence set of models: models for whom the Bayes Factor comparing to the highest Z-score is more than or equal to 0.35.

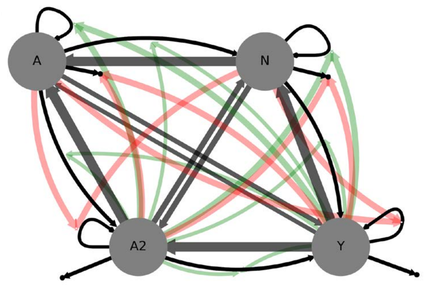

FIG. 6. Weighted model actions. Each parameter is scored based on the normalized evidence values for the models in which it is included. Each parameter (arrow) is thus given a weight (between 0 and 1).

Future directions: model selection using *in silico* data to predict experimental design

- Generate *in silico* data, using one or more "high confidence" model (see Figure 5(i)).
- Perform model selection repeatedly using different combinations of datapoints: different numbers of datapoints as well as different interval timing.
- The simplest timepoints and intervals that can result in a smaller confidence set will be the basis for future experimental design.

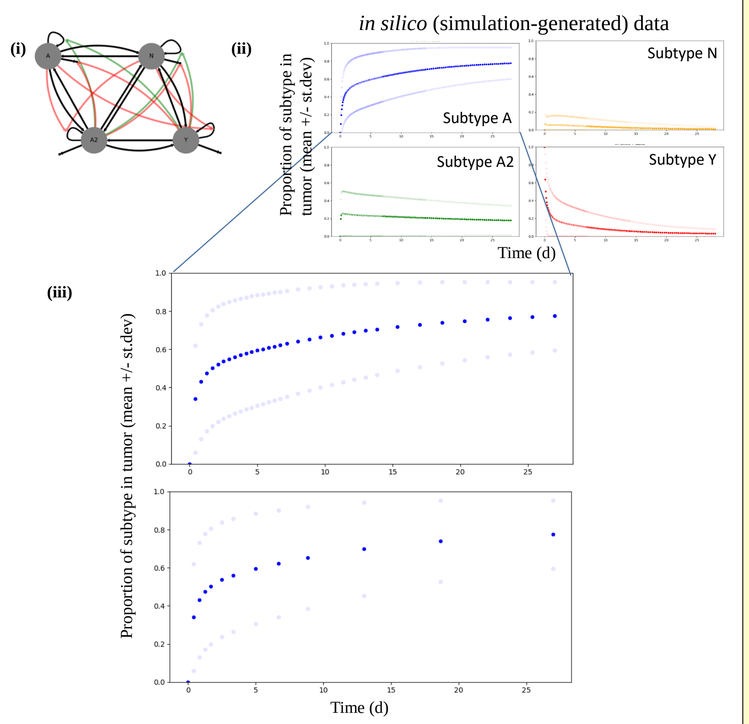

FIG. 7. Performing model selection on *in silico* data. Top: (i) simulating tumor growth using one of the high confidence models gives (ii) proportions of each subtype in a tumor over time. Bottom (iii): subtype A simulations as an example for the process. We will plan to take tumor subtype proportion measurements at what we find to be the optimal timepoints, by performing model selection twice using each of these data sets, choosing the timepoints for whichever model selection outcome provides the smallest (and thus most certain) confidence set compared to number of points. For example, if model selection performed on each of these gives the same confidence set of likely models, we would choose the lower timepoints, since we predict we can get the same information on cell behavior with fewer measurements.

Acknowledgements

Funding

- T32GM007347 F30CA247078 T32LM012412
- NIH National Cancer Institute (VQ, CL; U54, U01)

Special thanks to Dr. Vito Quaranta, Dr. Carlos F. Lopez, and Dr. Leonard Harris for mentorship and advice.