Spreaders and Sponges Define Metastasis in Lung Cancer: A Markov Chain Monte Carlo Mathematical Model

Paul K. Newton, Jeremy Mason, Kelly Bethel, Lyudmila Bazhenova, Jorge Nieva, Larry Norton, and Peter Kuhn

Abstract

The classic view of metastatic cancer progression is that it is a unidirectional process initiated at the primary tumor site, progressing to variably distant metastatic sites in a fairly predictable, although not perfectly understood, fashion. A Markov chain Monte Carlo mathematical approach can determine a pathway diagram that classifies metastatic tumors as "spreaders" or "sponges" and orders the timescales of progression from site to site. In light of recent experimental evidence highlighting the potential significance of self-seeding of primary tumors, we use a Markov chain Monte Carlo (MCMC) approach, based on large autopsy data sets, to quantify the stochastic, systemic, and often multidirectional aspects of cancer progression. We quantify three types of multidirectional mechanisms of progression: (i) self-seeding of the primary tumor, (ii) reseeding of the primary tumor from a metastatic site (primary reseeding), and (iii) reseeding of metastatic tumors (metastasis reseeding). The model shows that the combined characteristics of the primary and the first metastatic site to which it spreads largely determine the future pathways and timescales of systemic disease. Cancer Res; 73(9); 2760–9. © 2013 AACR.

Introduction

The classic view of metastatic progression, framed in part by the "seed-and-soil" hypothesis of Paget (1), is that cancer spreads from the primary tumor site to distant metastatic locations in a unidirectional way. The "seeds" responsible for the spread are circulating tumor cells (CTC; refs. 2–4) that detach from the primary tumor, enter the bloodstream and lymphatic system (3), and travel to new distant locations. If conditions are favorable, this initiates a complex (5–7) and not well understood metastatic cascade, ultimately leading to tumor growth at distant anatomic sites if their "soil" is hospitable (1). The exclusively unidirectional nature of this process has been challenged recently in a series of articles (8–13), which use mouse models to show a mechanism by which CTCs from the primary tumor can re-enter the primary, a process called "self-seeding" (12). These authors further comment that "it is tempting to speculate that self-seeding might occur not only at the primary tumor site, but also at distinct metastatic sites, each site being a nesting ground." The possibility of metastasis from metastases has also been discussed (14, 15). While the underlying "agent" responsible for the spread of cancer is the CTC, the disease progression pathways in different patients can be both predictable (from a statistical viewpoint), but often unpredictable and surprisingly distinct in patients with nominally the same disease (16, 17), prompting the question "how can metastatic pathways be predictable and unpredictable at the same time" (10)?

Motivated in part by these questions, we develop a Markov chain/Monte Carlo (MCMC) stochastic mathematical model for cancer progression to identify and quantify the multidirectional pathways and timescales associated with metastatic spread for primary lung cancer.

While stochastic in nature, our model shows that a defining aspect of both pathway selection and timescale determination is whether the disease spreads from the primary tumor to a metastatic site that is either a "spreader" (adrenal gland and kidney) or a "sponge" (regional lymph nodes, liver, bone). In contrast to the traditional view of cancer metastasis as a unidirectional process starting at the primary site and spreading to distant sites as time progresses, our model supports and quantifies the view that there are important multidirectional aspects to metastatic progression. These fall under 3 general...
Quick Guide to Equations and Assumptions

**Assumptions of the model:**
1. The disease progression starts with an isolated tumor in the lung position.
2. The progression dynamics follow a Markov stochastic process (19), moving from site \(i\) to site \(j\) according to a transition probability \(P_{ij}\) that depends only on those 2 sites, not on the past history of how it arrived at site \(i\).
3. The Markov transition matrix of our model is calculated so that the steady-state vector of the Markov chain model corresponds to the metastatic distribution of tumors found from the autopsy data set described (20).

**Key equations:**
1. A Markov chain dynamical system (18, 19) is defined by the equation:
   \[
   \tilde{v}_{k+1} = \tilde{v}_k A \quad (k = 0, 1, 2, \ldots),
   \]
   where \(A\) is called the Markov transition matrix and \(\tilde{v}_k\) is called the state vector. The entries of the state vector give the probabilities of metastatic tumors developing at the 50 different sites in our model. The sum of the entries must be 1. The entries of the Markov transition matrix are the transition probabilities \(P_{ij}\) from site \(i\) to site \(j\). In our model, \(A\) is a \(50 \times 50\) square matrix and \(\tilde{v}_k\) is a vector in \(\mathbb{R}^{50}\).
2. \(\tilde{v}_0 = (0, 0, \ldots, 1, 0, 0, \ldots)\) is our initial state vector, which has a 1 placed in the 23rd position, corresponding to the "lung" entry.
3. \(\tilde{v}_\infty = \lim_{k \to \infty} \tilde{v}_0 A^k\) is called the steady-state vector associated with the Markov chain. It can be obtained by solving the eigenvalue problem: \(\tilde{v}_\infty (A - I) = 0\). Therefore, the steady-state vector is a left eigenvector of the Markov transition matrix.

classes: (i) self-seeding of the primary tumor, (ii) reseeding of the primary tumor from a metastatic site (primary reseeding), and (iii) reseeding of metastatic tumors (metastasis reseeding).

Using a discrete Markov chain (19) system of equations applied to a large autopsy dataset of untreated patients with cancer (20), we quantify the likelihoods of the top metastatic pathways in terms of probabilities and conduct Monte Carlo computer simulations of cancer progression that statistically reflect the autopsy data about (non-Gaussian) distribution of disease. The stochastic Markov chain dynamical system takes place on a metastatic network–based model of disease progression that we construct based on available autopsy data over large populations of patients. To obtain our baseline model, we use the data described in an autopsy analysis (20) in which metastatic tumor distributions in a population of 3827 untreated deceased cancer victims were recorded: 163 of these had primary lung cancer of some type, distributing a total of 619 metastatic tumors across 27 different sites. Information on lung cancer type in this data set is not possible to obtain as the samples were collected before the widespread use of immunohistochemistry (1914–1943), without which, the subcategorization of non–small cell lung cancer is unreliable. However, it is probably safe to assume that the distribution of lung cancer type was not significantly different than current distributions, roughly 40% adenocarcinoma, 30% squamous cell carcinoma, 9% large cell carcinoma, and 21% small cell carcinoma.

**Materials and Methods**

**Structure of the lung cancer multistep diagram**

The 27 metastatic sites in the diagram shown in Fig. 1 are organized in ring formation, with 20 sites surrounding lung on the inner ring and the remaining 7 sites organized on the outermost ring, each connected to a site from the inner ring. The sites listed on the inner ring are called “first-order” sites—they have direct edge connections from the lung, with edge probabilities decreasing from 1200 clockwise around the ring. The most heavily weighted edge, hence the most likely first step of metastatic disease, is the transition from lung to regional lymph nodes [LN (reg)]. The least heavily weighted, hence least likely first step, is the transition from lung to skeletal muscle shown just to its left. The 7 sites making up the outermost ring are called “second-order” sites, also organized with edge probabilities decreasing in clockwise order. These sites are classified as "second-order" due to the fact that they have 2-step probabilities via a first-order site that are equal or higher in probability than any direct one-step probability from the lung. In short, for disease to spread to a second-order site from lung, it most probably passes via a first-order site.

The general structure of the concentric diagram, with lung placed at the center, highlights the underlying classical unidirectional view of disease progression. However, the diagram also highlights the 3 key mechanisms of multidirectional progression: (i) self-seeding of the primary tumor shown in the diagram as a self-loop in the seventh position, with an edge weight of 5.2% and (ii) reseeding of the primary tumor from a first-order site, shown as arrows directed back to the center. Because we are using an ensemble average of 1,000 trained lung cancer matrices to produce this diagram, the reseeding edges are all roughly comparable in weight (8%), (iii) metastasis reseeding of first-order sites shown as a self-loop back to each metastatic site. The strongest metastasis reseeders are lymph nodes (regional and distant), followed by liver, adrenal, bone, and kidney.
From this diagram, we can obtain the 2-step pathway probabilities from the lung, by direct multiplication of the 2 edges making up any of the 2-step paths starting from lung. The 729 distinct two-step paths from the lung, the top ones of which are shown in Fig. 2, produce the statistical distribution $\tilde{v}_2$ produced by the Markov chain model. We calculate all of these and rank them in decreasing order in the next subsections. By comparing the probability distributions $\tilde{v}_2$ and $\tilde{v}_\infty$ (shown in Supplementary Fig. S2), we can see that after 2 steps, the distribution has nearly converged to the steady-state, so we expect our rankings of 2-step pathways not to change much if we compare them to the top 3-step and higher step pathways.

Supplementary Figure S1 shows a (ensemble) convergence and nonconvergence plot associated with our search algorithm to calculate the Markov transition matrix based on the baseline
dataset (20). What is significant is the nonconvergence of our algorithm when we constrain our searches to not allow for any multidirectional edges. In other words, when we forced our algorithm to not allow edges directly back to a site (no self-metastases nor primary reseeding), either separately or together, the algorithm would not converge to a solution. In contrast, the algorithm, in general, converged quickly to a solution when all connections were allowed and produces a transition matrix with many multidirectional connections from site to site.

The autopsy datasets

The data in (20) compiles the metastatic tumor distributions in a population of 3,827 deceased cancer victims, none of whom received chemotherapy or radiation, hence the model can be said to be based on the natural progression of the disease, although mastectomy for many breast cancer primaries was most likely conducted at that time. In addition, brain metastases are likely underrepresented by this dataset as brain autopsies probably were not universally conducted at that time. The autopsies were conducted between 1914–1943 in 5 separate affiliated centers, with an ensemble distribution of 41 primary tumor types and 30 distinct metastatic locations. The total number of distinct primary and metastatic tumor locations is 50, which sets the size of our square Markov transition matrix (50 × 50) as well as the number of entries in the Markov state-vector \( v \). The data offer no direct information on the time history of the disease, either for individual patients comprising the ensemble or in ensemble format. The data we use, therefore, only contain information on the “long-time” distribution of metastatic tumors, where long-time is associated with end of life, a timescale that varies significantly from patient to patient. The model does, however, allow us to infer time histories from autopsy data based on the logic that if more metastatic tumors show up in a population of patients at a specific site, then on average, they would develop earlier in the progression history. Although this association is not perfect, if does allow us to extract meaningful temporal inferences from our Markov chain model. Details of how we infer the correct ensemble Markov transition matrix are described in reference 18.

We use the dataset in 2 distinct ways to construct our model. First, we associate the distribution of metastatic tumors (after appropriate normalization) for primary patients with lung cancer with the steady-state (long-time) probability distribution of our Markov chain (19). From this, we compute the “transition matrix” for our Markov chain (ensemble averaged) that produces this steady state. As the problem is mathematically undetermined, the calculation procedure requires an initial “candidate” transition matrix obtained from the autopsy data and discussed in (18), which is then systematically iterated until a numerical convergence criterion is satisfied. Interestingly, we also show that when our search algorithm is constrained so as to not allow any multidirectional edges in the directed graph associated with the transition matrix, no self-consistent model can be produced (i.e., the search algorithm does not converge). Then, we update our baseline model with the more targeted dataset described (25) of 137 patients with adenocarcinoma of the lung (stage I and II), all treated with complete lung resection, and show how the baseline model is able to adapt to this assimilated data set.

Results

Cancer metastasis as a stochastic multistep process

The ensemble averaged lung cancer transition matrix associated with the Markov chain model (see Fig. 1) depicts the complete metastatic pathway diagram (18). Each of the 2,500 entries, \( a_{ij} \), of the 50 × 50 transition matrix determines the probability of the disease (modeled as a random walker over the network) spreading from site “i” to site “j” in an effectively multistep process before the statistical tumor distribution of the autopsy dataset is filled out. The diagram rank orders (in decreasing clockwise order) all of the possible pathways emanating from the lung. One-step paths are defined by the edges leading directly out from the lung—the sum of these outgoing edges must be one. The single most likely one-step path of disease progression from the lung is to the regional lymph nodes, shown at the top of the diagram, with a probability of 15.1%, followed by the lung to adrenal gland path, with probability of 13.2%. On the diagram ordering the first steps out of the lung, we also show the “self-seeding” step directly back to the lung, represented by the edge from lung looping back to itself, with edge probability 5.2%.

Two-step paths are made up of an edge from the lung to another site (or back to itself), followed by the edge from that site to a second site. There are 729 two-step paths emanating from the lung. The probability of taking a particular two-step path from the lung is obtained by multiplying the weights of the 2 edges making up the path. The sum of all of these 2-step path probabilities must be 1, and so on for 3-step paths, 4-step paths, etc. We focus on quantifying all of the two-step paths in this article, because as shown in Supplementary Fig. S2 (See Supplementary Material), after 2 iterations of the Markov chain (\( k = 2 \)), the state-vector has nearly converged to the steady-state target vector for metastatic tumors making metastatic progression for lung cancer effectively a 2-step process. In Fig. 2, we show all of the 2-step paths emanating from the lung passing through each of the 8 most probable metastatic sites. To obtain the probability of cancer progression on 1 of these 2-step paths, one multiplies the products of the 2 edges making up the 2-step path.

Rank-ordering the two-step metastatic pathways toward the final state of the disease

We list the top multidirectional 2-step pathways obtained from our model in Table 1. The first entries of Table 1 list the top 10 reseeding pathways back to the lung from a first-order site, along with the running cumulative values. We highlight from this list several points. First, lymph nodes, adrenal gland, and liver are the most important intermediate sites that reseed back to the lung. Their cumulative probability value (3.8%) accounts for more than half of the total cumulative value from the entire list (6.2%). This total cumulative value is slightly greater than, but roughly comparable in size to the lung to lung reseeding path value of 5.2%, indicating that cells that reseed to the lung land there.
with roughly equal probabilities of having arrived via an intermediate site (see Table 1) versus directly from the lung. The second half of Table 1 lists the top 10 2-step reseeding pathways back to a metastatic site, a mechanism we call "metastasis reseeding." From this table, we can see that for lung cancer, lymph nodes and adrenal gland are the most active metastasis reseeders, followed by liver, bone, and kidney.

Metastatic sites as spreaders or sponges
A careful analysis of the top 30 2-step pathways allows us to compute the key probabilistic quantity of interest associated with each 2-step path which characterizes each site as a sponge or a spreader. The quantity is the ratio of probability out ($P_{out}$) over probability in ($P_{in}$) to each of the sites. If $P_{out} > P_{in}$, the site is a spreader, whereas if $P_{in} > P_{out}$, we characterize it as a sponge. The ratio ($P_{out}/P_{in}$) of their
exiting and incoming probabilities, in the case of a spreader, gives us what we call the amplification factor, as it is larger than one, whereas in the case of a sponge, we call the ratio the absorption factor, as it is less than 1. Using these quantities, the top 2 spreaders are the adrenal gland and kidney, with amplification factors of 1.91 (adrenal gland) and 2.86 (kidney). The total number of 2-step pathways into and out of the adrenal gland is 10, whereas the total into and out of kidney is only 3. For these reasons, we identify the adrenal gland as the key distant anatomic spreader of primary lung cancer.

The sponges associated with primary lung cancer are the regional lymph nodes, liver, and bone. Their respective absorption factors are 0.74 (regional lymph nodes), 0.87 (liver), and 0.75 (bone). The total number of 2-step pathways into and out of the regional lymph nodes is 16, compared with 8 into and out of the liver, and 5 into and out of bone. For these reasons, we identify the regional lymph nodes as the key anatomical sponge associated with primary lung cancer, followed by both bone and liver.

**The spatial pathways of lung cancer**

To compare the relative importance of 2-step unidirectional pathways versus 2-step multidirectional pathways, we list the top 30 two-step pathways in decreasing order in Supplementary Table S1. The top metastatic pathway (of any type) is the lung → lymph node (reg) → lymph node (reg) metastasis reseeding pathway, whereas the top unidirectional pathway is the lung → adrenal → lymph node (reg) path. Looking at all of the multidirectional pathways, it is clear that the lymph nodes and adrenal gland are the key metastatic sites responsible for multidirectional spread, whereas lymph nodes, adrenal gland, and liver are important sites responsible for unidirectional spread. In general terms, lymph nodes, adrenal gland, and liver feature very prominently as intermediate metastatic sites in many of the 2-step pathways.
The information can then be combined into a reduced 2-step diagram for progression, shown in Fig. 3. The diagram shows the centrality of lymph nodes and adrenal gland as key first metastatic sites, with many incoming and outgoing edges. The figure also captures all of the information about the spreader or sponge character of each site, with red indicating the color of the key spreaders (adrenal gland, kidney) and blue indicating the color of sponges (lung, regional lymph nodes, liver, bone). Amplification and absorption factors are shown in each of the ovals.

### Timescales of progression: enhancing the Kaplan–Meier approach

Our model gives a useful measure of metastatic progression timescale, called first-passage time from lung to any given site, defined as the number of edges a "random walker" leaving the lung must traverse to first arrive at that site. Monte Carlo simulations of random walk paths from the lung are conducted computationally to obtain mean first-passage times (averages over 10,000 runs) to every other site in the model. The mean first-passage times (mfpts) act as a proxy timescale (model-based) for metastatic progression. It is a model-based relative measure of the time that it takes for a primary tumor to metastasize to a secondary site, or, roughly speaking, a model-based measure of the timescale associated with successful extravasation and colonization (6). Timescales associated with metastatic disease are typically quantified by so-called Kaplan–Meier survival curves (23, 24), which follow a cohort of patients from presentation until death, plotting the survival percentage associated with the cohort. Alternative methods have been proposed, but by and large, tracking survival of a cohort of patients remains the industry-standard way of tracking progression. There is very little in the literature that tracks the timescale of progression from metastatic site to metastatic site (15–17, 21, 22).

Mean first-passage times from lung to each of the other sites are shown in Fig. 4. The sites are ordered from shortest to longest mean first-passage time from lung. In dark, we show the baseline (untreated patients) model using the data set (20). The dashed-dot line is a linear curve fit to the first 9 sites, showing a clear linear increasing regimen (roughly the top 16 sites), followed by a group of sites where mean first-passage times increase nonlinearly. The first 9 sites used in the reduced model set the basic linear timescales of progression for the high probability metastatic locations. Times increase following the general linear formula \( \text{mfpt} = a \cdot t + b \), where \( a = 2.56, b = 2.07 \) for the baseline (untreated) model, where “\( a \)” is the slope and “\( b \)” is the \( y \)-intercept. In this formula, larger slopes indicate longer overall mean first-passage times from lung to metastatic sites. Spread to regional lymph nodes is fastest (with a normalized value of 1), followed by normalized times to distant lymph nodes (1.47), adrenal (1.72), and liver (1.75). One should interpret these timescales to indicate that it takes roughly 75% longer for cancer to metastasize to adrenal gland and liver than to regional lymph nodes. Self- seeding back to lung has a normalized mean first-passage time of 2.30, which is faster.

### Table 1. Top 2-step pathway probabilities

<table>
<thead>
<tr>
<th>Top reseeding pathways back to lung</th>
<th>Transition probability</th>
<th>Cumulative values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung ( \rightarrow ) Lymph (reg) ( \rightarrow ) Lung</td>
<td>0.01214</td>
<td>0.02256</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Adrenal ( \rightarrow ) Lung</td>
<td>0.01042</td>
<td>0.03208</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Lymph (dist) ( \rightarrow ) Lung</td>
<td>0.00952</td>
<td>0.03853</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Liver ( \rightarrow ) Lung</td>
<td>0.00645</td>
<td>0.04386</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Kidney ( \rightarrow ) Lung</td>
<td>0.00533</td>
<td>0.04853</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Bone ( \rightarrow ) Lung</td>
<td>0.00467</td>
<td>0.05228</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Pleura ( \rightarrow ) Lung</td>
<td>0.00375</td>
<td>0.05595</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Pancreas ( \rightarrow ) Lung</td>
<td>0.00367</td>
<td>0.05883</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Heart ( \rightarrow ) Lung</td>
<td>0.00288</td>
<td>0.06156</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Lung ( \rightarrow ) Lung</td>
<td>0.00273</td>
<td>0.06429</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Top metastasis reseeders</th>
<th>Transition probability</th>
<th>Cumulative values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung ( \rightarrow ) Lymph (reg) ( \rightarrow ) Lymph (reg)</td>
<td>0.02819</td>
<td>0.04287</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Lymph (dist) ( \rightarrow ) Lymph (dist)</td>
<td>0.01468</td>
<td>0.05510</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Adrenal ( \rightarrow ) Adrenal</td>
<td>0.01223</td>
<td>0.06268</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Liver ( \rightarrow ) Liver</td>
<td>0.00758</td>
<td>0.06632</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Bone ( \rightarrow ) Bone</td>
<td>0.00364</td>
<td>0.06946</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Kidney ( \rightarrow ) Kidney</td>
<td>0.00314</td>
<td>0.07152</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Pleura ( \rightarrow ) Pleura</td>
<td>0.00206</td>
<td>0.07320</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Pancreas ( \rightarrow ) Pancreas</td>
<td>0.00168</td>
<td>0.07418</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Spleen ( \rightarrow ) Spleen</td>
<td>0.00098</td>
<td>0.07513</td>
</tr>
<tr>
<td>Lung ( \rightarrow ) Heart ( \rightarrow ) Heart</td>
<td>0.00095</td>
<td>0.07513</td>
</tr>
</tbody>
</table>

NOTE: Top 2-step reseeding pathways back to lung: Primary \( \rightarrow \) First-order site \( \rightarrow \) Primary. Top reseeding pathways back to metastatic site: Primary \( \rightarrow \) First-order site \( \rightarrow \) Back to first-order site. Cumulative values (obtained by adding the previous transition probabilities) are listed in third column.
than to most of the first-order sites, but over twice the time as the lung to regional lymph node timescale.

**Assimilating new autopsy data of adenocarcinoma lung cancer patients undergoing complete resection**

Figure 4 (more details are shown in Supplementary Table S1) also shows metastatic pathways and mean first-passage times using the model with assimilated data from (25), an autopsy data set tracking a cohort of patients with adenocarcinoma of the lung (ACL) who underwent complete lung resection. Of these, 35 survived more than 30 days after resection, 22 were classified as stage I, and 13 as stage II. We assimilated their metastatic tumor distribution from an autopsy study into our baseline (untreated population) model, recalculated the Markov transition matrix and all mean first-passage times. The results are shown in Fig. 4 (and the middle and right columns of Supplementary Table S1). Stage I are shown in medium dark, stage II in light gray.

Comparing the columns of Supplementary Table S1, the main change in the spatial pathways shows up in the fifth entry down, where the Lung → Adrenal → LN (Dist) pathway drops in probability on the list of the stage I treated patients but not as much as for the stage II treated patients. Lung resection seems to alter this important pathway, particularly for stage I patients, making it less likely to occur, perhaps by disruption of lymphatic connections between the primary tumor and ipsilateral adrenal gland. The overall probabilities of each of the pathways in the treated population also decrease from the untreated population.

The effect of treatment on the overall mean first-passage times is shown in Fig. 4. The corresponding curve fit to the first 9 sites follow the same general linear trend as in the untreated population, $\text{mfpt} = a \cdot t + b$, but with $a = 2.68, b = 1.55$ (stage I, medium dark); $a = 2.54, b = 1.91$ (stage II, light gray). The conclusions we can draw are clear: mean first-passage times increase overall with the stage I treated cohort, shown by the increase in slope over the untreated slope, but not with the stage II treated cohort. Interestingly, the mfpt back to lung in the treated cohort actually decreases with treatment. As lung is classified as a sponge in our model, this does not seem to have a negative overall effect on the general trend of increasing passage times with treatment. In contrast, the mfpt back to adrenal gland (the key spreader) with the treated cohort increases. This enhances the overall increase in mfpts for the treated cohort. The mean first-passage times increase most in the subgroup of stage I patients, indicating that complete lung resection is more effective in this group compared with the stage II subgroup. To summarize, our model shows that lung resection for patients with ACL seems to generally increase...
overall mfpts of metastases for stage I patients, and it does this by (i) altering a key pathway from lung to adrenal gland to lymph nodes (distal), (ii) increasing mean first-passage times to the adrenal gland (spreader), (iii) decreasing mean first-passage times back to the lung (sponge), and (iv) reducing the overall top pathway probabilities. Lung resection seems to have very little impact on stage II patients. The failure of resection to improve metastasis-free survival in stage II patients with lung cancer may occur because the regional lymph nodes act as a sponge (Fig. 3), potentially suppressing patients with lung cancer may occur because the regional lymph nodes act as a sponge (Fig. 3), potentially suppressing metastasis reseeding when not removed. However, because the risk of local disease is high in lung cancer, surgery remains the preferred treatment in stage II disease.

Discussion

Our model depicts cancer progression as effectively a multistep (2-step), multidirectional, stochastic process, spreading probabilistically from site to site in individual patients, but filling out a well-defined and predictable metastatic tumor distribution for large ensembles of patients. This stable, robust, and predictable ensemble tumor distribution available over large autopsy data sets is exploited to build a Markov transition matrix for lung cancer progression. We identify the top unidirectional and multidirectional metastatic pathways of primary lung cancer by means of a probabilistic comparison of all 2-step paths emanating from the lung. The results support the view that multidirectional pathways play an important role in cancer progression. We identify 3 main mechanisms of multidirectionality needed to obtain consistency with ensemble autopsy data: (i) primary tumor self-seeding, (ii) reseeding of the primary tumor from a metastatic tumor, and (iii) metastasis reseeding. Of these, the most important are metastasis reseeding of the lymph nodes (both regional and distant) and adrenal gland and primary lung reseeding via the regional lymph nodes. Also significant is metastasis reseeding of the liver and primary self-seeding of the lung, but neither seem to be as significant as passage of the disease through the regional lymph nodes.

While very few patients die from their first metastasis, the characterization of the first metastatic site as a spreader or sponge yields important insights into metastatic pathway selection and the determination of progression timescales for patients. The model may have implications for decisions surrounding surgical resection of oligometastatic disease (26) as one might predict different outcomes for patients whose solitary site of disease is a sponge or spreader. Historically, resection of isolated adrenal metastasis has entered clinical practice in lung cancer, and removal of this spreader site has benefited patients (27). Conversely, there has never been an established role for resection of isolated liver metastasis, a sponge site, despite there being a track record of success (28–32).

A careful inspection of the top 2-step pathways supports the dominance of unidirectional metastatic spread over multidirectional processes, which perhaps explains why the prevailing historical view is one of unidirectional spread (5). However, we should emphasize that our search algorithm for a Markov transition matrix could not converge to any solution when we constrained it so that multidirectional edges were ruled out but did converge consistently to an ensemble of transition matrices when unconstrained so that all possible paths were allowed (See Supplementary Material). In other words, we were not able to find a Markov transition matrix that produced a steady-state consistent with the autopsy data unless multidirectional edge connections were allowed. Therefore, we stress the viewpoint that multidirectional processes play a key role in pathway progression.

Figure 4. Mean first-passage times from lung to each of the metastatic sites. Dark gray shows the baseline (untreated population) model, medium gray shows the baseline model with assimilated stage I resections, and light gray shows baseline model with assimilated stage II resections. Lines are linear curve fits to first 9 entries. Error bars show 1 SD from the mean. See text for details.
selection and timescale determination for metastatic lung cancer.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Disclaimer
The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Cancer Institute of the NIH.

Authors' Contributions
Conception and design: P.K. Newton, J. Mason, K. Bethel, L. Bazhenova, L. Norton, P. Kuhn
Development of methodology: P.K. Newton, J. Mason
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. J. Nieva
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): P.K. Newton, J. Mason, K. Bethel, L. Bazhenova, J. J. Nieva, P. Kuhn
Writing, review, and/or revision of the manuscript: P.K. Newton, J. Mason, K. Bethel, L. Bazhenova, J. J. Nieva, L. Norton, P. Kuhn
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P. Kuhn
Study supervision: P.K. Newton, P. Kuhn

Grant Support
The project described was supported by Award Number U54CA143906 from the National Cancer Institute and the Bill and Melinda Gates Foundation through the Gates Millennium Fellowship Program.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 11, 2012; revised January 24, 2013; accepted February 18, 2013; published OnlineFirst February 27, 2013.

References
Spreadsers and Sponges Define Metastasis in Lung Cancer: A Markov Chain Monte Carlo Mathematical Model


*Cancer Res* 2013;73:2760-2769. Published OnlineFirst February 27, 2013.

**Updated version**
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-4488

**Supplementary Material**
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2013/02/27/0008-5472.CAN-12-4488.DC1

**Cited articles**
This article cites 29 articles, 3 of which you can access for free at:
http://cancerres.aacrjournals.org/content/73/9/2760.full.html#ref-list-1

**Citing articles**
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
/content/73/9/2760.full.html#related-urls

**E-mail alerts**
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.