MODELING EARLY INITIATION PROCESSES IN SMOKING-INDUCED LUNG ADENOCARCINOMAS

Grégory Vuillaume[1], Thomas Mueller[2], Marja Talikka[1], Yiming Cheng, Gaëlle Diserens[1], Manuel C. Peitsch[1], Julia Hoeng[1], and Frank Tobin[3]

[1]Philip Morris International R&D, Philip Morris Products S.A., Neuchâtel, Switzerland, gregory.vuillaume@pmi.com
[2]Philip Morris International R&D, Philip Morris Research Laboratories GmbH, Cologne, Germany
[3]Tobin Consulting LLC, Newton Square, Pennsylvania, US

ABSTRACT

Most cancer models focus on the growth and progression of a tumor once a neoplasm has first formed. This work focuses on building a mathematical model of the early initiation processes for the earliest neoplasm for lung adenocarcinomas induced by smoking.

The biology was triaged to identify the key biological behaviors causing the phenotype transition from never-smoker lung cells to the earliest phenotype considered a neoplasm. The biology is then translated into a set of nonlinear ordinary differential equations (ODEs). Finally, 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 reproduce the key biological phenomena, without producing any unacceptable one.

1. INTRODUCTION

Smoking induces a variety of complex molecular, histological, and physiological changes in the lungs, eventually leading to adenocarcinoma formation in a subset of smokers. A better understanding of these complex interactions is needed. Mathematical modeling is a powerful approach to developing a single integrated hypothesis that brings together diverse biological phenomena with high-throughput systems biology data and biomarkers measured in *in vitro, in vivo*, and in clinical studies.

The objective is to build a mathematical model of the early initiation processes of the development of lung adenocarcinomas. While most cancer models focus on the growth, promotion, and progression of the tumor, this model ends with the formation of the earliest neoplasm. Because the model acts as a central integrative hypothesis, it must be accurate enough to account for the major phenomenology involved in these initiation processes, be able to reproduce all the experimental data, and explain the timings of tumorigenesis based on smoking consumption patterns and demographic differences.

2. MODEL BUILDING

The process of building the model consists of four main steps: description of the main biological processes involved in disease development, translation of these processes into the corresponding mathematical equations, acquisition of quantitative data, and calibration of the mathematical model to *best* match all the available data and biological knowledge.

2.1. Biological Conceptual Model

The biological model that provides the framework for the mathematical model is composed of several inducible tissue states (phenotypes) that are affected by smoking, demographic factors, and two properties that represent the state of the entire lung as an organ (Figure 1).

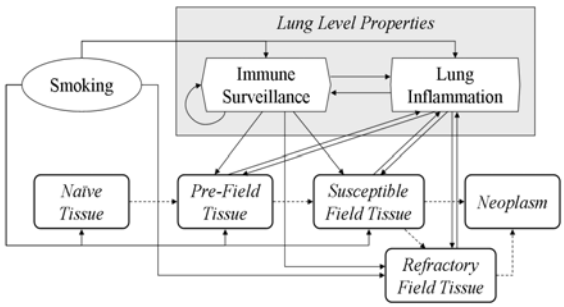

Figure 1. Tissue states and organ properties during active smoking

The effects of smoking are both direct and indirect and impact the entire lung as well as its constituent tissues. From the standpoint of tumorigenesis, the most direct impacts of smoking are an increase in lung inflammation and a decrease in immune surveillance capabilities [1]. In this model, lung inflammation and immune surveillance are considered aggregate organ-level concepts, each incorporating a multitude of constituent processes. This aggregation was done specifically to limit the model complexity and the ensuing need for quantitative data that would arise during the calibration step. Lung inflammation represents a pro-progression concept and immune surveillance represents an anti-progression concept.

When smoking is initiated, naïve, healthy lung tissue is exposed to smoke and is rapidly converted to pre-field tissue, a stressed tissue state which, with ongoing smoking, represents an inflammatory phenotype of the lung.

The dynamics in pre-field tissue imply that tissue homeostasis covering defense and repair mechanisms is sufficiently maintained to combat smoke-dependent lesions, [2], but becomes gradually compromised as pre-field tissue progresses from an early stage to a late-stage pre-field phenotype.

With prolonged, chronic smoking, late-stage pre-field phenotype undergoes a tissue state transition, and early susceptible field tissue begins to form. This tissue state has the phenotype of a chronically inflamed tissue, harbors critical genomic, genetic, and epigenetic alterations, and becomes the foundation for a favorable environment for malignant transformation [3]. With continued smoking, and when sufficiently damaged, a subset of *susceptible* field tissue (susceptible to reversion/healing) further evolves to a *refractory* field tissue state (refractory to reversion/healing), which is considered a pre-neoplastic state, e.g., as represented by atypical adenomatous hyperplasia (AAH). While all other tissue types are revertible during cessation, refractory tissue is not. The refractory field cells guarantee that the model outcomes reflect the elevated lung cancer risk for a number of years after smoking has ceased, as evidenced by epidemiological studies [4]. These cells have an impaired healing capacity and can only be eliminated by immune surveillance function which may remove smoking induced perturbations. Both susceptible and refractory field tissue phenotypes can progress as smoking continues and the model ends when either field type reaches a truly neoplastic state.

The tissue states and their progression staging from early to late phenotypes are represented by constituent properties. These properties vary with each tissue state to properly handle the very different characteristics of each state. As with the organ-level properties, a decision was made to choose the granularity of the properties such that the biology is simplified to the most relevant, most basic characteristics affecting tumor initiation and each *lumped* property consists of an aggregate of many biological phenomena. These tissue properties represent the ability of the cell to maintain its integrity and resist both progression as well as the accumulated damages of smoking. As with the organ-level lung inflammation and immune surveillance effects on tumorigenesis, each tissue type has two countervailing attributes – the ability to sustain 'normalcy' and the damage due to smoking that promotes progression towards malignancy. These tissue properties are impacted by smoking, lung properties, and each other. Table 1 shows the properties needed to distinguish the seminal properties of each tissue type. Figure 2 shows all the pre-field properties and related influences during an active smoking situation. Similar influence diagrams exist for each tissue type, in both smoking and cessation situations.

Table 1. Model tissue properties

	Naïve	*Pre-field*	*Field susceptible*	*Field refractory*
Progression inhibiting		Yes	Yes	Yes
Progression promoting		Yes	Yes	Yes
Growth			Yes	Yes
Death			Yes	Yes
Comments	Stable Homeostatic	Stressed Homeostatic	Damaged Net growth	Highly damaged Net growth

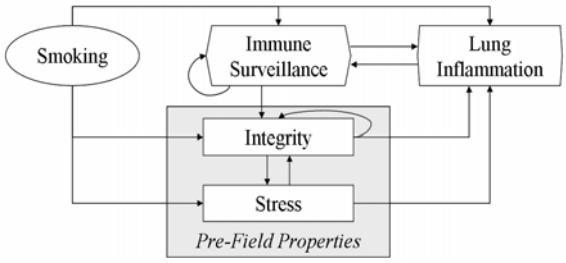

Figure 2. Pre-field properties and influences during active smoking

The model treats the lung and the target tissues as spatially uniform and, within each tissue, focuses on the most advanced tissue phenotypes that can develop into a neoplasm, as it is the initiation timing that is of most interest. Different tissue states can coexist, and the tissue state transitions are dictated by the staging of each state. Each tissue is considered to progress from an early to late staging phenotype with the staging determined by the evolution of the properties that define that tissue. Once each property has reached its end state value, the tissue is considered in its late stage and, only then, may it transition to the next tissue state.

The model can also handle smoking cessation. The conceptual foundation of Figures 1 and 2 remains, but the smoking perturbation ceases. As long as a neoplasm has not yet occurred, the anti-progression elements of the model (immune surveillance at the organ level, cellular integrity/tumor suppression properties) will eventually start to reverse the overall progression of the lung. Cessation itself is not an instantaneous process, as it may take up to a year for smoking products to clear the lung [5]. However, field refractory tissue is not revertible and there is a period of time when it can continue to progress. The balance between the immune system's ability to kill off such cells versus their progression 'momentum' will determine if they become neoplastic during cessation. The dynamics of the lung are such that it takes approximately 1-2 years for the anti-tumorigenic elements of the lung to start reversing the initiation process. After approximately 20 years, the lung will return to a quasi-normal state, but with *scars* of smoking still present [6].

2.2. Mathematical Model

This section describes the steps required to translate the biological model into a mathematical model. X represents any given model variable and, with the exception

of rates (growth/death, tissue transition), variables are restricted to the [0,1] interval.

2.2.1. Handling Smoking

The model distinguishes between the consumption of smoking products, primarily cigarettes, from the actual biological effects of smoking. Smoke is a complex mixture of gases, particulates, and thousands of compounds [1], and would be very hard to model in detail except as an aggregate concept. In addition, as discussed above with cessation, the lungs have a "memory" of smoking since tar, metabolites, and particulates do not clear immediately [5]. To handle all these situations, we do not directly use the time-varying dosage of the smoking product, *S(t)*, as the input variable in the model, but a function, the smoking potency *SP(t)*, which handles the memory of the lung and the clearance phenomena:

$$SP(t) = \int_{-\tau}^{0} S(t+s)\phi(S(t+s),s)\,ds \tag{1}$$

where τ is the clearance period and ϕ is a normalized memory function. For our initial computational experiments, we have been using τ = 90 days. *SP(t)* also has the advantage of converting all smoking products into a standardized form (by modifying ϕ), so that the same model can handle different smoking products – cigarettes, filtered or not, cigars, etc.

2.2.2. Differential Equations

Each biological concept is considered a variable evolving through an associated ODE. The tissues are also described by the number of cells in that tissue as a way to handle the phenotype transitions and the proliferative growth that may occur in the field tissues. These cell counts are also represented as ODEs.

Each property ODE right hand side contains a term for each of the influences associated with the model concept of that variable as stipulated in the biological influence described in the previous section:

$$X_j'(t) = \sum_{k=1}^{M_j} s_{jk}\alpha_{jk}X_{jk}(t) \tag{2}$$

where M_j is the number of influences, X_{jk}, on X_j, with strengths α_{jk} and signs

$$s_{jk} = \begin{cases} 1 & \text{if } X_{jk} \text{ is a positive influence on } X_j \\ -1 & \text{if } X_{jk} \text{ is a negative influence on } X_j \end{cases} \tag{3}$$

The parameters, α_{jk}, are strictly positive (through the use of the s_{jk}'s) and will be determined by optimization. Finally, note that when the influence X_{jk} represents a rate (growth/death/transition), the term $X_{jk}(t)$ in equation (2) is multiplied by $N_j(t)$, where $N_j(t)$ is the number of cells to which the rate is applied.

The initial conditions of the ODEs are defined by a non-smoker at time = 0, so that the immune surveillance is fully functioning at the beginning of the simulation.

For each tissue type, we introduce the concept of staging – a measure of the progression that goes from 0 (early state of the phenotype) to 1 (the most progressed version of that phenotype possible). This can be done for the tissue overall as well as for each property defined for that tissue. Property staging is defined as:

$$Stage\left(X(t), X_e, X_l\right) = \frac{X(t) - X_e}{X_l - X_e}. \tag{4}$$

where X_e is the least progressed value of property X in the tissue and X_l is the corresponding most progressed value. The overall tissue stage is then defined as a weighted average of the staging of its properties and will vary with each tissue type. For instance, pre-field tissue has two properties, stress (*PFS*) and integrity (*PFI*), and thus the pre-field staging is defined by:

$$\begin{aligned} PF\_Stage(t) = {} & \omega_1 Stage\left(PFS(t), 1, 0\right) \\ & + \omega_2 Stage\left(PFI(t), 0, 1\right) \end{aligned} \tag{5}$$

where ω_1, ω_2 are normalized weights for the biological importance of the properties to the progression (uniform weights $\omega_1 = \omega_2 = 0.5$).

While the ODEs appear to be linear, they are not, because of several subtleties. First, most of the tissue and organ properties are defined for finite ranges, usually [0,1], so saturation effects and boundary limits modify the ODE solution behaviors. Second, tissue cell count changes involve second order terms. Namely, they occur from growth/death effects and from transitions into or out of the tissue phenotype, when that or the previous tissue has reached its terminal staging:

$$\begin{aligned} N_j(t)' = {} & \left\{G_j(t) - D_j(t)\right\} N_j(t) \\ & + T_{j-1} N_{j-1}(t)\left[Staging_{j-1}(t) == 1\right] \\ & - T_j N_j(t)\left[Staging_j(t) == 1\right] \end{aligned} \tag{6}$$

where $N_j(t)$ is the number of cells in tissue state j, $G_j(t)$ and $D_j(t)$ are growth and death rates, T_j is the transition rate from phenotype j to $j+1$, and $Staging_j$ is the assessment of the overall progressive state (the staging) of tissue type j. At the end of this process, the model contains approximately 20 ODEs involving about 50 parameters.

2.2.3. Neoplastic Test

When either field-susceptible or field-refractory tissues have progressed as much as they can, i.e., their staging is 1, then a neoplasm is considered to have formed. This follows from the biological definition of field tissue, which states that the most progressed form of that tissue is really the earliest neoplastic formation.

2.2.4. Constraints

In addition to setting the basis for the ODEs, we use the biological knowledge to identify key biological phenomena that the model should reproduce, as well as unacceptable behaviors that the model should avoid. These are translated into formal mathematical constraints that will be used during the model calibration process (see

section 2.3). As simple examples, during active smoking periods, we expect lung inflammation (*LI*) to increase, immune surveillance (*IS*) to decrease, and the susceptible field tissue to be proliferating. These behaviors translate to the following constraints:

$$LI'(t) > 0, IS'(t) < 0, FSG(t)/FSD(t) > 1 \text{ if } SP(t) > 0 \quad (7)$$

where *FSG(t)* and *FSD(t)* represent the growth and death rates of susceptible field cells, respectively.

2.3. Model Calibration

Model calibration is done by using constrained optimization to simultaneously find a single set of model parameters (the α's) that provides a *"good"* fit to *all* the experimental data sets and accurately reproduces the key biological phenomena, without producing any unacceptable behaviors ('that satisfy all the mathematical constraints such as (7)'). More specifically, we solve:

$$\begin{aligned} &O(\Lambda) = \min_{\Lambda} O(\Lambda) \\ &g(\Lambda) = 0, \quad h(\Lambda) > 0 \\ &O(\Lambda) = \sum_{j}^{\#expts} \omega_j O_j\left(M_j\left(\vec{X}, t, \Lambda\right), D_j\right) \end{aligned} \quad (8)$$

where the objective function is a weighted sum over the distance (not necessarily an L_2 metric) between all the experimentally collected data and the model outputs corresponding to each of those experiments. Note, we have indicated that the model might vary with each data set to appropriately correspond to each experiment, yet there is still one global parameter set shared by each of these variants. The weights are determined by the "trust" in each experiment and to keep the various terms approximately equal to prevent any experiment from unduly dominating the objective function.

We note that there is an implicit dependence upon the data sets, D_j, in the objective function. Adding or removing data may change the resulting parameters and the subsequent model behaviors. This is also true for the qualitative knowledge encapsulated in the constraints. It is therefore critical that only high-quality experimental data be used. This naturally leads to the question of what data are acceptable for the purposes of building the model. And that, in turn, comes back to the assumption used in building the model. These essentially define a "scope" for the data that cannot be violated. Several key criteria for the data must be met: they must come from lung tissue related to the initiation of adenocarcinomas and they must be from the initiation processes, not from tumor-related processes. Because many of the model variables are aggregate concepts, it is necessary to have surrogates for them. These criteria are difficult to interpret in practice due to subtleties and the lack of adequate description of experimental protocols.

The resulting parameter set provides the *bes*t fit between the model outputs and the available data. This implies trying to find the global minimum as best as can be done with all the caveats of noisy and possibly inconsistent data. The constraints become critically important to limit the parameters to a feasible set of 'good enough'. Given that the model encompasses the available biological knowledge, the resulting calibrated model can be considered a central mechanism hypothesis for lung adenocarcinoma initiation induced by smoking.

3. DISCUSSION

Early initiation processes in smoking-induced lung adenocarcinomas are complex and involve multiple processes and positive and negative feedback loop pairs. At the lung level, the lung inflammation and immune surveillance properties have this role. In pre-field tissue, and the two field tissues, the cellular integrity/tumor suppression and the cellular damage/tumor promoter properties act in this capacity. It is necessary to have both the progression-advancing and progression-reverting properties to handle both the smoking-induced tumorigenesis and the cessation consequences of reversion of tumorigenesis. As a consequence of this balance of positive and negative effects, interpreting single experimental results in the context of disease initiation becomes very difficult, as experimental data usually focus on a quite limited aspect of the disease development. In addition, the effective rates of change of these different processes may well be different.

Of particular interest are simulations of different smoking patterns (age of starting to smoke, daily dose, potential age of smoking cessation, etc.) to the time to first neoplasm formation. With this model, it should be possible to understand more precisely the impact of various smoking scenarios on lung adenocarcinoma initiation. In addition, the model provides a mechanistic explanation for different smoking or cessation scenarios.

4. REFERENCES

[1] M. R. Stampfli and G. P. Anderson, "How cigarette smoke skews immune responses to promote infection, lung disease and cancer," *Nat Rev Immunol,* vol. 9, pp. 377-84, May 2009.

[2] K. Aoshiba and A. Nagai, "Oxidative stress, cell death, and other damage to alveolar epithelial cells induced by cigarette smoke," *Tob Induc Dis,* vol. 1, pp. 219-26, 2003.

[3] A. J. Schetter, N. H. Heegaard, and C. C. Harris, "Inflammation and cancer: interweaving microRNA, free radical, cytokine and p53 pathways," *Carcinogenesis,* vol. 31, pp. 37-49, Jan 2010.

[4] E. Benhamou, S. Benhamou, A. Auquier, and R. Flamant, "Changes in patterns of cigarette smoking and lung cancer risk: results of a case-control study," *Br J Cancer,* vol. 60, pp. 601-4, Oct 1989.

[5] S. S. Hecht, S. G. Carmella, M. Chen, J. F. Dor Koch, A. T. Miller, S. E. Murphy, J. A. Jensen, C. L. Zimmerman, and D. K. Hatsukami, "Quantitation of urinary metabolites of a tobacco-specific lung carcinogen after smoking cessation," *Cancer Res,* vol. 59, pp. 590-6, Feb 1 1999.

[6] J. Beane, P. Sebastiani, G. Liu, J. S. Brody, M. E. Lenburg, and A. Spira, "Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression," *Genome Biol,* vol. 8, p. R201, 2007.