**Integrating bioinformatics into multiscale models for hepatocellular carcinoma (U01CA212007)**

**Model Credibility Plan**

Our aim is to develop new mathematical techniques to model hepatocellular carcinoma (HCC) at multiple scales (molecular, cellular, tumor, organ, whole patient) to elucidate the mechanisms driving the progression of the cancer and to predict response to targeted therapies. In effort to create credible and reproducible models we are implementing the following strategies based on the ten simple rules (TSR) proposed by the model credibility committee:

**Rule 1 – Define context clearly.** We are developing the model in the context of HCC and have avoided the over-generalization of model to “cancer”. We will keep the specifics of the cancer type in all our future publications and presentations along with clear definition of the context.

**Rule 2 – Use appropriate data.** At the molecular and cellular scales, features are inferred from bioinformatic analyses and experimental data. At the tumor and organ scales, models are calibrated to *in vivo* mouse models of HCC and clinical data. The majority of the needed experimental data is going to be generated by the *in vitro* and *in vivo* experiments in the co-PI labs (Ewald and Tran labs). The Fertig lab will perform appropriate preprocessing of all high dimensional data to remove technical artifacts and batch effects prior to analysis.

**Rule 3 – Evaluate within context.** The model will be verified and validated, followed by uncertainty quantification, and sensitivity analysis within the application and context of the HCC. Our codes are developed and managed by multiple researchers in and across the computational labs involved in the project to ensure that they are error free. We are also designing the experiments with model validation in mind to reserve experiments or scenarios that could be used for model validation instead of calibration.

**Rule 4 – List limitation explicitly.** We plan to clearly communicate model limitations and context with users and readers during dissemination. As an example, we have and will continue to list the model limitations explicitly in the Discussion section of our papers.

**Rule 5 – Use version control.** Code sharing version control is done through local Git servers (run on a local Synology Network-Attached Storage server) as well as cloud solutions such as Github that also facilitates dissemination of the models in the public domain (<https://github.com/HCCMultiscale>).

**Rule 6 – Document adequately.** Our codes will be well-documented and will be accompanied by detailed explanation of the elements of the model as well as a user guide on how to run the provided models. Based on our experience with staff transitions during the development of large models, we have established good practices in the lab to maintain a detailed documentation through in-code comments and accompanying user guides.

**Rule 7 – Disseminate broadly.** We have and will publicly provide the model along with the publication of the results. Additionally, we will provide the codes on our lab website, Github, and through MSM wiki platforms.

**Rule 8 – Get independent reviews.** In the first year of the project models most of the revisions will be done internally by different members of the lab and local groups. We will establish connection with MSM member(s) in the March 2020 meeting to perform independent verification of our codes as they develop.

**Rule 9 – Test competing implementation.** We have and will utilize development of competing implementations to explain if and why a specific mechanism is the driver of the process of interest.

**Rule 10 – Conform to standard.** We have and will export our system biology models in SBML format for ease of use by groups using a variety of tools. All bioinformatics methods developed as part of this project will be distributed as R/Bioconductor packages.

We have published a model of HGF/c-Met signaling pathway in the context of liver and HCC and included a segment on model credibility in the Discussion section to explicitly address the TSRs. The model is available through SBML supplementat from the journal website. Further, we have submitted a tutorial manuscript detailing our multiscale modeling efforts and have made our code available on GitHub.

We have also performed bioinformatics analyses of gene expression and proteomics data of *in vivo* data from the Tran lab. Code for these analyses is available on our group github site (<https://github.com/HCCMultiscale>) and has been debugged internally by multiple bioinformaticians in the group. We will prepare a publication when appropriate biological validation of these results has been completed, and the data and reports generated with this code will be released to the public domain at that time.