A mathematical model of ctDNA shedding predicts tumor detection size

Stefano Avanzini[1], David M. Kurtz[2], Jacob J. Chabon[3,4], Sharon Seiko Hori[1,5], Sanjiv Sam Gambhir[1,4,5,6,7], Ash A. Alizadeh[2,3,4], Maximilian Diehn[3,4,8], Johannes G. Reiter[1,4,6,9*]

[1] Canary Center for Cancer Early Detection, Department of Radiology, Stanford University School of Medicine, Palo Alto, CA 94304, USA.
[2] Division of Oncology, Division of Hematology, Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
[3] Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
[4] Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305, USA.
[5] Molecular Imaging Program at Stanford, Department of Radiology, Stanford University School of Medicine, Stanford, CA 94305, USA.
[6] Bio-X Program, Stanford University, Stanford, CA 94305, USA.
[7] Department of Bioengineering and Department of Materials Science & Engineering, Stanford University School of Medicine, Palo Alto, CA 94305, USA.
[8] Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA 94305, USA.
[9] Department of Biomedical Data Science, Biophysics Program, Stanford University, Stanford, CA 94305, USA.

Abstract

Early cancer detection aims to find tumors before they progress to an uncurable stage. Prospective studies with tens of thousands of healthy participants are ongoing to determine whether asymptomatic cancers can be accurately detected by analyzing circulating tumor DNA (ctDNA) from blood samples. We developed a stochastic mathematical model of tumor evolution and ctDNA shedding to investigate the potential and the limitations of ctDNA-based cancer early detection tests. We inferred ctDNA shedding rates of early stage lung cancers and calculated that a 15 mL blood sample contains on average only 1.5 genome equivalents of ctDNA for lung tumors with 1 billion cells (size of ~1 cm^3). We considered two clinically different scenarios: cancer screening and cancer relapse detection. For monthly relapse testing with a sequencing panel covering 20 tumor-specific mutations, we found a median detection size of 0.24 cm^3 corresponding to a lead time of 160 days compared to imaging-based relapse detection. For annual screening, we found a median detection size of 2.8-4.8 cm^3 depending on the sequencing panel size and on the mutation frequency. The expected detection sizes correspond to lead times of 390-520 days compared to current median lung tumor sizes at diagnosis. This quantitative framework provides a mechanistic interpretation of ctDNA-based cancer detection approaches and helps to optimize cancer early detection strategies.

[*] correspondence should be addressed to johannes.reiter@stanford.edu

Introduction

Early stage cancer patients are more likely to be cured than advanced stage cancer patients (Etzioni et al. 2003; Song et al. 2018; Srivastava et al. 2019; Mattox et al. 2019). For example, the five-year survival rate of lung cancer patients who are diagnosed at a localized stage is 57% while for those diagnosed with distant metastases is only 5% (Siegel, Miller, and Jemal 2020). Unfortunately, only 16% of lung cancers are diagnosed at a localized stage. Recently, multiple studies presented new minimally invasive approaches to detect cancer-specific biomarkers in blood samples (Bettegowda et al. 2014; Newman et al. 2014; Phallen et al. 2017; Cohen et al. 2018; Mouliere et al. 2018; Shen et al. 2018; Cristiano et al. 2019; Ulz et al. 2016). These studies focused on cell-free DNA (cfDNA) which is released into the blood stream when cells undergo apoptosis (Wan et al. 2017; Heitzer et al. 2019). Most cfDNA in the bloodstream derives from normal cells, while a small proportion can derive from tumor cells and is known as circulating tumor DNA (ctDNA). Smaller tumors are therefore harder to detect because fewer tumor cells undergo apoptosis and their ctDNA makes up less than 0.01% of the cfDNA in the blood (Abbosh et al. 2017; Newman et al. 2016).

Despite this challenge, previous studies successfully demonstrated that 30-100% of symptomatic tumors (mostly larger than 3 cm^3) can be detected from a 10 – 15 mL blood sample (Cohen et al. 2018; Phallen et al. 2017). However, assessing whether blood-based tests can also detect still asymptomatic tumors at sizes smaller than 3 cm^3 with a sufficiently high specificity to reduce cancer mortality requires elaborate clinical trials with tens of thousands of participants (Aravanis, Lee, and Klausner 2017). While the first trials are already on the way, we lack mechanistic frameworks to predict the expected detection size for a given sequencing panel and screening frequency such that the clinical trials can focus on approaches with the highest success probability. For example, how would the performance of a screening test change for a subpopulation with tumors with half as many mutations (e.g., lung cancers of non-smokers vs. smokers)? Motivated by these fundamental questions, we developed a stochastic mathematical model of cancer evolution and biomarker shedding to study the potential and the limitations of blood-based cancer early detection tests across various scenarios (Hori and Gambhir 2011; Hori et al. 2017; Vogelstein et al. 2013; Greaves and Maley 2012). This framework will help to predict the performance of ctDNA-based tumor detection approaches and thereby inform the design of future clinical trials to find cancers earlier.

Results

Mathematical model of cancer evolution and ctDNA shedding

We first consider early stage lung cancers with a typical tumor volume doubling time of 181 days leading to a net growth rate of $r = \ln(2)/181 \approx 0.4\%$ per day (Winer-Muram et al. 2002). Lung cancer cells approximately divide with a birth rate of $b = 0.14$ per day (Rew and Wilson 2000), and die with a death rate of $d = b - r = 0.136$ per day (**Fig. 1**). For now, we assume that tumor cells release ctDNA into the bloodstream exclusively during apoptosis with a ctDNA shedding probability of q_d per cell death. Later, we will explore ctDNA shedding during

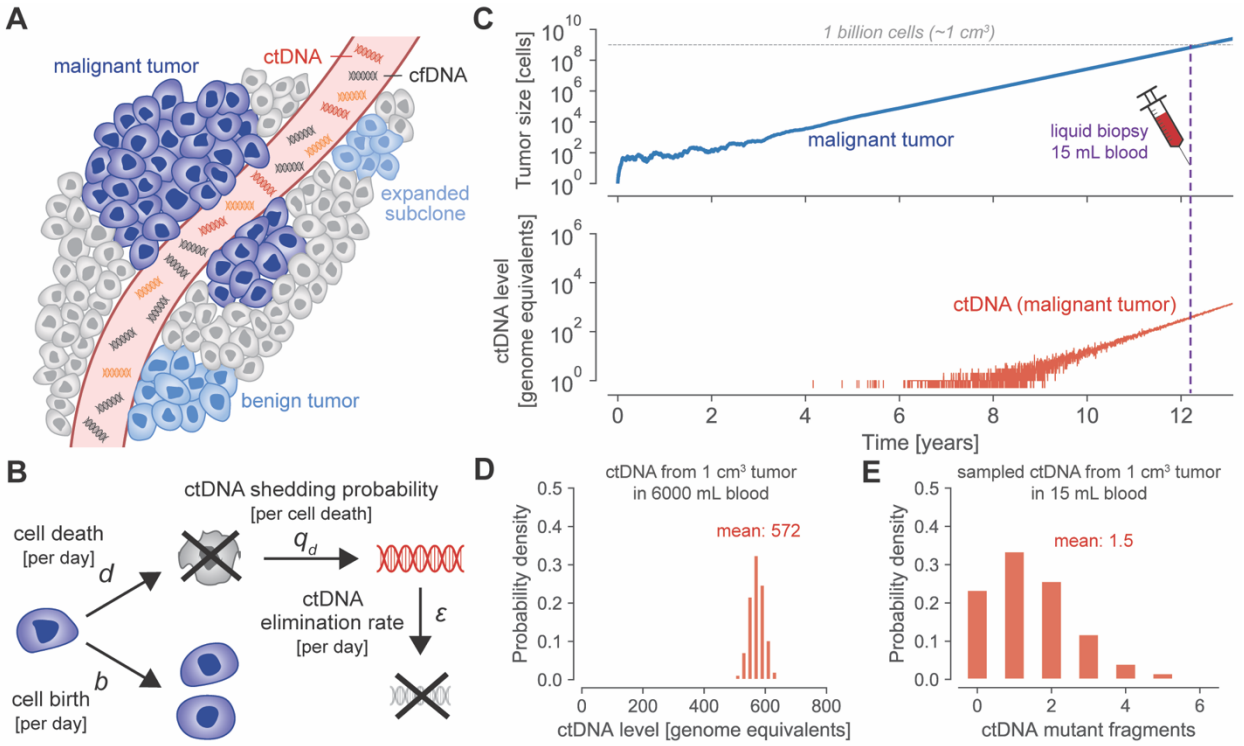

Fig. 1: Evolutionary dynamics of ctDNA shed by a growing cancer. A | Cells shed cell-free DNA (cfDNA) into the bloodstream. Tumor cells shed cell-free circulating tumor DNA (ctDNA) into the bloodstream. **B |** Mathematical model of cancer evolution and ctDNA shedding. Tumor cells divide with birth rate *b* and die with death rate *d* per day. During cell apoptosis, cells shed ctDNA into the bloodstream with probability q_d. Cell-free DNA is eliminated from the bloodstream with rate ε per day according to the half-life time of ctDNA, $t_{1/2}$ = 30 minutes. **C |** A primary tumor starts to grow at time zero with a growth rate of $r = b - d = 0.4\%$ ($b = 0.14$, $d = 0.136$), typical for early-stage lung cancers (tumor doubling time of 181 days). The tumor sheds ctDNA into the bloodstream according to the product of the cell death rate *d* and the shedding probability per cell death q_d. **D |** Distribution of ctDNA diploid genome equivalents (GE) in the entire bloodstream shed by the primary tumor at the time (purple dashed line in panel **C**) when the tumor reaches 1 billion cells (≈1 cm³; gray dashed line in panel **C**) leading to a mean tumor concentration of 0.19 GE per plasma mL and a tumor fraction of 0.02% at a DNA concentration of 6.3 ng per plasma mL. **E |** Probability distribution of ctDNA mutant fragments shed by the primary tumor present in a liquid biopsy of 15 mL of blood when the malignant tumor reaches 1 billion cells (assuming the covered heterozygous mutation is present in all cancer cells). Parameter values: ctDNA shedding probability per cell death $q_d = 1.4 \times 10^{-4}$ (**Supplementary Information**).

other processes such as necrosis or proliferation (Wan et al. 2017; Mouliere et al. 2018). By reanalyzing ctDNA sequencing data of an early stage lung cancer cohort (Abbosh et al. 2017), we inferred a mean shedding probability of $q_d \approx 1.4 \times 10^{-4}$ diploid genome equivalents (GE) per cell death (90% confidence interval, CI: $2.7 \times 10^{-5} - 7.0 \times 10^{-4}$; **fig. S1**; **Supplementary Information**). For a ctDNA half-life time of $t_{1/2} = 30$ minutes (Wan et al. 2017), we calculated a ctDNA elimination rate of $\varepsilon \approx 33.3$ per day. We illustrate a typical realization of this evolutionary process in **Fig. 1C**. A malignant tumor grows exponentially with a growth rate of $r = 0.4\%$ and releases ctDNA into the bloodstream with a shedding rate of $d \cdot q_d$. At a primary tumor size of 1 cm³ (1 billion cells), we find on average 572 ctDNA GE circulating in the bloodstream (**Fig. 1D**). A 15 mL blood sample contains approximately 1.5 ctDNA GE (**Fig. 1E**). At a plasma DNA concentration of 6.3 ng per plasma mL, 1.5 GE per plasma mL correspond to a

tumor fraction of 0.02% (assuming 6.6 pg per GE; **Supplementary Information**). The unit of GE can also be interpreted as the expected number of ctDNA fragments that exhibit a specific somatic heterozygous mutation. Hence, the number of ctDNA genome equivalents in a sample represents a biological limitation to detect a specific mutation in the tumor. In comparison, given a number of DNA fragments covering a specific genomic region, the number of mutated DNA fragments can be converted to a variant allele frequency (VAF) representing a technological detection limitation due to sequencing errors (Newman et al. 2016; Phallen et al. 2017).

Next we aimed to calculate the expected number of ctDNA genome equivalents, C, circulating in the bloodstream for any tumor size M and derived a surprisingly simple closed-form expression. The number of ctDNA genome equivalents circulating when the tumor reaches a size of M cells follows a Poisson distribution with a mean of $C = M \cdot d \cdot q_d/(\varepsilon + r)$ (for $M \gg 1$ and $d \cdot q_d \ll 1$; **Supplementary Text S1**). For a tumor with 1 billion cells, we calculated a mean of $C \approx 572$ ctDNA GE which perfectly matched the results from the exact computer simulations of the above defined branching process (**Fig. 1D**).

To further demonstrate the generality of this framework and the accuracy of our analytical results, we considered tumors with different sizes, growth rates, and cell turnover rates. As expected, a tumor with 0.5 billion cells leads to roughly half the number of circulating biomarkers ($C \approx 286$ ctDNA GE; **Fig. 2A**). More surprisingly, a slowly growing lung cancer ($r = 0.1\%$) leads to a significantly higher number of 585 GE than a faster growing cancer ($r = 4\%$) with 420 GE at the same size of 1 cm^3 (**Fig. 2B**; assuming that the faster growth is achieved by a lower death rate). If instead the faster growth is achieved by a higher birth rate and equal death rates, we find a smaller difference (571.6 vs. 572.3 GE). In comparison with colorectal cancer cells that divide on average with a birth rate of $b = 0.25$ ($d = 0.246$ for the same growth rate) (Bozic et al. 2010), we observed that the higher cell turnover rate of colorectal tumors leads to an almost 2-fold (1036 vs 572 GE) increase in the amount of ctDNA in the bloodstream because of the increased rate of cells undergoing apoptosis (**Fig. 2C**). Generally, the tumor growth dynamics, the ctDNA half-life time, and the ctDNA shedding rate strongly influence ctDNA levels (**fig. S2**). The analytical results were validated by perfectly matching exact simulation results across all considered scenarios (full lines vs bars in histogram of **Fig. 2**, **A–C**; **fig. S3**; **tables S1** and **S2**).

Using this mathematical framework of cancer evolution (Durrett 2015; Beerenwinkel et al. 2015; Altrock, Liu, and Michor 2015; Wodarz and Komarova 2014), we can predict the expected tumor detection size for an early detection test based on somatic point mutations in ctDNA (Phallen et al. 2017; Cohen et al. 2018). To compute realistic tumor detection sizes, we considered various sources of biological and technical errors. Given the number of wildtype and mutant fragments, we calculated the probability that a mutation arose from sequencing errors assuming a sequencing error rate of $1.5 \cdot 10^{-5}$ per base-pair (Newman et al. 2016)

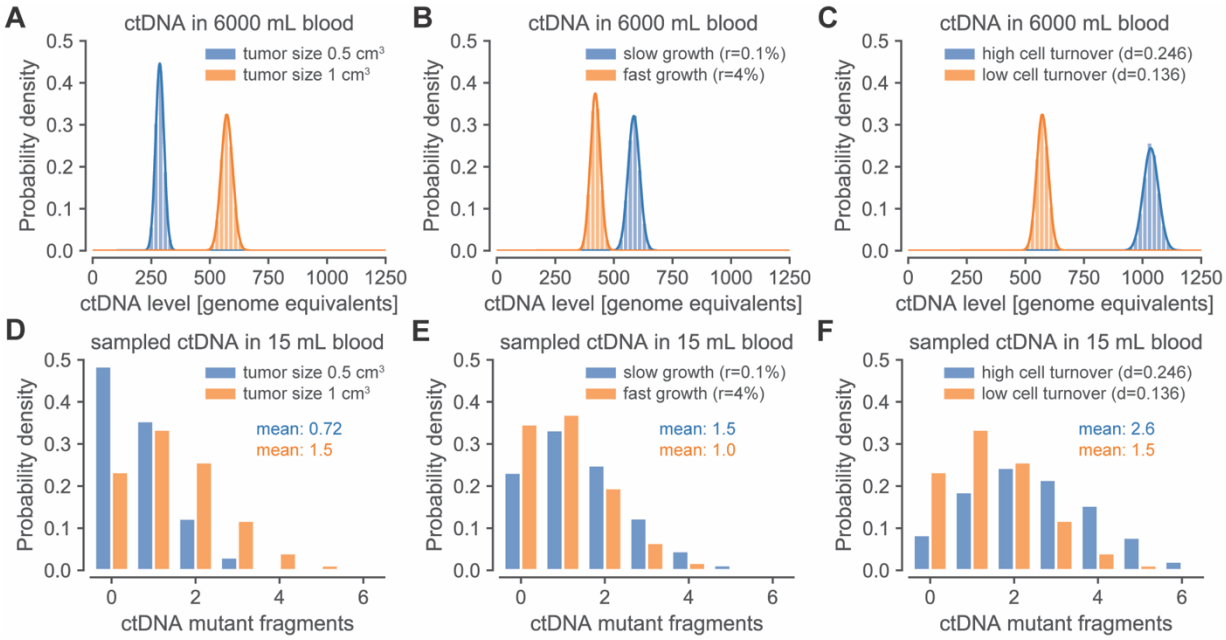

Fig. 2: Tumor growth rate and cell turnover strongly affect the amount of ctDNA. Top panels (**A-C**) depict diploid genome equivalents of ctDNA present in entire bloodstream when a lung tumor reaches a given size. Bars illustrate distribution of genome equivalents of ctDNA based on 10^4 simulation realizations. Full lines illustrate asymptotic results (Equation (2); **Supplementary Text S1**) and perfectly agree with simulation results (bars). Bottom panels (**D-F**) depict genome equivalents of ctDNA present in a 15 mL blood sample. **A** | Tumors with 500 million cells (≈ 0.5 cm^3) lead to low levels of ctDNA present in the bloodstream. **B** | Fast growing tumors lead to a lower level of ctDNA when they reach a size of 10^9 cells (≈ 1 cm^3) because fewer cell deaths decrease the amount of released ctDNA. **C** | High cell turnover rates (e.g. colorectal cancer) lead to a higher level of ctDNA at a given tumor size (1 billion cells) compared to lower cell turnover rates (e.g. lung cancer). **D-F** | Probability distribution of ctDNA genome equivalents present in a liquid biopsy of 15 mL blood at the time when the tumor reaches the given sizes. Parameter values: birth rate b = 0.14 per cell per day and death rate d = 0.136 per cell per day (for fast growth d = 0.1, for slow growth d = 0.139); b = 0.25 and d = 0.246 for high cell turnover; ctDNA half-life time $t_{1/2}$ = 30 minutes; ctDNA shedding probability per cell death q_d = 1.4 x 10^{-4}.

(**Supplementary Information**). To comprehend the variation of DNA concentration in plasma samples, we reanalyzed previously measured plasma DNA concentrations from Cohen et al. (Cohen et al. 2018). We used a Gamma distribution with a median of 5.2 ng of DNA per plasma mL (= 788 GE per plasma mL) to model the variability of plasma DNA concentration (**fig. S4B**; **Supplementary Information**). This analysis also revealed that plasma DNA concentrations increased in advanced cancer patients more than expected by the ctDNA amount shed from larger tumors (**fig. S4E**). Although the plasma DNA concentration increased with age in the cohort of Cohen et al. ($R^2 = 0.18$, $P = 3.3 \cdot 10^{-4}$; **fig. S4F**) and the healthy patients were on average younger than the cancer patients (mean of 49 vs. 63 years), this age difference explains less than 20% of the observed plasma DNA concentration increase in cancer patients.

We distinguish two types of early cancer detection tests because of their distinct clinical use-cases and requirements: i) cancer screening (somatic mutations of the tumor are not known a priori) and ii) cancer relapse detection (somatic mutations are known a priori by sequencing a sample of the primary tumor). We start with the fundamentally simpler detection problem (ii)

where the mutations are known and therefore relatively small custom sequencing panels can be used to detect a relapsing tumor.

Tumor relapse detection if mutations are known a priori

For cancer relapse detection, we considered an aggressive lung tumor growing with $r = 1\%$ per day (doubling time of 69 days) and assumed a sequencing panel that covers 20 tumor-specific mutations (Abbosh et al. 2017; McDonald et al. 2019; Newman et al. 2016; Tie et al. 2016; Reinert et al. 2019; Khan et al. 2018). Requiring that at least one of these 20 tumor-specific mutations needs to be called present in the plasma sample to infer that the tumor relapsed, we find an AUC (area under the curve) for the ROC (receiver operating characteristic) curve of 86% for tumors with 0.2 cm^3 (**Fig. 3A**). At a specificity of 99.5%, we observed a sensitivity of 20% for tumors with 0.2 cm^3. Repeatedly applying this virtual early detection test to a relapsing lung tumor led to a median detection size of 190 mm^3 for monthly sampling and 530 mm^3 for quarterly sampling (**fig. S5**). Important to note is that although the same test with a specificity of

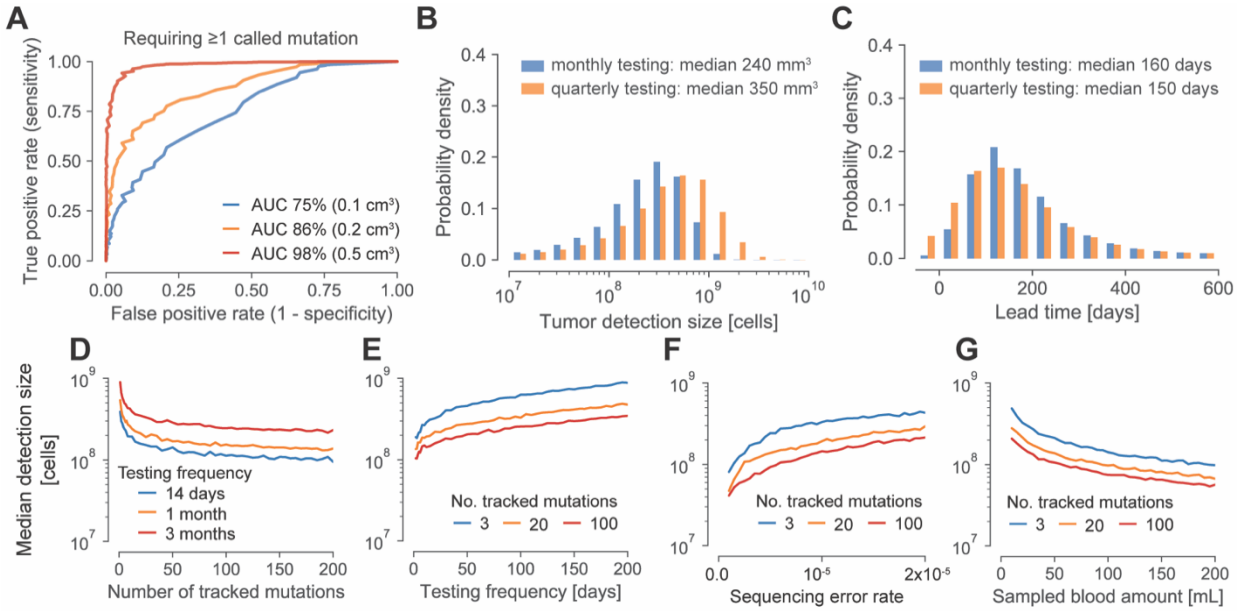

Fig. 3: Expected tumor relapse detection size and lead time compared to current clinical relapse detection. A | ROC (Receiver Operating Characteristic) curves for tumors with 100 million cells (≈ 0.1 cm^3, blue line), 200 million cells (≈ 0.2 cm^3, orange line), or 500 million cells (≈ 0.5 cm^3, red line) when 20 clonal tumor-specific mutations are tracked for relapse detection and one of these 20 mutations needs to be called for a positive test. Sensitivities of 7.3%, 20%, and 63% for tumors with 100 million, 200 million, and 500 million cells at a specificity of 99.5%. AUC, area under the curve. **B-G |** For better comparability, positive detection test thresholds were set such that if the test is repeated over a full year, a false positive rate of 5% is obtained. **B |** Expected tumor detection size distributions for monthly and quarterly repeated relapse detection tests (sequencing panel covers 20 mutations). **C |** Expected lead time distributions compared to imaging-based approaches applied at the same frequency with a detection limit of 1 cm^3 for monthly and quarterly repeated relapse detection tests (sequencing panel covers 20 clonal mutations). **D-G |** Median tumor detection sizes over the number of clonal mutations covered by the sequencing panel (panel **D**), the testing frequency (panel **E**), the sequencing error rate (panel **F**), and the sampled blood amount (panel **G**). Parameter values typical for lung cancer (if not differently specified): birth rate b = 0.14 per cell per day; death rate d = 0.136 per cell per day; ctDNA half-life time $t_{1/2}$ = 30 minutes; ctDNA shedding probability per cell death q_d = 1.4 x 10^{-4}; sequencing error rate per base-pair 1.5 x 10^{-5}; 15 mL blood sampled per test; DNA median concentration 5.2 ng per plasma mL.

99.5% has been applied for both sampling frequencies, the monthly sampling produces 0.06 false-positive test results over 12 months of relapse testing while the quarterly sampling only produces 0.02 false-positives over 12 months. For a fair comparison, we adjusted the mutation calling thresholds such that 0.05 false-positives are expected for both sampling frequencies over 12 months of relapse testing. With this adjustment, the median detection size of quarterly testing decreased to 350 mm^3. Monthly relapse testing still led to a 31% smaller median detection size of 240 mm^3 (**Fig. 3B**).

To further assess ctDNA-based relapse detection approaches, we computed the expected lead time to imaging-based relapse detection when applied at the same frequency (ignoring the differences in the costs). We conservatively assumed a radiological detection limit of exactly 1 cm^3 and a specificity of 100% above that limit. The median lead time of monthly ctDNA testing compared to imaging was 160 days (**Fig. 3C**). Quarterly ctDNA testing yielded a similar lead time of 150 days because of our assumption that imaging is performed at the same frequency. These predicted lead times for early stage lung cancer are similar to the reported median of ~160 days by Chaudhuri et al. (Chaudhuri et al. 2017) and slightly longer than the reported median of 70 days (range, 10 – 346 days) by Abbosh et al. (2017), likely due to less frequent relapse testing and a lower average number of clonal mutations covered by the sequencing panel (range 3-26).

Next, we investigated how the sequencing panel size, sampling frequency, blood sample amount, sequencing error, and the number of called mutations for detection affect the tumor detection size. Although we kept the expected number of false-positives per year again constant at 0.05, the median tumor detection size strongly decreased up to a sequencing panel size of approximately 40 and then continued to minimally decrease (**Fig. 3D**). As expected, the median tumor detection size also decreased with an increasing sampling frequency (**Fig. 3E**). Weekly and more frequent relapse testing led to a large drop of the median detection size. Decreasing the sequencing error rate strongly decreased the detection size as it simultaneously increases the sensitivity and specificity (**Fig. 3F**). Increasing the amount of sampled blood, led to a similarly strong decrease of the expected detection size (**Fig. 3G**). Moreover, we found that requiring multiple called mutations for a positive detection test additionally helped to increase the sensitivity at the same specificity (Wan et al. 2019) (**fig. S6**).

Tumor detection without a priori known mutations

If the mutations in the tumor are not known a priori, cancer detection becomes fundamentally more complex. Two major considerations for ctDNA-based cancer early detection are the expected number of mutations per tumor covered by the sequencing panel and the underlying sequencing error rate per base-pair of the assay. The expected number of somatic mutations covered by the sequencing panel can be maximized by focusing on recurrently mutated regions such that many more mutations per sequenced Mega-base are observed than expected from the average lung cancer mutation frequency of ~10 mutations per Mega-base. For example,

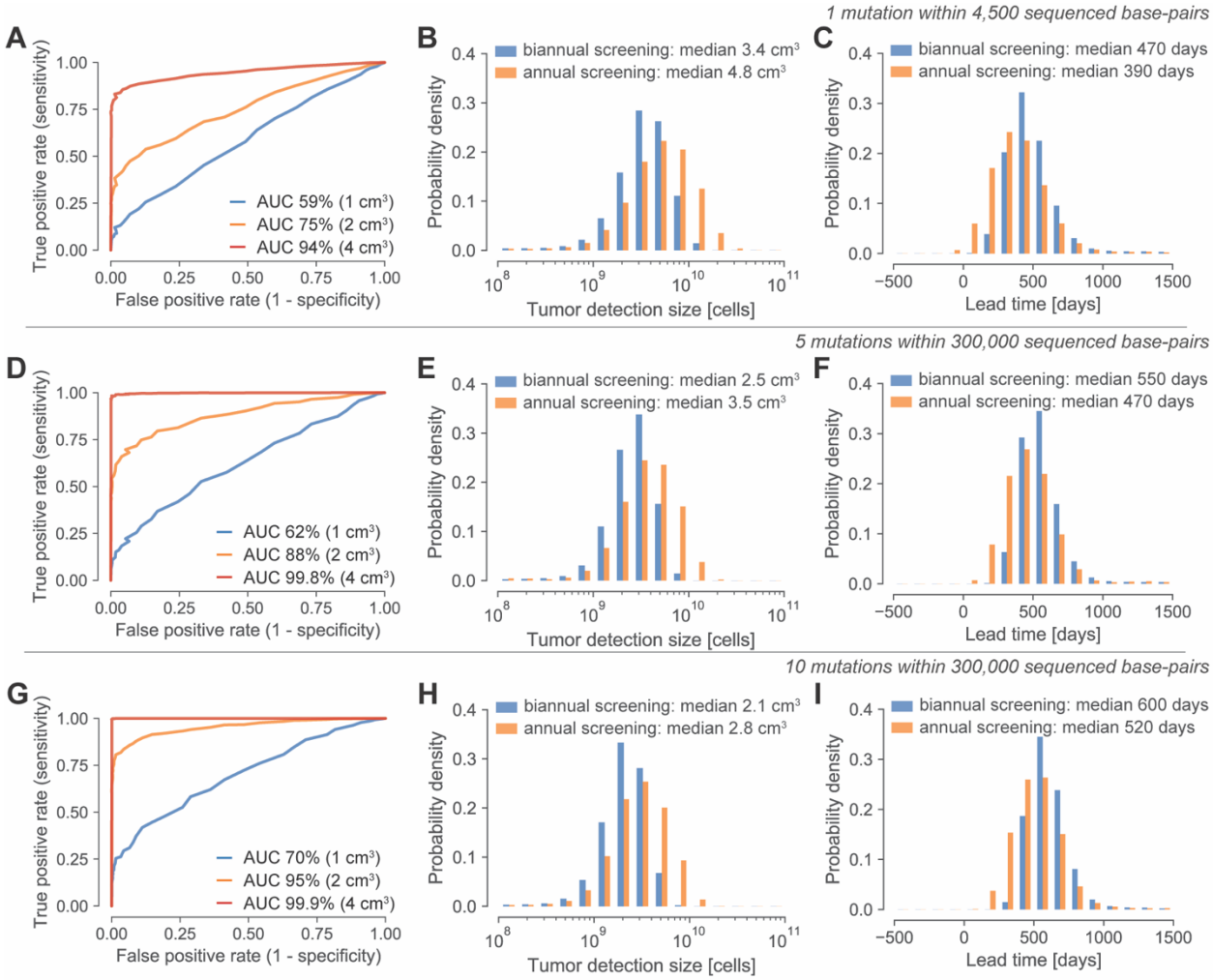

Fig. 4: Expected tumor detection size and lead time distributions for different screening scenarios. A,D,G | ROC (Receiver Operating Characteristic) curves for different sequencing panels used on plasma samples of subjects with eight benign lesions (each 34 million cells) and a malignant tumor with 1 billion cells (≈ 1 cm^3, blue line), 2 billion cells (≈ 2 cm^3, orange line), or 4 billion cells (≈ 4 cm^3, red line). AUC, area under the curve. **A |** Sensitivities of 8.0%, 32%, and 79% for tumors with 1 billion, 2 billion, and 4 billion cells at a test specificity of 99%. **B,E,H |** Tumor detection size distributions for biannually (blue bars) and annually (orange bars) repeated virtual screening tests for different sequencing panels and number of mutations as expected for cancers of never-smoking subjects (panel **E**) and subjects with a smoking history (panel **H**). Number of expected mutations for the smaller panel is similar for cancers of smokers and non-smokers (panel **B**). Positive screening test thresholds were set such that 0.01 false-positives are expected per year of screening. **C,F,I |** Lead time distributions for biannually (blue bars) and annually (orange bars) repeated virtual screening tests compared to current clinical diagnosis times calculated from detection sizes in the SEER database and assuming a typical early stage lung cancer growth rate of 0.4% per day. **D |** Sensitivities of 11%, 56%, and 98% for tumors with 1 billion, 2 billion, and 4 billion cells at a test specificity of 99%. **G |** Sensitivities of 19%, 76%, and 99.9% for tumors with 1 billion, 2 billion, and 4 billion cells at a test specificity of 99%. Parameter values typical for early stage lung cancer: birth rate b = 0.14 per cell per day; death rate d = 0.136 per cell per day; ctDNA half-life time $t_{1/2}$ = 30 minutes; ctDNA shedding probability per cell death q_d = 1.4 x 10^{-4}; sequencing error rate per base-pair 1.5 x 10^{-5}; 15 mL blood sampled per test; DNA median concentration 5.2 ng per plasma mL.

CAPP-Seq (spanning ~300,000 base-pairs) and CancerSEEK (spanning ~4,500 base-pairs) cover on average 9.1 and 1.1 mutations per early stage lung cancer in the Abbosh et al. cohort (Newman et al. 2016; Cohen et al. 2018; Abbosh et al. 2017). However, expanded clones in normal tissue and benign tumors frequently exhibit the same recurrent mutations as cancer cells

(Martincorena and Campbell 2015; Yokoyama et al. 2019; Makohon-Moore et al. 2018), and hence DNA shed from these cells can hamper the specificity of a cancer screening test (Dudley et al. 2019). Fortunately, most benign lesions are smaller than malignant tumors and since benign cells typically replicate with a lower rate than malignant cells, they are expected to release comparatively less ctDNA into the bloodstream (Rew and Wilson 2000). Moreover, we assume that white blood cells were sequenced separately to filter mutations related to clonal hematopoiesis of indeterminate potential present in hematopoietic stem cells (Cohen et al. 2018).

We consider again an early stage lung tumor growing with $r = 0.4\%$ per day. For a sequencing panel covering on average one mutation per lung cancer across 4,500 base-pairs, we computed AUC values for the ROC curves of 59%, 75%, and 94% for tumors at sizes of 1 cm^3, 2 cm^3, and 4 cm^3, respectively (**Fig. 4A**). Based on previous lung cancer screening studies, we assumed that eight benign lung nodules with each 34 million cells (diameter of ~4 mm) replicated at a constant size with a cell birth/death rate of $b_{bn} = d_{bn} = 0.07$ per day. Per benign nodule, one mutation was assumed to be covered by the sequencing panel (McWilliams et al. 2013; Rew and Wilson 2000). At a specificity of 99%, we computed a sensitivity of 32% for tumors with 2 cm^3 (diameter of 16 mm). Repeating this virtual early detection test annually for a growing tumor, we obtain a median detection size of 4.8 cm^3 (**Fig. 4B**). This detection size would be 79% smaller than the current median detection size of approximately 22.5 cm^3 (diameter of 35 mm; assuming that tumors are approximately spherical) for lung cancers reported in the SEER database from 2005 to 2015 (NCI Seer 2018). Comparing the computed detection size and the SEER median size at diagnosis, we calculated a lead time to current diagnosis times of 390 days for a typical growth rate of early stage lung cancer (**Fig. 4C**). For a screening test repeated every 6 months, we computed a median detection size of 3.4 cm^3 (adjusted specificity to 99.5% to obtain the same number of 0.01 false-positives per year of screening). Faster growing tumors led to larger detection sizes. For example, for more than twice as fast-growing tumors ($r = 1\%$, tumor volume doubling time of 70 days), we computed a median detection size of 15 cm^3 for annual screening (**fig. S7A**). Note that these tumors grow from 0.1 cm^3 to 3.8 cm^3 in just one year.

In comparison, a sequencing panel covering 300,000 base-pairs (e.g., CAPP-Seq) led to AUC values for the ROC curves of 88% and 95% for lung tumors with a size of 2 cm^3 in never-smoking subjects and in subjects with a smoking history, respectively (**Fig. 4, D** and **G**). We separately analyzed smokers and non-smokers here because in contrast to the above evaluated panel focused on driver gene mutations, the expected number of mutations covered by the larger panel for cancers of never-smoking subjects was lower than for cancers of subjects with a smoking history (5 vs. 10 mutations; **table S3**). For an annually repeated screening test, the expected median detection size was 3.4 cm^3 for lung tumors of never-smokers and 2.8 cm^3 for tumors of subjects with a smoking-history (**Fig. 4, E** and **H**). These detection sizes correspond to lead times compared to current tumor sizes at diagnosis of 470 and 520 days for tumors of never-smokers and smokers, respectively (**Fig. 4, F** and **I**).

While these predicted tumor detection sizes are encouraging, a major challenge for every cancer detection test is the low incidence rate of cancers. Relatively common cancers such as lung cancer have yearly incidence rates of approximately 1 in 2,000 (Siegel, Miller, and Jemal 2020). Because more than 98% of lung cancers occur in people above 50 (~30% of the US population), the incidence rate increases to 1 in 600 in this age group. Hence, approximately six false-positives would be expected for one true-positive at a test specificity of 99%. We calculated a PPV (positive predictive value) of 5.5% and an NPV (negative predictive value) of 99.9% for a lung tumor with 2 billion cells. For heavy smokers in their late sixties the incidence rate increases to 1 in 120 and the same test would have a PPV of 22.6% and an NPV of 99.4% for lung cancers with 2 billion cells. Screening for many common cancer types with one joint detection test will help to improve the PPV but might require larger sequencing panels potentially increasing biological and sequencing artefacts.

ctDNA shedding during apoptosis, necrosis, and proliferation

So far we assumed that ctDNA shedding occurs exclusively during apoptosis. We can generalize our framework such that the effective shedding rate λ is given by the sum of ctDNA fragments shed during apoptosis, necrosis, and proliferation: $\lambda = d \cdot q_d + q_n + b \cdot q_b$ where q_b denotes the shedding probability per cell division and q_n denotes the shedding rate from necrosis per unit of time. We show that independent of the three shedding processes, the amount of ctDNA when the tumor reaches a size of M cells remains approximately Poisson-distributed with a mean of $C = M \cdot \lambda/(\varepsilon + r)$ (assuming $M \gg 1$ and $\lambda \ll 1$; **Supplementary Text S1**) where the effective shedding rate simply represents the sum of the operating ctDNA shedding processes (**fig. S8**; **table S1**).

Discussion

Our mathematical framework provides a theoretical upper-bound for the performance of mutation-based ctDNA cancer early detection tests. A critical parameter for these estimates is the normalized ctDNA shedding rate per cell. Based on ctDNA data of early stage lung cancer, the stochastic model suggests that decreasing the sequencing error rate, increasing the amount of sampled plasma, increasing the sequencing panel size, and increasing the sampling rate can drastically decrease the expected tumor relapse detection size at the same normalized annual false-positive rate (**Fig. 3**). The virtual screening computations indicate that lung tumors would be detected when they reach a diameter of 2.1 mm in an annual screening program with a sequencing panel of 4,500 base-pairs (**Fig. 4B**). According to the lung cancer staging system for tumor sizes, 7% of the detected cases would be classified as T1a (≤1 cm), 36% as T1b (>1 but ≤2 cm), and 51% as T1c (>2 but ≤3 cm) (Detterbeck et al. 2017). Only 6% of the detected tumors would have reached sizes for stages T2a (>3 but ≤4 cm) or beyond. Although these calculations suggest that most tumors can only be detected when they reach sizes of billions of

cells, detecting some tumors before they become symptomatic and shifting some diagnosis to an earlier stage can have an enormous impact on cancer mortality (Mattox et al. 2019).

This study has several limitations. First, our understanding of ctDNA shedding and its variance across lung tumors and tumor types is still in its infancy and more studies such as TRACERx with ctDNA and CT volumetric analysis are required to inform the shedding rate inference. Second, the ctDNA shedding dynamics of precursor lesions and how their presence interferes with cancer early detection are largely unknown. We conservatively assumed that the ctDNA shedding probability per cell death in benign lesions is the same as in malignant tumors. Nonetheless, it is plausible that benign cells exhibit a lower shedding rate because they have not invaded surrounding tissue. Third, our analysis was limited to point mutations present in ctDNA. Including additional cancer-associated characteristics of ctDNA or other biomarkers can help to further decrease the expected detection size (Cohen et al. 2018; Shen et al. 2018; Mouliere et al. 2018; Cristiano et al. 2019; Ulz et al. 2016; Heitzer et al. 2019). Last, while we validated our mathematical results through exact computer simulations, the predicted tumor detection sizes will need to be validated in large prospective studies.

A major challenge of cancer early detection is the stochastic nature of cancer initiation and progression. While new technologies can detect smaller and smaller tumors, the optimal treatment of asymptomatic tumors is often unclear and needs to be balanced with the risk for overtreatment as well as undertreatment (Srivastava et al. 2019; Mattox et al. 2019). For example, in more than 20% of smokers suspicious lesions can be found by low-dose computed tomographic screening. Nevertheless, lung cancer was detected only in ~0.6% of screened smokers within a 10-year follow-up (The National Lung Screening Trial Research Team 2011; de Koning et al. 2020). Similarly, around 33% of humans harbor precursor lesions in their pancreas but the life-time risk of developing pancreatic cancer is 1.6% (NCI Seer 2018; Makohon-Moore and Iacobuzio-Donahue 2016). Hence, most precursor lesions do not progress to cancer within the life-time of humans (Makohon-Moore et al. 2018; Lang et al. 2020). Because cancer incidence strongly depends on factors such as age, genetic predisposition, life-style, or exposure to mutagens (e.g., sun, smoke, etc.), screening programs often focus on high-risk individuals to decrease the chances of overtreatment. Our results show that cancer screening and surveillance strategies can be further optimized and personalized by comprehensive mathematical models of cancer evolution (de Koning et al. 2014; Curtius et al. 2020; Ryser et al. 2016; Lahouel et al. 2020; Lang et al. 2020).

Supplementary Information

Parameter selection and inference

We used previously measured values of cell division rates and tumor volume doubling times to obtain the tumor growth rates and death rates. The estimated average time between cell divisions of lung cancer cells is 7 days (Rew and Wilson 2000), resulting in a cell birth rate of $b = 0.14$ per day. Given a volume doubling time of ~180 days of stage I lung cancers (Winer-Muram et al. 2002), we find a tumor growth rate $r = \ln(2)/180 \approx 0.4\%$ per day and a death rate of $d = b - r = 0.136$ per day. Since benign lung nodules have doubling times of ≥ 500 days (Winer-Muram et al. 2002), we explore the effects of nodules at fixed sizes and assume that cells in benign lung nodules replicate with rates of $b_{\mathrm{bn}} = d_{\mathrm{bn}} = 0.07$ per day (Rew and Wilson 2000). The estimated average time between cell divisions of colorectal cancer cells is 4 days (Rew and Wilson 2000), resulting in a cell birth rate of $b = 0.25$ per day.

To estimate the ctDNA shedding rate per cancer cell, we reanalyzed data from the TRACERx study (Abbosh et al. 2017). Tumor volumes were estimated from preoperative CT scans (see Supplementary Table 1 of Abbosh et al. (2017)). We calculated the mean VAF (variant allele frequency) of somatic mutations in each liquid biopsy across all mutations that were identified as clonal in the primary tumor (see Supplementary Table 5 of Abbosh et al. (2017)). We studied the correlation between tumor volume and the mean VAF of clonal mutations as well as the mean mutant fragment number of clonal mutations in 24 patients. We excluded all patients where cancer cells were found in lymph nodes because those cancer cells were not part of the tumor volume estimation. The mean VAF of clonal mutations strongly correlated with tumor volume (Spearman $\rho = 0.53$, $P = 0.0083$; **fig. S1A**). This correlation is slightly weaker than reported in the original TRACERx study because the mean frequency was calculated from all clonal mutations including those mutations that were undetectable in the plasma sample. Tumor volume showed a stronger correlation with the mean mutant fragment number of clonal mutations (Spearman $\rho = 0.66$, $P < 0.001$; **fig. S1B**), possibly because plasma DNA concentrations vary strongly even in healthy patients and thereby complicate a strictly frequency-based analysis (**fig. S4**). Linear regression predicts 0.19 mutant fragments of clonal heterozygous mutations per plasma mL per cm^3 of tumor tissue.

To infer the shedding rate λ, we used the estimated 0.19 mutant fragments per plasma mL and set up the following equilibrium equation of $\lambda \cdot 10^9$ cells per $\mathrm{cm}^3 = 570 \cdot \varepsilon$, assuming 570 ($= 0.19 \cdot 3000$) diploid genome equivalents in 3000 mL of plasma. We followed previous estimates of the ctDNA half-life time of $t_{1/2} = 30$ minutes and explored the effects of half-life times from 20 to 120 minutes (**fig. S2**) (Wan et al. 2017). We calculated the ctDNA elimination rate as $\varepsilon = {}^{\ln 2}/_{t_{1/2}} \cdot 60 \cdot 24 \approx 33.3$ fragments per day. We found a shedding rate of $\lambda = 1.9 \cdot$

10^{-5} GE per cell per day leading to a shedding probability of $q_d = \lambda/_d = 1.4 \times 10^{-4}$ GE per cell death, assuming that ctDNA is exclusively shed during cell apoptosis. To estimate the shedding rate of cells in other tissue types, we can rescale the shedding probability by the corresponding cell death rate. For example, we estimate the shedding rate of cells in benign lung nodules as $\lambda_{bn} = d_{bn} \cdot q_d \approx 9.8 \cdot 10^{-6}$ GE per cell per day.

In silico sampling and sequencing of plasma DNA

To compare different screening strategies, we employed the virtual sampling and sequencing of plasma DNA. For each liquid biopsy, ctDNA fragments in 15 mL of blood were sampled from 6000 mL of blood according to a binomial distribution. For the virtual detection tests in **Figs. 3, 4** and **figs. S5-S7**, plasma cfDNA concentrations were sampled from a Gamma distribution (mean and median of 6.3 and 5.2 ng per plasma mL) as illustrated in **fig. S4B**. To calculate the cfDNA GE, we assumed a weight of 6.6 pg per GE (Heitzer et al. 2019) . The sequencing error rate per base-pair was set to $e_{seq} = 0.0015\%$ (Newman et al. 2016; Phallen et al. 2017). For each mutation covered by the sequencing panel, we account for sequencing errors by calculating a *P* value as $1 - \sum_{i=0}^{k-1} \binom{n}{i} \cdot {e_{seq}}^{i} \cdot \left(1 - e_{seq}\right)^{n-i}$ where *k* denotes the number of observed fragments supporting the mutation and *n* denotes the total number of fragments covering the mutation's genomic position.

Supplementary Materials

Supplementary Text S1: Mathematical modeling.

Fig. S1. Amount of ctDNA correlates with tumor volume in early stage lung cancer.

Fig. S2. Detection probability of ctDNA fragments in a 15 mL liquid biopsy at given tumor size depends on many parameters.

Fig. S3. Comparison between computer simulations, exact and asymptotic theoretical distributions conditioned on tumor survival.

Fig. S4. Plasma DNA concentration increases with disease progression.

Fig. S5. Expected tumor detection size and lead time compared to current clinical relapse detection at a fixed specificity of 99.5% per test.

Fig. S6. Expected tumor detection size and lead time compared to current clinical relapse detection at a fixed specificity of 99.5% per test.

Fig. S7. Expected tumor detection size distributions for fast growing lung cancers.

Fig. S8. Operating ctDNA shedding processes determine the effective shedding rate.

Table S1. Mean and variance for the distribution of the circulating biomarker C_t^A at time t shed by a tumor conditioned on its survival until time t across different shedding processes.

Table S2. Mean and variance for the distribution of the circulating biomarker C_t^A at time t shed by a tumor conditioned on its survival until time t across a wide range of shedding rates.

Table S3. Mean number of mutations of lung cancers covered by CancerSEEK and CAPP-Seq.

References

Abbosh, Christopher, Nicolai J Birkbak, Gareth A Wilson, Mariam Jamal-Hanjani, Tudor Constantin, Raheleh Salari, John Le Quesne, et al. 2017. "Phylogenetic CtDNA Analysis Depicts Early-Stage Lung Cancer Evolution." *Nature* 545 (7655): 446–51.

Altrock, Philipp M, Lin L Liu, and Franziska Michor. 2015. "The Mathematics of Cancer: Integrating Quantitative Models." *Nature Reviews Cancer* 15 (12): 730–45.

Aravanis, Alexander M, Mark Lee, and Richard D Klausner. 2017. "Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection." *Cell* 168 (4): 571–74. https://doi.org/10.1016/j.cell.2017.01.030.

Beerenwinkel, Niko, Roland F Schwarz, Moritz Gerstung, and Florian Markowetz. 2015. "Cancer Evolution: Mathematical Models and Computational Inference." *Systematic Biology* 64 (1): e1–25. https://doi.org/10.1093/sysbio/syu081.

Bettegowda, Chetan, Mark Sausen, Rebecca J Leary, Isaac Kinde, Yuxuan Wang, Nishant Agrawal, Bjarne R Bartlett, et al. 2014. "Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies." *Science Translational Medicine* 6 (224): 224ra24. https://doi.org/10.1126/scitranslmed.3007094.

Bozic, Ivana, Tibor Antal, Hisashi Ohtsuki, Hannah Carter, Dewey Kim, Sining Chen, Rachel Karchin, Kenneth W Kinzler, Bert Vogelstein, and Martin A Nowak. 2010. "Accumulation of Driver and Passenger Mutations during Tumor Progression." *Proc Natl Acad Sci USA* 107 (43): 18545–50. https://doi.org/10.1073/pnas.1010978107.

Chaudhuri, Aadel A, Jacob J Chabon, Alexander F Lovejoy, Aaron M Newman, Henning Stehr, Tej D Azad, Michael S Khodadoust, et al. 2017. "Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling." *Cancer Discovery* 7 (12): 1394–1403. https://doi.org/10.1158/2159-8290.CD-17-0716.

Cohen, Joshua D, Lu Li, Yuxuan Wang, Christopher Thoburn, Bahman Afsari, Ludmila Danilova, Christopher Douville, et al. 2018. "Detection and Localization of Surgically Resectable Cancers with a Multi-Analyte Blood Test." *Science* 357: 926–930. https://doi.org/10.1126/science.aar3247.

Cristiano, Stephen, Alessandro Leal, Jillian Phallen, Jacob Fiksel, Vilmos Adleff, Daniel C. Bruhm, Sarah Østrup Jensen, et al. 2019. "Genome-Wide Cell-Free DNA Fragmentation in Patients with Cancer." *Nature* 570 (7761): 385–89. https://doi.org/10.1038/s41586-019-1272-6.

Curtius, Kit, Anup Dewanji, William D Hazelton, Joel H Rubenstein, and E Georg Luebeck. 2020. "Optimal Timing for Cancer Screening and Adaptive Surveillance Using

Mathematical Modeling." *BioRxiv*. https://doi.org/10.1101/2020.02.11.927475.

Detterbeck, Frank C, Daniel J Boffa, Anthony W Kim, and Lynn T Tanoue. 2017. "The Eighth Edition Lung Cancer Stage Classification." *Chest* 151 (1): 193–203.

Dudley, Jonathan C, Joseph Schroers-Martin, Daniel V Lazzareschi, William Y Shi, Simon B Chen, Mohammad Shahrokh Esfahani, Dharati Trivedi, et al. 2019. "Detection and Surveillance of Bladder Cancer Using Urine Tumor DNA." *Cancer Discovery* 9 (4): 500–509. https://doi.org/10.1158/2159-8290.CD-18-0825.

Durrett, Richard. 2015. *Branching Process Models of Cancer*. Springer.

Etzioni, Ruth, Nicole Urban, Scott Ramsey, Martin McIntosh, Stephen Schwartz, Brian Reid, Jerald Radich, Garnet Anderson, and Leland Hartwell. 2003. "Early Detection: The Case for Early Detection." *Nature Reviews Cancer* 3 (4): 243.

Greaves, Mel, and Carlo C Maley. 2012. "Clonal Evolution in Cancer." *Nature* 481 (7381): 306–13. https://doi.org/10.1038/nature10762.

Heitzer, Ellen, Imran S Haque, Charles E S Roberts, and Michael R Speicher. 2019. "Current and Future Perspectives of Liquid Biopsies in Genomics-Driven Oncology." *Nature Reviews Genetics* 20 (2): 71–88. https://doi.org/10.1038/ s41576-018-0071-5.

Hori, Sharon Seiko, and Sanjiv Sam Gambhir. 2011. "Mathematical Model Identifies Blood Biomarker--Based Early Cancer Detection Strategies and Limitations." *Science Translational Medicine* 3 (109): 109ra116.

Hori, Sharon Seiko, Amelie M Lutz, Ramasamy Paulmurugan, and Sanjiv Sam Gambhir. 2017. "A Model-Based Personalized Cancer Screening Strategy for Detecting Early-Stage Tumors Using Blood-Borne Biomarkers." *Cancer Research* 77 (10): 2570–84. https://doi.org/10.1158/0008-5472.CAN-16-2904.

Khan, Khurum H, David Cunningham, Benjamin Werner, Georgios Vlachogiannis, Inmaculada Spiteri, Timon Heide, Javier Fernandez Mateos, et al. 2018. "Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial." *Cancer Discovery* 8 (10): 1270–85.

Koning, Harry J de, Carlijn M van der Aalst, Pim A de Jong, Ernst T Scholten, Kristiaan Nackaerts, Marjolein A Heuvelmans, Jan-Willem J Lammers, et al. 2020. "Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial." *New England Journal of Medicine*. https://doi.org/10.1056/NEJMoa1911793.

Koning, Harry J de, Rafael Meza, Sylvia K Plevritis, Kevin ten Haaf, Vidit N Munshi, Jihyoun Jeon, Saadet Ayca Erdogan, et al. 2014. "Benefits and Harms of Computed Tomography Lung Cancer Screening Strategies: A Comparative Modeling Study for the US Preventive Services Task Force." *Annals of Internal Medicine* 160 (5): 311–20.

Lahouel, Kamel, Laurent Younes, Ludmila Danilova, Francis M Giardiello, Ralph H Hruban, John Groopman, Kenneth W Kinzler, Bert Vogelstein, Donald Geman, and Cristian Tomasetti. 2020. "Revisiting the Tumorigenesis Timeline with a Data-Driven Generative Model." *Proc Natl Acad Sci USA* 117 (2): 857–64.

Lang, Brian M, Jack Kuipers, Benjamin Misselwitz, and Niko Beerenwinkel. 2020. "Predicting Colorectal Cancer Risk from Adenoma Detection via a Two-Type Branching Process

Model." *PLOS Computational Biology* 16 (2): e1007552.

Makohon-Moore, Alvin P, and Christine A Iacobuzio-Donahue. 2016. "Pancreatic Cancer Biology and Genetics from an Evolutionary Perspective." *Nature Reviews Cancer* 16 (9): 553–65. https://doi.org/10.1038/nrc.2016.66.

Makohon-Moore, Alvin P, Karen Matsukuma, Ming Zhang, Johannes G Reiter, Jeffrey M Gerold, Yuchen Jiao, Lisa Sikkema, et al. 2018. "Precancerous Neoplastic Cells Can Move through the Pancreatic Ductal System." *Nature* 561 (7722): 201–5.

Martincorena, Iñigo, and Peter J Campbell. 2015. "Somatic Mutation in Cancer and Normal Cells." *Science* 349 (6255): 1483–89.

Mattox, Austin K, Chetan Bettegowda, Shibin Zhou, Nickolas Papadopoulos, Kenneth W Kinzler, and Bert Vogelstein. 2019. "Applications of Liquid Biopsies for Cancer." *Science Translational Medicine* 11 (507): eaay1984.

McDonald, Bradon R, Tania Contente-Cuomo, Stephen-John Sammut, Ahuva Odenheimer-Bergman, Brenda Ernst, Nieves Perdigones, Suet-Feung Chin, et al. 2019. "Personalized Circulating Tumor DNA Analysis to Detect Residual Disease after Neoadjuvant Therapy in Breast Cancer." *Science Translational Medicine* 11 (504): eaax7392. https://doi.org/10.1126/scitranslmed.aax7392.

McWilliams, Annette, Martin C Tammemagi, John R Mayo, Heidi Roberts, Geoffrey Liu, Kam Soghrati, Kazuhiro Yasufuku, et al. 2013. "Probability of Cancer in Pulmonary Nodules Detected on First Screening CT." *New England Journal of Medicine* 369 (10): 910–19. https://doi.org/10.1056/NEJMoa1214726.

Mouliere, Florent, Dineika Chandrananda, Anna M Piskorz, Elizabeth K Moore, James Morris, Lise Barlebo Ahlborn, Richard Mair, et al. 2018. "Enhanced Detection of Circulating Tumor DNA by Fragment Size Analysis." *Science Translational Medicine* 10 (466): eaat4921. https://doi.org/10.1126/scitranslmed.aat4921.

NCI Seer. 2018. "Surveillance, Epidemiology, and End Results (SEER) Research Data 1973-2015 - ASCII Text Data." 2018. https://seer.cancer.gov.

Newman, Aaron M, Scott V Bratman, Jacqueline To, Jacob F Wynne, Neville C W Eclov, Leslie A Modlin, Chih Long Liu, et al. 2014. "An Ultrasensitive Method for Quantitating Circulating Tumor DNA with Broad Patient Coverage." *Nature Medicine* 20 (5): 548.

Newman, Aaron M, Alexander F Lovejoy, Daniel M Klass, David M Kurtz, Jacob J Chabon, Florian Scherer, Henning Stehr, et al. 2016. "Integrated Digital Error Suppression for Improved Detection of Circulating Tumor DNA." *Nature Biotechnology* 34 (5): 547.

Phallen, Jillian, Mark Sausen, Vilmos Adleff, Alessandro Leal, Carolyn Hruban, James White, Valsamo Anagnostou, et al. 2017. "Direct Detection of Early-Stage Cancers Using Circulating Tumor DNA." *Sci Transl Med* 9 (403): eaan2415. https://doi.org/10.1126/scitranslmed.aan2415.

Reinert, Thomas, Tenna Vesterman Henriksen, Emil Christensen, Shruti Sharma, Raheleh Salari, Himanshu Sethi, Michael Knudsen, et al. 2019. "Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients with Stages I to III Colorectal Cancer." *JAMA Oncology*. https://doi.org/10.1001/jamaoncol.2019.0528.

Rew, D A, and G D Wilson. 2000. "Cell Production Rates in Human Tissues and Tumours and

Their Significance. Part II: Clinical Data." *European Journal of Surgical Oncology : The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* 26 (4): 405–17. https://doi.org/10.1053/ejso.1999.0907.

Ryser, Marc D, Mathias Worni, Elizabeth L Turner, Jeffrey R Marks, Rick Durrett, and E Shelley Hwang. 2016. "Outcomes of Active Surveillance for Ductal Carcinoma in Situ: A Computational Risk Analysis." *Journal of the National Cancer Institute* 108 (5): djv372.

Shen, Shu Yi, Rajat Singhania, Gordon Fehringer, Ankur Chakravarthy, Michael H A Roehrl, Dianne Chadwick, Philip C Zuzarte, et al. 2018. "Sensitive Tumour Detection and Classification Using Plasma Cell-Free DNA Methylomes." *Nature* 563 (7732): 579.

Siegel, Rebecca L, Kimberly D Miller, and Ahmedin Jemal. 2020. "Cancer Statistics, 2020." *CA: A Cancer Journal for Clinicians* 70 (1): 7–30.

Song, Mingyang, Bert Vogelstein, Edward L Giovannucci, Walter C Willett, and Cristian Tomasetti. 2018. "Cancer Prevention: Molecular and Epidemiologic Consensus." *Science* 361 (6409): 1317–18.

Srivastava, Sudhir, Eugene J Koay, Alexander D Borowsky, Angelo M De Marzo, Sharmistha Ghosh, Paul D Wagner, and Barnett S Kramer. 2019. "Cancer Overdiagnosis: A Biological Challenge and Clinical Dilemma." *Nature Reviews Cancer* 19: 349–58. https://doi.org/10.1038/s41568-019-0142-8.

The National Lung Screening Trial Research Team. 2011. "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening." *New England Journal of Medicine* 365 (5): 395–409. https://doi.org/10.1056/NEJMoa1102873.

Tie, Jeanne, Yuxuan Wang, Cristian Tomasetti, Lu Li, Simeon Springer, Isaac Kinde, Natalie Silliman, et al. 2016. "Circulating Tumor DNA Analysis Detects Minimal Residual Disease and Predicts Recurrence in Patients with Stage II Colon Cancer." *Science Translational Medicine* 8 (346): 346ra92. https://doi.org/10.1126/scitranslmed.aaf6219.

Ulz, Peter, Gerhard G Thallinger, Martina Auer, Ricarda Graf, Karl Kashofer, Stephan W Jahn, Luca Abete, et al. 2016. "Inferring Expressed Genes by Whole-Genome Sequencing of Plasma DNA." *Nature Genetics* 48 (10): 1273.

Vogelstein, Bert, Nickolas Papadopoulos, Victor E Velculescu, Shibin Zhou, Luis A Diaz, and Kenneth W Kinzler. 2013. "Cancer Genome Landscapes." *Science* 339 (6127): 1546–58.

Wan, Jonathan C M, Katrin Heider, Davina Gale, Suzanne Murphy, Eyal Fisher, James Morris, Florent Mouliere, et al. 2019. "High-Sensitivity Monitoring of CtDNA by Patient-Specific Sequencing Panels and Integration of Variant Reads." *BioRxiv*, 759399. https://doi.org/10.1101/759399.

Wan, Jonathan C M, Charles Massie, Javier Garcia-Corbacho, Florent Mouliere, James D Brenton, Carlos Caldas, Simon Pacey, Richard Baird, and Nitzan Rosenfeld. 2017. "Liquid Biopsies Come of Age: Towards Implementation of Circulating Tumour DNA." *Nature Reviews Cancer* 17 (4): 223–38.

Winer-Muram, Helen T, S Gregory Jennings, Robert D Tarver, Alex M Aisen, Mark Tann, Dewey J Conces, and Cristopher A Meyer. 2002. "Volumetric Growth Rate of Stage I Lung Cancer Prior to Treatment: Serial CT Scanning." *Radiology* 223 (3): 798–805.

Wodarz, Dominik, and Natalia L Komarova. 2014. *Dynamics of Cancer: Mathematical*

Foundations of Oncology. World Scientific Publishing, Singapore.

Yokoyama, Akira, Nobuyuki Kakiuchi, Tetsuichi Yoshizato, Yasuhito Nannya, Hiromichi Suzuki, Yasuhide Takeuchi, Yusuke Shiozawa, et al. 2019. "Age-Related Remodelling of Oesophageal Epithelia by Mutated Cancer Drivers." *Nature* 565 (7739): 312–17. https://doi.org/10.1038/s41586-018-0811-x.

Acknowledgments: We thank Tibor Antal, David Cheek, and Michael Nicholson for valuable discussions and feedback. Some of the computing for this project was performed on the Stanford Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. **Funding**: This work was supported by the National Institutes of Health grants 4R00CA22999102 (J.G.R.). **Author contributions**: J.G.R. conceived and supervised the study. S.A. and J.G.R. developed the mathematical model and performed the computational analysis. S.A., D.M.K., J.J.C., J.G.R. analyzed data. All authors discussed and interpreted data. J.G.R. and S.A. wrote the original manuscript. All authors reviewed and edited the manuscript.