Gaelle Diserens, Gregory Vuillaume, Thomas Mueller, Marja Talikka, Yiming Cheng, Julia Hoeng
Philip Morris International R&D, Philip Morris Products S.A., Neuchâtel, Switzerland
e-mail: Gaelle.Diserens@contracted.pmi.com, gregory.vuillaume@pmintl.com
Philip Morris International R&D, Philip Morris Research Laboratories GmbH, Cologne, Germany
Frank Tobin
Tobin Consulting LLC, Newtown Square, Pennsylvania, US

Modeling Early Initiation Processes in Smoking-Induced Lung Adenocarcinomas

While most cancer models focus on the development of the tumor itself, our objective is to build a mathematical model of the early initiation processes of the development of lung adenocarcinomas induced by smoking. Our goal is to produce a model that is accurate enough to account for the major phenomenology involved in these initiation processes, that is able to reproduce all the experimental data, and that can explain the timings of tumorigenesis based on demographic differences.

We have approached the model building in four steps. First, the poorly understood biology was triaged to identify the key biological behaviors causing the phenotype transition from normal cells (prior to any smoke exposure) to the earliest phenotype that could be considered a neoplasm. Second, the biology was translated into a nonlinear ODE model that can reasonably explain the effects of smoking and that is neither too complex nor too simplistic. The resulting rate equations for the phenotype dynamics contain first and second order terms. The model is augmented with constraint functions that have a dual role they can be used for checking that the simulation results obey the modeling assumptions and they can be used in the optimization step to insure more reasonable parameters.

The third modeling step consists of the acquisition and analysis of quantitative biological data to calibrate the model. Because the amount of quantitative data within the scope of the model is limited, we have adopted a rigorous surrogate strategy. This allows us to use both clinical and animal data (including omics). The use of animal data requires care to make sure that both the dose and the age of the animals can be properly incorporated into a human model that extends across an entire adult lifespan. Finally, a strategy of constrained optimization is used to obtain a single set of model parameters that simultaneously provides a good fit to all the experimental data sets and accurately reproduces the key biological phenomena, without producing any unacceptable ones.

The model is currently being built and so far contains approximately 20 differential equations involving 50 parameters. We will discuss the model building process, some of the associated mathematical and computational challenges, the need for good data collection practices, and the value of a formal mathematical language for the expression of complex biological knowledge.