

## CANCER

# A mathematical model of ctDNA shedding predicts tumor detection size

Stefano Avanzini<sup>1</sup>, David M. Kurtz<sup>2</sup>, Jacob J. Chabon<sup>3,4</sup>, Everett J. Moding<sup>4,5</sup>, Sharon Seiko Hori<sup>1,6</sup>, Sanjiv Sam Gambhir<sup>1,4,6,7,8\*</sup>, Ash A. Alizadeh<sup>2,3,4</sup>, Maximilian Diehn<sup>3,4,5</sup>, Johannes G. Reiter<sup>1,4,7,9†</sup>

Early cancer detection aims to find tumors before they progress to an incurable stage. To determine the potential of circulating tumor DNA (ctDNA) for cancer detection, we developed a mathematical model of tumor evolution and ctDNA shedding to predict the size at which tumors become detectable. From 176 patients with stage I to III lung cancer, we inferred that, on average, 0.014% of a tumor cell's DNA is shed into the bloodstream per cell death. For annual screening, the model predicts median detection sizes of 2.0 to 2.3 cm representing a ~40% decrease from the current median detection size of 3.5 cm. For informed monthly cancer relapse testing, the model predicts a median detection size of 0.83 cm and suggests that treatment failure can be detected 140 days earlier than with imaging-based approaches. This mechanistic framework can help accelerate clinical trials by precomputing the most promising cancer early detection strategies.

## INTRODUCTION

Patients with early-stage cancer are more likely to be cured than patients with advanced-stage cancer (1–4). For example, the 5-year survival rate of patients with lung cancer who are diagnosed at a localized stage is 57%, while for those diagnosed with distant metastases, it is only 5% (5). Unfortunately, only 16% of lung cancers are diagnosed at a localized stage. Recently, multiple studies presented new minimally invasive approaches based on cell-free DNA (cfDNA) to detect cancer from blood samples (6–15). Most cfDNA in the bloodstream is derived from normal cells, while a small proportion is derived from tumor cells and is known as circulating tumor DNA (ctDNA). Presumably, ctDNA is shed by tumor cells undergoing apoptosis or necrosis (16–18). Small tumors would therefore be harder to detect because fewer tumor cells undergo cell death and shed ctDNA into the bloodstream.

Previous studies showed that 30 to 100% of symptomatic tumors (mostly larger than 3 cm<sup>3</sup>) can be detected from a 10- to 15-ml blood sample (8, 9, 14). However, assessing whether blood-based tests can also detect still asymptomatic tumors at sizes smaller than 3 cm<sup>3</sup> with a sufficiently high specificity to reduce cancer mortality requires elaborate clinical trials with tens of thousands of participants. While such trials are already under way, we lack the mechanistic frameworks necessary to predict the expected size of tumors that would be detected with a given sequencing approach and sampling frequency (19). Such frameworks would enable investigators

to a priori choose a sequencing and sampling strategy with the highest success probability for a given screening population. 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 nonsmokers versus 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. This mechanistic framework will help to predict the performance of ctDNA-based tumor detection approaches and thereby inform and optimize 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 (20). Lung cancer cells approximately divide with a birth rate of  $b = 0.14$  per day (21) and die with a death rate of  $d = b - r = 0.136$  per day (Fig. 1A). For now, we assume that each tumor cell releases ctDNA into the bloodstream during apoptosis with a ctDNA shedding probability of  $q_d$  per cell death. This assumption implies that the amount of ctDNA linearly correlates with tumor burden, and the slope of the linear regression has to be 1 in logarithmic space.

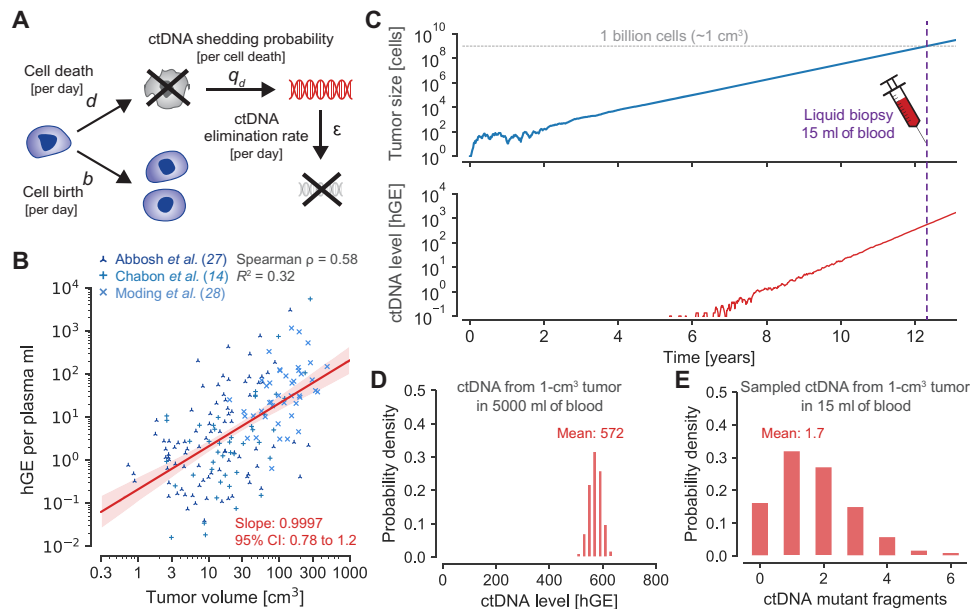
By reanalyzing ctDNA sequencing data and tumor volumes of 176 patients with stage I to III non-small cell lung cancer of three cohorts (14, 22, 23), we found that haploid genome equivalents (hGE) per plasma ml indeed correlate with tumor volume with a slope of 0.9997 [95% confidence interval (CI), 0.78 to 1.2;  $R^2 = 0.32$ ; red line in Fig. 1B; Materials and Methods]. We found similar linear regression slopes and intercepts in the separated three cohorts (fig. S1). For the combined dataset, linear regression predicted 0.21 hGE per plasma ml for 1 cm<sup>3</sup> of tumor volume (95% CI, 0.15 to 0.28; 95% prediction interval, 0.0033 to 13 hGE per plasma ml for a fixed slope of 1; fig. S2B). On the basis of these analyses, we inferred a mean shedding probability of  $q_d \approx 1.4 \times 10^{-4}$  hGE per cell death

Copyright © 2020  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim to  
original U.S. Government  
Works. Distributed  
under a Creative  
Commons Attribution  
License 4.0 (CC BY).

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

\*Deceased.

†Corresponding author. Email: johannes.reiter@stanford.edu



**Fig. 1. Evolutionary dynamics of ctDNA shed by a growing cancer.** (A) 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  $q_d$ . ctDNA is eliminated from the bloodstream with rate  $\epsilon$  per day according to the half-life time of ctDNA,  $t_{1/2} = 30$  min. (B) hGE per plasma ml correlate with tumor volume with a slope of 0.9997 in 176 patients with lung cancer ( $R^2 = 0.32$ ;  $P = 2.6 \times 10^{-16}$ ). Red shaded region depicts 95% confidence interval (CI). Linear regression predicts 0.21 hGE per plasma ml for a tumor volume of  $1 \text{ cm}^3$ , leading to a shedding probability of  $q_d = 1.4 \times 10^{-4}$  hGE per cell death (Materials and Methods). (C) A 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). 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 hGE in the entire bloodstream at the time [purple dashed line in (C)] when the tumor reaches 1 billion cells [ $\approx 1 \text{ cm}^3$ ; gray dashed line in (C)], leading to a mean of 0.21 hGE per plasma ml and a tumor fraction of 0.022% at a mean DNA concentration of 6.3 ng per plasma ml. (E) Probability distribution of ctDNA mutant fragments present in a liquid biopsy of 15 ml of blood when the tumor reaches 1 billion cells (assuming that the covered somatic heterozygous mutation is present in all cancer cells).

(95% CI,  $1.0 \times 10^{-4}$  to  $1.9 \times 10^{-4}$  for a slope of 1; fig. S2). In other words, approximately 0.014% of a cancer cell's genome is shed into the bloodstream after it undergoes apoptosis. For a ctDNA half-life time of  $t_{1/2} = 30$  min (16), we calculated a ctDNA elimination rate of  $\epsilon \approx 33$  per day. We illustrate a typical realization of this evolutionary process in Fig. 1C (movie S1). A tumor grows exponentially with a growth rate of  $r = 0.4\%$  per day and releases ctDNA into the bloodstream with a shedding rate of  $d \cdot q_d$ . At a primary tumor size of  $1 \text{ cm}^3$  [ $\sim 1$  billion cells (24)], we find, on average, 572 ctDNA hGE circulating in the bloodstream (Fig. 1D). A 15-ml blood sample contains approximately 1.7 ctDNA hGE (Fig. 1E). At a mean plasma DNA concentration of 6.3 ng per plasma ml (Materials and Methods), 0.21 ctDNA hGE per plasma ml correspond to a tumor fraction of 0.022% (assuming 6.6 pg per diploid genome). The unit of hGE can also be interpreted as the expected number of ctDNA fragments that exhibit a specific somatic heterozygous mutation. Hence, the number of ctDNA hGE 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 (8, 25).

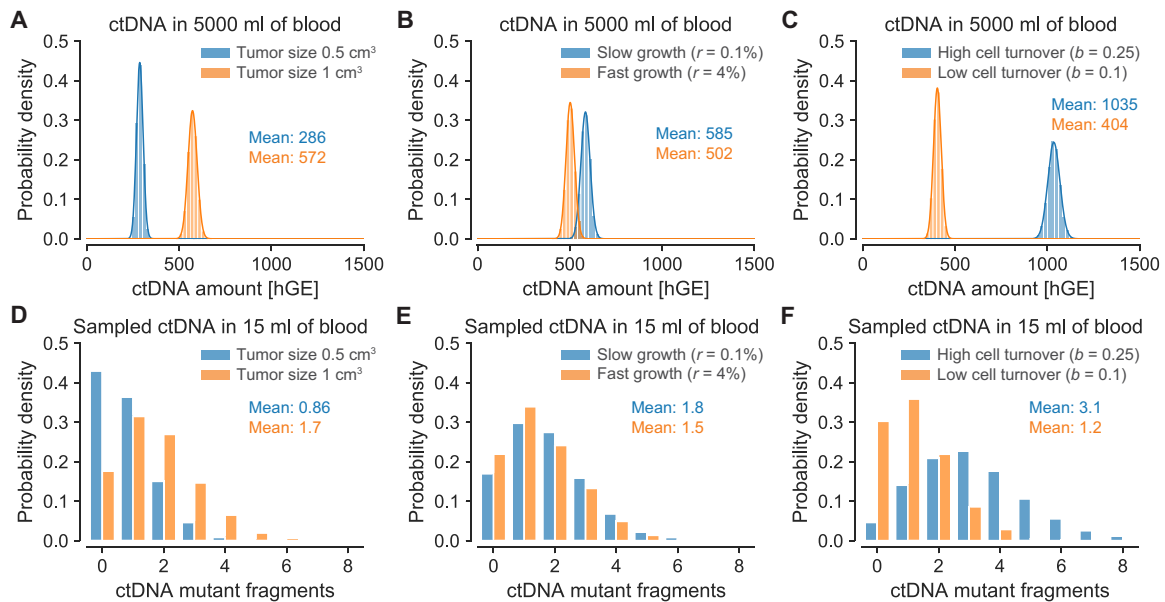
Next, we aimed to calculate the expected number of ctDNA haploid genome equivalents,  $C$ , circulating in the bloodstream for any tumor size  $M$  and derived a simple closed-form expression. The

number of ctDNA hGE circulating when the tumor reaches a size of  $M$  cells follows a Poisson distribution with a mean of

$$C = M \cdot d \cdot q_d / (\epsilon + r) \tag{1}$$

(for  $M \gg 1$  and  $d \cdot q_d \ll 1$ ;  $r = b - d$ ; note S1). For a tumor with 1 billion cells, we calculated a mean of  $C \approx 572$  ctDNA hGE, which perfectly matched the results from the exact computer simulations of the above defined branching process (Fig. 1D). Similarly, for a liquid biopsy of 15 ml of blood (0.3% of 5000 ml), we calculated a mean of 1.7 hGE by multiplying  $C$  with the fraction of the sampled blood (Fig. 1E).

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 half the number of circulating biomarkers ( $C \approx 286$  hGE; Fig. 2A). More unexpectedly, a slowly growing lung cancer ( $r = 0.1\%$ ) leads to a substantially higher number of biomarkers ( $C \approx 585$  hGE) than a faster-growing cancer ( $r = 4\%$ ) with 502 hGE at the same size of  $1 \text{ cm}^3$ , assuming that the faster growth is achieved by proportionally increased birth and decreased death rates (Fig. 2B). If instead the faster growth is achieved by a higher birth rate and an equal death rate, we find a smaller difference (584.2 versus 584.9 hGE). If cancer cells divide with a birth rate of  $b = 0.25$  (e.g., colorectal cancer cells), the higher cell turnover rate would lead to a more than twofold (1035 versus 404 hGE) increase in ctDNA compared with a



**Fig. 2. Tumor growth rate and cell turnover strongly affect the amount of ctDNA.** (A to C) ctDNA hGE present in the entire bloodstream when a lung tumor reaches a given size. Bars illustrate distribution of hGE of ctDNA based on 10,000 simulation realizations. Full lines illustrate asymptotic results (Eq. 1; note S1) and perfectly agree with simulation results (bars). (D to F) ctDNA mutant fragments present in a 15-ml blood sample. (A) Tumors with half the cells (0.5 versus 1 billion cells) lead to half the ctDNA level in the bloodstream (birth rate  $b = 0.14$  per cell per day and death rate  $d = 0.136$  per cell per day). (B) Assuming that the growth rate proportionally changes the birth and death rates, fast-growing tumors lead to a lower level of ctDNA when they reach a size of 1 billion cells ( $\approx 1$  cm<sup>3</sup>) because fewer cell deaths decrease the amount of released ctDNA (slow growth:  $b = 0.14$  and  $d = 0.139$ ; fast growth:  $b = 0.1595$  and  $d = 0.1195$ ). (C) Higher cell turnover rates lead to a higher level of ctDNA at a given tumor size (1 billion cells) compared with lower cell turnover rates because of the increased rate of cells undergoing apoptosis (if the underlying shedding probability per cell death is the same; high cell turnover:  $b = 0.25$  and  $d = 0.246$ ; low cell turnover:  $b = 0.1$  and  $d = 0.096$ ). Parameter values: ctDNA half-life time  $t_{1/2} = 30$  min; ctDNA shedding probability per cell death  $q_d = 1.4 \times 10^{-4}$  hGE.

lower cell turnover rate ( $b = 0.1$ ; e.g., breast cancer cells; assuming the same growth rate) because of the increased rate of cells undergoing apoptosis despite the same underlying shedding probability per cell death (Fig. 2C). In general, the tumor growth dynamics, the ctDNA half-life time, and the ctDNA shedding rate strongly influence ctDNA levels (fig. S3). The analytical results were validated by perfectly matching exact simulation results across all considered scenarios (full lines versus bars in the histogram of Fig. 2, A to C; fig. S4 and tables S1 and S2).

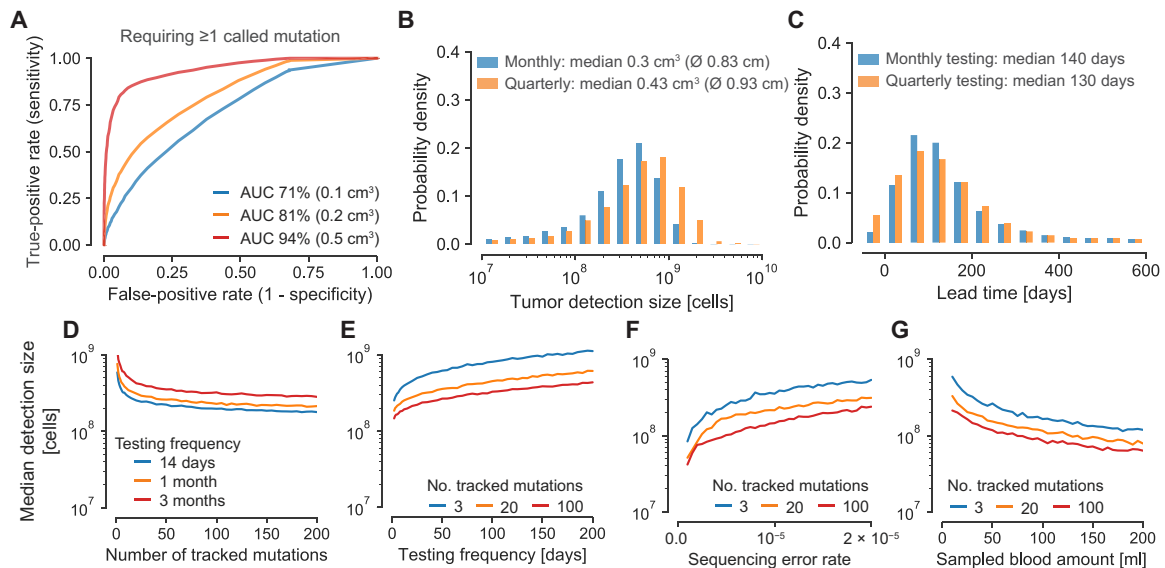
Using this mathematical framework of cancer evolution (26–30), we can predict the expected tumor detection size for an early detection test based on somatic point mutations in ctDNA (8, 9). To compute realistic tumor detection sizes, we considered various sources of biological and technical errors. Given the number of wild-type and mutant fragments, we calculated the probability that a mutation arose from sequencing errors assuming a sequencing error rate of  $10^{-5}$  per base pair (8, 25, 31). To comprehend the variation of DNA concentration in plasma samples, we reanalyzed previously measured plasma DNA concentrations from Cohen *et al.* (9). We used a Gamma distribution with a median of 5.2 ng of DNA per plasma ml ( $=788$  hGE per plasma ml) to model the variability of plasma DNA concentration (fig. S5B; Materials and Methods). This analysis also revealed that plasma DNA concentrations increased in patients with advanced-stage cancer more than expected by the ctDNA amount shed from larger tumors alone (fig. S5E).

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 (22, 25, 32–36). Requiring that at least one of these 20 tumor-specific mutations needs to be called as significantly 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 81% for tumors with  $0.2$  cm<sup>3</sup> (diameter of 0.73 cm; Fig. 3A). At a specificity of 99.5%, we observed a sensitivity of 12% for tumors with  $0.2$  cm<sup>3</sup>, assuming a sequencing efficiency of 50% [i.e., only 50% of the DNA fragments can be assessed (14); see fig. S6A for 100% sequencing efficiency]. Repeatedly applying this virtual early detection test to a relapsing lung tumor led to a median detection size of  $0.28$  cm<sup>3</sup> for monthly sampling and  $0.69$  cm<sup>3</sup> for quarterly sampling (fig. S6D). Important to note is that although the same test with a specificity of 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 an objective comparison, we adjusted the mutation calling thresholds such that 0.05 false positives are expected for both sampling frequencies



**Fig. 3. Expected tumor relapse detection size and lead time compared with current clinical relapse detection.** (A) ROC curves for tumors with 100 million cells ( $\approx 0.1 \text{ cm}^3$ ; blue line), 200 million cells ( $\approx 0.2 \text{ cm}^3$ ; orange line), or 500 million cells ( $\approx 0.5 \text{ cm}^3$ ; 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. (B to G) For better comparability, positive detection test thresholds were set such that if the test is repeated multiple times, a combined FPR of 5% is obtained over all tests per year. (B) Expected tumor detection size distributions for monthly and quarterly repeated relapse detection tests (sequencing panel covers 20 mutations).  $\emptyset$  indicates diameter of spherical tumor. (C) Expected lead time distributions compared with imaging-based approaches applied at the same frequency with a detection limit of  $1 \text{ cm}^3$  for monthly and quarterly repeated relapse detection tests (sequencing panel covers 20 clonal mutations). (D to G) Median tumor detection sizes over the number of clonal mutations covered by the sequencing panel (D), the testing frequency (E), the sequencing error rate (F), and the sampled blood amount (G). Parameter values (if not differently specified): birth rate  $b = 0.14$  per cell per day; death rate  $d = 0.13$  per cell per day; ctDNA half-life time  $t_{1/2} = 30$  min; ctDNA shedding probability per cell death  $q_d = 1.4 \times 10^{-4}$  hGE; sequencing efficiency of 50%; sequencing error rate per base pair  $10^{-3}$ ; 15 ml of blood sampled per test; DNA median concentration  $5.2 \text{ ng}$  per plasma ml. In all scenarios, at least one mutation needs to be called for a positive test. ROC, receiver operating characteristic; AUC, area under the curve.

over 12 months of relapse testing (Materials and Methods). With this adjustment, the median detection size of quarterly testing decreased to  $0.43 \text{ cm}^3$  (Fig. 3B). Monthly relapse testing still led to a 30% smaller median detection size of  $0.3 \text{ cm}^3$  (see fig. S7 for results with 100% sequencing efficiency).

To further assess ctDNA-based relapse detection, we computed the expected lead time to imaging-based relapse detection when applied at the same frequency. We conservatively assumed a radiological detection limit of exactly  $1 \text{ cm}^3$  and a specificity of 100% above that limit. The median lead time of monthly ctDNA testing compared to imaging was 140 days (Fig. 3C). Quarterly ctDNA testing yielded a similar lead time of 130 days because of our assumption that imaging is performed at the same frequency. These predicted lead times for early-stage lung cancer coincide with the reported median of  $\sim 160$  days by Chaudhuri *et al.* (37), the  $\sim 120$  days by Moding *et al.* (23), and the slightly shorter reported median of 70 days (range, 10 to 346 days) by Abbosh *et al.* (22), likely due to two required detected mutations, less frequent relapse testing, and a lower average number of clonal mutations covered by the sequencing panel (range, 3 to 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 25 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. A decreasing sequencing error rate and an increasing amount of sampled blood led to strong decreases of the expected detection sizes (Fig. 3, F and G). Moreover, we found that requiring multiple called mutations for a positive detection test can increase the sensitivity at the same specificity (fig. S8). However, the sensitivity does not increase monotonically with the required number of called mutations and strongly depends on the panel size, plasma DNA concentration, and ctDNA abundance (fig. S9; see Materials and Methods for further details).

### 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 of the genome such that many more mutations per sequenced megabase are observed than expected from the average lung cancer mutation frequency of  $\sim 10$  mutations per megabase. For example, CAncer Personalized Profiling by deep Sequencing (CAPP-Seq) (spanning  $\sim 300,000$  base pairs) and CancerSEEK (spanning  $\sim 2000$  base pairs) cover, on average, 9.1 and 1.1 mutations per early-stage lung cancer, respectively, in the TRACERx cohort (table S3) (9, 22, 25).

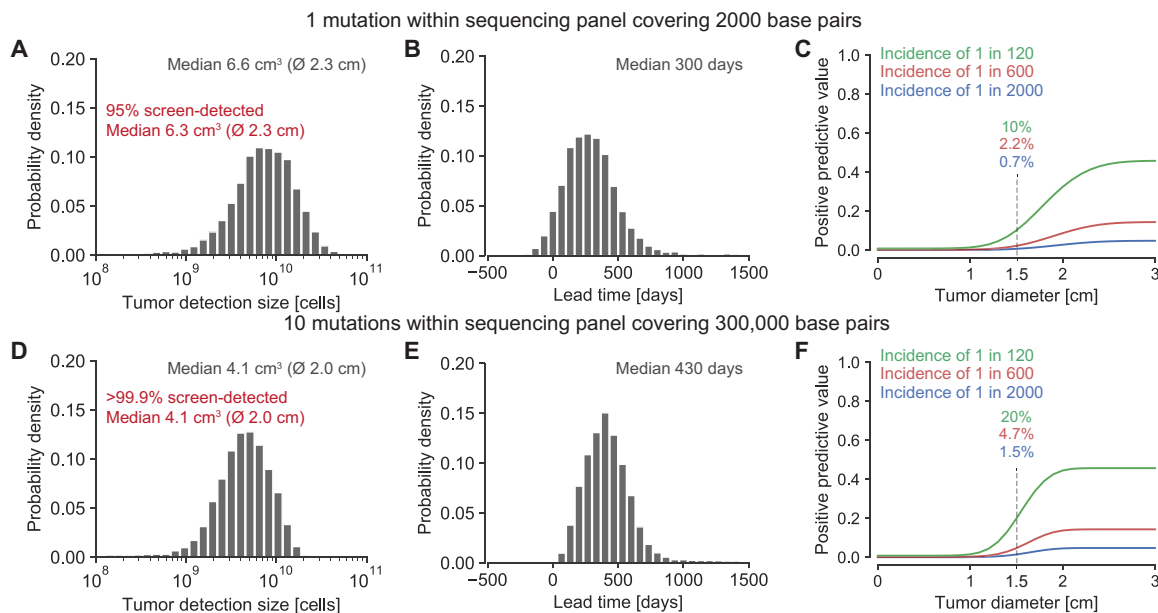
We again consider an early-stage lung tumor growing with  $r = 0.4\%$  per day. For a sequencing panel covering, on average, one

mutation per lung cancer across 2000 base pairs, we computed sensitivities of 4.3, 17, and 54% at a specificity of 99% for tumors at sizes of 1, 2, and 4 cm<sup>3</sup>, respectively. Repeating this virtual early detection test annually for a growing tumor, we obtained a median detection size of 6.6 cm<sup>3</sup> (diameter of 2.3 cm; Fig. 4A). This detection size would be 70% smaller than the current median detection size of approximately 22.5 cm<sup>3</sup> (diameter of 3.5 cm; assuming that tumors are approximately spherical) for lung cancers reported in the SEER database from 2005 to 2015 (38). Comparing the computed detection size and the SEER median size at diagnosis, we calculated a lead time to current diagnosis times of 300 days for a typical growth rate of early-stage lung cancer (Fig. 4B). For 4.6% of cancers, we observed a negative lead time. In other words, those cancers became symptomatic (i.e., reached typical diagnosis size of 22.5 cm<sup>3</sup>) before they were detected by 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 69 days), only 52% of cancers would be detected before becoming symptomatic, and their median detection size would be 8.6 cm<sup>3</sup> (diameter of 2.5 cm; fig. S10). Note that these tumors grow from 0.1 to 3.8 cm<sup>3</sup> in just 1 year and are therefore very hard to detect with a screening program.

In comparison, for a sequencing panel covering 300,000 base pairs (e.g., CAPP-Seq), we computed sensitivities of 24 and 39% at a specificity of 99% for tumors with a size of 2 cm<sup>3</sup> in never-smoking subjects and in subjects with a smoking history, respectively. We separately analyzed smokers and nonsmokers because in contrast to the above evaluated panel that focused on driver gene mutations, the expected number of mutations covered by the larger panel for

cancers of never-smoking subjects was lower than that for cancers of subjects with a smoking history (5 versus 10 mutations; table S3). For an annually repeated screening test, the expected median detection size was 5.1 cm<sup>3</sup> (diameter of 2.1 cm) for lung tumors of never-smokers and 4.1 cm<sup>3</sup> (diameter of 2.0 cm) for tumors of subjects with a smoking history (Fig. 4D). These detection sizes correspond to lead times compared to current tumor sizes at diagnosis of 360 and 430 days for tumors of never-smokers and smokers, respectively (Fig. 4E). Similarly high lead times of up to 450 days were observed in the CCGA (Circulating Cell-free Genome Atlas) study (15). In contrast to the informed detection scenario where the sensitivity increased when more than one called mutation was required for detection, in the uninformed detection scenarios, the sensitivity decreased when more than one called mutation was required for detection (fig. S11 versus fig. S9).

For the above calculations, we assumed that the detected mutations in cfDNA were unique to cancers. However, expanded clones in blood or in normal tissue and benign lesions frequently exhibit somatic mutations, which can be observed in cfDNA and hence hamper the specificity of a cancer screening test (14, 36, 39, 40). We therefore assumed that white blood cells are sequenced separately to remove somatic mutations arising due to clonal hematopoiesis. Moreover, previous studies showed that, on average, eight benign lung nodules with a diameter of ~0.4 cm exist in subjects undergoing lung cancer screening (41). Because most benign lesions are smaller than malignant tumors and because benign cells typically replicate with a lower rate than malignant cells, they are expected to release comparatively less ctDNA into the bloodstream (21). Our calculations



**Fig. 4. Expected tumor detection size and lead time distributions for screening with different sequencing panels.** (A) Tumor detection size distribution for annually repeated virtual screening tests with a 2000–base pair sequencing panel covering one somatic mutation per lung cancer. (B and E) Lead time distributions for annually repeated virtual screening tests compared with current clinical diagnosis times calculated from detection sizes in the SEER database, assuming an early-stage lung cancer growth rate of  $r = 0.4\%$  per day. (C and F) PPV (positive predictive value) for a cancer early detection test in populations with different lung cancer incidence rates. (D) Tumor detection size distribution for annually repeated virtual screening tests with a 300,000–base pair sequencing panel covering 10 somatic mutations per lung cancer of a smoker (table S3). 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  $t_{1/2} = 30$  min; ctDNA shedding probability  $q_d = 1.4 \times 10^{-4}$  hGE per cell death; sequencing error rate per base pair  $10^{-5}$ ; sequencing efficiency 50%; 15 ml of blood sampled per test; plasma normal DNA median concentration 5.2 ng per ml; 99% test specificity.

suggest that these nodules would only slightly increase the expected detection size—even if benign cells exhibit the same ctDNA shedding probability and somatic mutation load as malignant cells (fig. S12).

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 2000 (5). Because more than 98% of lung cancers occur in people above 50 (~30% of the U.S. 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 2.2% and an NPV (negative predictive value) of 99.9% for a lung tumor with a diameter of 1.5 cm using a sequencing panel of 2000 base pairs (Fig. 4C). For heavy smokers in their late 60s, the incidence rate increases to 1 in 120, and the same test would have a PPV of 10% and an NPV of 99.3% for 1.5-cm lung tumors. In comparison, a PPV of 3.8% was reported for low-dose computed tomography (CT) lung cancer screening (42).

### 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  $\lambda$  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 (10, 16). 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  $C = M \cdot \lambda / (\epsilon + r)$  (assuming  $M \gg 1$  and  $\lambda \ll 1$ ; note S1), where the effective shedding rate represents the sum of the operating ctDNA shedding processes (fig. S13 and table S1).

## DISCUSSION

Our mathematical framework provides a theoretical estimate for the performance of mutation-based ctDNA detection across clinical scenarios. For example, ctDNA becomes increasingly important for identifying actionable mutations when tumor tissue is not available. On the basis of our stochastic model, we estimate that the probability of a false negative for a particular actionable mutation clonally present in tumors with diameters of 1, 1.5, and 2 cm is 82, 44, and 9.3%, respectively (requiring a specificity of 99%). A critical parameter for these estimates is the ctDNA shedding rate per cancer cell, which can vary by multiple magnitudes among patients (Fig. 1B). The stochastic model indicates 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 (FPR) (Fig. 3). The virtual screening computations indicate that lung tumors would be detected when they reach a diameter of 2.3 cm in an annual screening program with a sequencing panel of 2000 base pairs (Fig. 4A). According to the lung cancer staging system for tumor sizes, 12% of the screen-detected cases would be classified as T1a ( $\leq 1$  cm), 21% as T1b ( $>1$  but  $\leq 2$  cm), and 55% as T1c ( $>2$  but  $\leq 3$  cm). Only 12% of the screen-detected tumors would have reached sizes for stage T2a ( $>3$  but  $\leq 4$  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 diagnoses to an earlier stage can have an enormous impact on cancer mortality (4, 43).

This study has several limitations. First, our understanding of ctDNA shedding and its variance across lung tumors and other tumor types remains limited. More studies such as those by Abbosh *et al.*, Chabon *et al.*, and Moding *et al.* correlating ctDNA levels with tumor volume and other clinicopathological features in patients with lung cancer are required to inform shedding rate inferences in other tumor types (14, 22, 23). The high concordance in both the slope and the intercept of the linear regression analysis across the three aforementioned lung cancer cohorts is reassuring (Fig. 1B and fig. S1). Other biological factors such as tumor histology and stage additionally influence ctDNA levels (14) and could be included in future models when more data become available. Second, the ctDNA shedding dynamics of precursor lesions and how their presence interferes with cancer early detection are largely unknown (fig. S12). 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 (9–15, 17). For example, the detection of differentially methylated regions or of copy number variants can be modeled with our framework (11, 44). Last, we verified our mathematical results through exact computer simulations; however, the predicted tumor detection sizes will need to be validated in large clinical studies. According to the model, every tumor eventually becomes detectable—but some will become symptomatic first (e.g., those cancers with negative lead times in Fig. 4). We note that tumor-specific ctDNA mutant fragments were detected in all TRACERx subjects, but no mutations were called because of the high specificity requirements in 40% (38 of 96) of subjects (22).

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 and undertreatment (3, 4). 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 (45, 46). Similarly, around 33% of humans harbor precursor lesions in their pancreas, but the lifetime risk of developing pancreatic cancer is 1.6% (38, 47). Hence, most precursor lesions do not progress to cancer within the lifetime of humans (39, 48). Because cancer incidence strongly depends on factors such as age, genetic predisposition, lifestyle, or exposure to mutagens (e.g., sun, smoke), 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 and biomarker shedding (19, 28, 48–51).

## MATERIALS AND METHODS

### 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 (21), 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 (20), 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. For the analysis of tumors with a high cell turnover rate (Fig. 2C), we assumed an average time between cell divisions of colorectal cancer cells of 4 days (21), resulting in a cell birth rate of  $b = 0.25$  per day. For tumors with a low cell turnover rate, we assumed an average time between cell divisions of breast cancer cells of 10 days (21), resulting in a cell birth rate of  $b = 0.1$  per day. For the analysis with coexisting benign lesions (fig. S12), we explored the effects of nodules at fixed sizes due to their doubling times of  $\geq 500$  days and assumed that cells in benign lung nodules replicate with rates of  $b_{bn} = d_{bn} = 0.07$  per day (20, 21).

To estimate the ctDNA shedding rate per cancer cell, we reanalyzed data from 176 patients with stage I to III non-small cell lung cancer studied in Chabon *et al.* (14), Abbosh *et al.* (22), and Moding *et al.* (23). In Chabon *et al.*, metabolic tumor volumes were computed from whole-body  $^{18}\text{F}$ -FDG PET (positron emission tomography)-CT scans in 81 subjects. For 46 of these subjects, tumor-specific ctDNA mutations were reported. Reanalyzed gross tumor volume from Abbosh *et al.* and Moding *et al.* was estimated from preoperative CT scans. hGE per plasma ml was calculated from the mean VAF times the plasma cfDNA concentration in a subject divided by 0.0033 ng (weight of haploid human genome). Estimates can be found in Supplementary Table 3 of Chabon *et al.* and Supplementary Table 2 of Moding *et al.* For the data in Abbosh *et al.*, the mean VAF of somatic mutations was calculated across all mutations (including those that were not called in the liquid biopsy) that were identified as clonal in the primary tumor [in subject CRUK0053, all mutations were considered because no clonality information was available; see Supplementary Table 5 of (22)].

To infer the shedding rate  $\lambda$ , we used the predicted 0.21 hGE per plasma ml for a tumor with a volume of  $1 \text{ cm}^3$  (assuming 1 billion cells per  $\text{cm}^3$ ; Fig. 1B) and set up the following equilibrium equation of  $\lambda \cdot 10^9 \text{ cells} = 577.5 \cdot \epsilon$ , assuming 577.5 hGE ( $= 0.21 \cdot 5000 \cdot 55\%$ ) in 5000 ml of blood with a plasma concentration of 55%. We used previous estimates of the ctDNA half-life time of  $t_{1/2} = 30$  min and explored the effects of half-life times from 20 to 120 min (fig. S3) (16). We calculated the ctDNA elimination rate as  $\epsilon = \ln 2/t_{1/2} \cdot 60 \cdot 24 \approx 33.3$  fragments per day. We found a shedding rate of  $\lambda = 1.9 \cdot 10^{-5}$  hGE per cell per day, leading to a shedding probability of  $q_d = \lambda/d = 1.4 \times 10^{-4}$  hGE 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  $\lambda_{bn} = d_{bn} \cdot q_d \approx 9.8 \cdot 10^{-6}$  hGE per cell per day.

We investigated how a distribution of values for the shedding probability  $q_d$  would influence our results (fig. S2). Normality tests performed on the residuals of the linear regression between ctDNA levels and tumor volume (Fig. 1B) were consistent with a normal distribution and supported the assumption of a slope 1 for the linear correlation in log-log space. By fixing the slope to 1, we inferred the 95% CI for the amount of ctDNA shed by a tumor of  $1 \text{ cm}^3$  per plasma ml (fig. S2B). Because the CI is symmetric in logarithmic scale with respect to the predicted value of 0.21 hGE per plasma ml, we modeled the variability of this parameter by fitting a normal distribution in log space to match the estimated 95% CI (fig. S2C). Through our equilibrium equation and estimated death rate of cancer cells, we converted this normal probability distribution into a lognormal density for the shedding probability  $q_d$  in linear space [mean,  $1.4 \times 10^{-4}$  hGE per cell death; SD (standard deviation),  $2.2 \times 10^{-5}$ ].

The average amount of ctDNA mutant fragments found in a 15-ml blood sample when the shedding probability follows a lognormal distribution is almost the same than when the shedding probability is fixed to the mean of the lognormal distribution (1.74 hGE versus 1.72 hGE; fig. S2E). To facilitate the analysis and interpretation of ctDNA shedding for cancer early detection, we assumed a fixed value of  $q_d$ .

### In silico sampling and sequencing of plasma DNA

To compare different screening strategies, we used virtual sampling and sequencing of DNA. For each liquid biopsy, ctDNA fragments in 15 ml of blood were sampled from 5000 ml of blood according to a binomial distribution. For the virtual detection tests in Figs. 3 and 4 and figs. S6 to S12, plasma cfDNA concentrations were sampled from a Gamma distribution (mean and median of 6.3 and 5.2 ng per plasma ml, respectively) as illustrated in fig. S5B. We assumed that 50% of cfDNA fragments are assessed by sequencing (except in figs. S6, A to C, and S7). To calculate the cfDNA genome equivalents, we assumed a weight of 6.6 pg per diploid genome (17). The sequencing error rate per base pair was set to  $e_{\text{seq}} = 10^{-5}$  (8, 25, 31). For each mutation covered by the sequencing panel, we account for sequencing errors by calculating a  $P$  value as  $1 - \sum_{i=0}^{k-1} \binom{n}{i} \cdot e_{\text{seq}}^i \cdot (1 - e_{\text{seq}})^{n-i}$ , 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.

To optimize the performance of ctDNA-based cancer detection, we calculated the sensitivity and specificity when more than one called mutation is required for detection. These calculations show a non-monotonically changing sensitivity for different numbers of called mutation thresholds to classify a test as positive (fig. S9C). This counterintuitive behavior is ascribable to the mutation calling  $P$  value threshold, which we adapt to keep the FPR approximately constant to objectively compare different testing approaches (fig. S9, A and B). The estimated sensitivity for a fixed plasma normal DNA concentration and a fixed ctDNA level changes according to a reverse sawtooth wave for an FPR of  $\leq 1\%$ , a sequencing error rate of  $10^{-5}$ , and a sequencing panel size of 20 where all mutations are clonally present in a relapsing tumor (fig. S9D). The sensitivity jumps occur at different numbers of required mutations for detection for the different plasma DNA concentrations. For example, for a plasma DNA concentration of 12.4 ng/ml (yellow curve in fig. S9D), we observe two jumps: (i) when instead of 3, only 2 mutant fragments become sufficient to reach a  $P$  value below the threshold for the desired FPR, and (ii) when instead of 2, only 1 mutant fragment becomes sufficient to reach a  $P$  value below the threshold for the given FPR. When the whole distribution of plasma DNA concentration is taken into account (fig. S5B), the sensitivity varies more smoothly but still increases locally when the number of mutant fragments required jumps downward (fig. S9E). These local maxima are also affected by the ctDNA amount (fig. S9F). Assuming 50 ctDNA hGE in the bloodstream, the sensitivity reaches a single maximum value (at 4 mutations required for detection) before decreasing to almost zero (blue curve). At a value of 100 ctDNA hGE, the sensitivity shows two local maxima, which shift as the amount of ctDNA increases. For instance, peaks are at 2 and 6 called mutations required for detection with 100 ctDNA hGE and at 3 and 9 called mutations with 300 ctDNA hGE. With 600 ctDNA hGE or more, the sensitivity saturates and stays constant at almost 100% before starting to decrease monotonically. These results can then be used to study sensitivity as

a function of tumor sizes (and hence of varying ctDNA levels). For an FPR of 1% and tumor diameters between roughly 0.73 cm ( $2 \times 10^8$  cells) and 1.0 cm ( $6 \times 10^8$  cells), 8 called mutations required for detection led to a higher sensitivity than 2, 4, and 1 mutations (in decreasing order; fig. S9G).

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/6/50/eabc4308/DC1>

## REFERENCES AND NOTES

- R. Etzioni, N. Urban, S. Ramsey, M. McIntosh, S. Schwartz, B. Reid, J. Radich, G. Anderson, L. Hartwell, Early detection: The case for early detection. *Nat. Rev. Cancer* **3**, 243–252 (2003).
- M. Song, B. Vogelstein, E. L. Giovannucci, W. C. Willett, C. Tomasetti, Cancer prevention: Molecular and epidemiologic consensus. *Science* **361**, 1317–1318 (2018).
- S. Srivastava, E. J. Koay, A. D. Borowsky, A. M. De Marzo, S. Ghosh, P. D. Wagner, B. S. Kramer, Cancer overdiagnosis: A biological challenge and clinical dilemma. *Nat. Rev. Cancer* **19**, 349–358 (2019).
- A. K. Mattox, C. Bettgowda, S. Zhou, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, Applications of liquid biopsies for cancer. *Sci. Transl. Med.* **11**, eaay1984 (2019).
- R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 7–30 (2020).
- C. Bettgowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Lubner, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, I.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. Max Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
- A. M. Newman, S. V. Bratman, J. To, J. F. Wynne, N. C. W. Eclow, L. A. Modlin, C. L. Liu, J. W. Neal, H. A. Wakelee, R. E. Merritt, J. B. Shrager, B. W. Loo Jr., A. A. Alizadeh, M. Diehn, An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).
- J. Phallen, M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, V. Anagnostou, J. Fiksel, S. Cristiano, E. Papp, S. Speir, T. Reinert, M.-B. W. Orntoft, B. D. Woodward, D. Murphy, S. Parpart-Li, D. Riley, M. Nesselbush, N. Sengamaly, A. Georgiadis, Q. K. Li, M. R. Madsen, F. V. Mortensen, J. Huiskens, C. Punt, N. van Grieken, R. Fijneman, G. Meijer, H. Husain, R. B. Scharpf, L. A. Diaz Jr., S. Jones, S. Angiuoli, T. Orntoft, H. J. Nielsen, C. L. Andersen, V. E. Velculescu, Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* **9**, eaan2415 (2017).
- J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. D. Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Brown, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettgowda, L. A. Diaz Jr., C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **357**, 926–930 (2018).
- F. Mouliere, D. Chandrananda, A. M. Piskorz, E. K. Moore, J. Morris, L. B. Ahlborn, R. Mair, T. Goranova, F. Marass, K. Heider, J. C. M. Wan, A. Supernat, I. Hudecova, I. Gounaris, S. Ros, M. Jimenez-Linan, J. Garcia-Corbacho, K. Patel, O. Østrup, S. Murphy, M. D. Eldridge, D. Gale, G. D. Stewart, J. Burge, W. N. Cooper, M. S. van der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **10**, eaat4921 (2018).
- S. Y. Shen, R. Singhanian, G. Fehringer, A. Chakravarthy, M. H. A. Roehrl, D. Chadwick, P. C. Zuzarte, A. Borgida, T. T. Wang, T. Li, O. Kis, Z. Zhao, A. Spreafico, T. d. S. Medina, Y. Wang, D. Roulois, I. Ettayebi, Z. Chen, S. Chow, T. Murphy, A. Arruda, G. M. O’Kane, J. Liu, M. Mansour, J. D. McPherson, C. O’Brien, N. Leigh, P. L. Bedard, N. Fleshner, G. Liu, M. D. Minden, S. Gallinger, A. Goldenberg, T. J. Pugh, M. M. Hoffman, S. V. Bratman, R. J. Hung, D. D. de Carvalho, Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579–583 (2018).
- S. Cristiano, A. Leal, J. Phallen, J. Fiksel, V. Adleff, D. C. Bruhm, S. Ø. Jensen, J. E. Medina, C. Hruban, J. R. White, D. N. Palsgrove, N. Niknafs, V. Anagnostou, P. Forde, J. Naidoo, K. Marrone, J. Brahmer, B. D. Woodward, H. Husain, K. L. van Rooijen, M. B. W. Orntoft, A. H. Madsen, C. J. H. van de Velde, M. Verheij, A. Cats, C. J. A. Punt, G. R. Vink, N. C. T. van Grieken, M. Koopman, R. J. A. Fijneman, J. S. Johansen, H. J. Nielsen, G. A. Meijer, C. L. Andersen, R. B. Scharpf, V. E. Velculescu, Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389 (2019).
- P. Ulz, G. G. Thallinger, M. Auer, R. Graf, K. Kashofer, S. W. Jahn, L. Abete, G. Pristauer, E. Petru, J. B. Geigl, E. Heitzer, M. R. Speicher, Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48**, 1273–1278 (2016).
- J. J. Chabon, E. G. Hamilton, D. M. Kurtz, M. S. Esfahani, E. J. Moding, H. Stehr, J. Schroers-Martin, B. Y. Nabet, B. Chen, A. A. Chaudhuri, C. L. Liu, A. B. Hui, M. C. Jin, T. D. Azad, D. Almanza, Y. J. Jeon, M. C. Nesselbush, L. Co Ting Keh, R. F. Bonilla, C. H. Yoo, R. B. Ko, E. L. Chen, D. J. Merriott, P. P. Massion, A. S. Mansfield, J. Jen, H. Z. Ren, S. H. Lin, C. L. Costantino, R. Burr, R. Tibshirani, S. S. Gambhir, G. J. Berry, K. C. Jensen, R. B. West, J. W. Neal, H. A. Wakelee, B. W. Loo Jr., C. A. Kunder, A. N. Leung, N. S. Lui, M. F. Berry, J. B. Shrager, V. S. Nair, D. A. Haber, L. V. Sequist, A. A. Alizadeh, M. Diehn, Integrating genomic features for non-invasive early lung cancer detection. *Nature* **580**, 245–251 (2020).
- M. C. Liu, G. R. Oxnard, E. A. Klein, C. Swanton, M. V. Seiden; CCGA Consortium; STRIVE investigators; GRAIL Inc.; Advisors, Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* **31**, 745–759 (2020).
- J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
- E. Heitzer, I. S. Haque, C. E. S. Roberts, M. R. Speicher, Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20**, 71–88 (2019).
- H. Schwarzenbach, D. S. B. Hoon, K. Pantel, Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437 (2011).
- N. Pashayan, P. D. P. Pharoah, The challenge of early detection in cancer. *Science* **368**, 589–590 (2020).
- H. T. Winer-Muram, S. G. Jennings, R. D. Tarver, A. M. Aisen, M. Tann, D. J. Conces, C. A. Broyer, Volumetric growth rate of stage I lung cancer prior to treatment: Serial CT scanning. *Radiology* **223**, 798–805 (2002).
- D. A. Rew, G. D. Wilson, Cell production rates in human tissues and tumours and their significance. Part II: clinical data. *Eur. J. Surg. Oncol.* **26**, 405–417 (2000).
- C. Abbosh, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. M. Granahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. A. Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. M. Quinn, P. A. Crossbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. Jimmy Liu, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann; The TRACERx consortium; The PEACE consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
- E. J. Moding, Y. Liu, B. Y. Nabet, J. J. Chabon, A. A. Chaudhuri, A. B. Hui, R. F. Bonilla, R. B. Ko, C. H. Yoo, L. Gojenola, C. D. Jones, J. He, Y. Qiao, T. Xu, J. V. Heymach, A. Tsao, Z. Liao, D. R. Gomez, M. Das, S. K. Padda, K. J. Ramchandran, J. W. Neal, H. A. Wakelee, B. W. Loo Jr., S. H. Lin, A. A. Alizadeh, M. Diehn, Circulating tumor DNA dynamics predict benefit from consolidation immunotherapy in locally advanced non-small-cell lung cancer. *Nat. Cancer* **1**, 176–183 (2020).
- S. S. Hori, S. S. Gambhir, Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. *Sci. Transl. Med.* **3**, 109ra116 (2011).
- A. M. Newman, A. F. Lovejoy, D. M. Klass, D. M. Kurtz, J. J. Chabon, F. Scherer, H. Stehr, C. L. Liu, S. V. Bratman, C. Say, L. Zhou, J. N. Carter, R. B. West, G. W. Sledge Jr., J. B. Shrager, B. W. Loo Jr., J. W. Neal, H. A. Wakelee, M. Diehn, A. A. Alizadeh, Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
- R. Durrett, *Branching Process Models of Cancer* (Springer, 2015).
- D. Wodarz, N. L. Komarova, *Dynamics of Cancer: Mathematical Foundations of Oncology* (World Scientific Publishing, Singapore, 2014).
- J. G. Reiter, A. P. Makohon-Moore, J. M. Gerold, A. Heyde, M. A. Attiyeh, Z. A. Kohutek, C. J. Tokheim, A. Brown, R. M. DeBlasio, J. Niyazov, A. Zucker, R. Karchin, K. W. Kinzler, C. A. Iacobuzio-Donahue, B. Vogelstein, M. A. Nowak, Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033–1037 (2018).
- N. Beerenwinkel, R. F. Schwarz, M. Gerstung, F. Markowetz, Cancer evolution: Mathematical models and computational inference. *Syst. Biol.* **64**, e1–e25 (2015).
- P. M. Altrock, L. L. Liu, F. Michor, The mathematics of cancer: Integrating quantitative models. *Nat. Rev. Cancer* **15**, 730–745 (2015).
- J. C. M. Wan, K. Heider, D. Gale, S. Murphy, E. Fisher, F. Mouliere, A. Ruiz-Valdepenas, A. Santonja, J. Morris, D. Chandrananda, A. Marshall, A. B. Gill, P. Y. Chan, E. Barker, G. Young, W. N. Cooper, I. Hudecova, F. Marass, R. Mair, K. M. Brindle, G. D. Stewart,



- J. E. Abraham, C. Caldas, D. M. Rassl, R. C. Rintoul, C. Alifrangis, M. R. Middleton, F. A. Gallagher, C. Parkinson, A. Durrani, U. M. Dermott, C. G. Smith, C. Massie, P. G. Corrie, N. Rosenfeld, ctDNA monitoring using patient-specific sequencing and integration of variant reads. *Sci. Transl. Med.* **12**, eaaz8084 (2020).
32. B. R. McDonald, T. Contente-Cuomo, S.-J. Sammut, A. Odenheimer-Bergman, B. Ernst, N. Perdigones, S.-F. Chin, M. Farooq, R. Mejia, P. A. Cronin, K. S. Anderson, H. E. Kosiorek, D. W. Northfelt, A. E. McCullough, B. K. Patel, J. N. Weitzel, T. P. Slavina, C. Caldas, B. A. Pockaj, M. Murtaza, Personalized circulating tumor DNA analysis to detect residual disease after neoadjuvant therapy in breast cancer. *Sci. Transl. Med.* **11**, eaax7392 (2019).
33. J. Tie, Y. Wang, C. Tomasetti, L. Li, S. Springer, I. Kinde, N. Silliman, M. Tacey, H.-L. Wong, M. Christie, S. Kosmider, I. Skinner, R. Wong, M. Steel, B. Tran, J. Desai, I. Jones, A. Haydon, T. Hayes, T. J. Price, R. L. Strausberg, L. A. Diaz Jr., N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **8**, 346ra92 (2016).
34. T. Reinert, T. V. Henriksen, E. Christensen, S. Sharma, R. Salari, H. Sethi, M. Knudsen, I. Nordentoft, H.-T. Wu, A. S. Tin, M. Heilskov Rasmussen, S. Vang, S. Shchegrova, A. Frydendahl Boll Johansen, R. Srinivasan, Z. Assaf, M. Balcioglu, A. Olson, S. Dashner, D. Hafez, S. Navarro, S. Goel, M. Rabinowitz, P. Billings, S. Sigurjonsson, L. Dyrskjot, R. Swenerton, A. Aleshin, S. Laurberg, A. Husted Madsen, A. S. Kannerup, K. Stribolt, S. Palmelund Krag, L. H. Iversen, K. Gotschalck Sunesen, C. H. J. Lin, B. G. Zimmermann, C. Lindbjerg Andersen, Analysis of plasma cell-free DNA by ultradeep sequencing in patients with stages I to III colorectal cancer. *JAMA Oncol.* **5**, 1124–1131 (2019).
35. K. H. Khan, D. Cunningham, B. Werner, G. Vlachogiannis, I. Spiteri, T. Heide, J. F. Mateos, A. Vatsioui, A. Lampis, M. D. Damavandi, H. Lote, I. S. Huntingford, S. Hedayat, I. Chau, N. Tunariu, G. Menestrasi, F. Trevisani, S. Rao, G. Anandappa, D. Watkins, N. Starling, J. Thomas, C. Peckitt, N. Khan, M. Khan, M. Ruggie, R. Begum, B. Hezelova, A. Bryant, T. Jones, P. Proszek, M. Fassan, J. C. Hahne, M. Hubank, C. Braconi, A. Sottoriva, N. Valeri, 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 Discov.* **8**, 1270–1285 (2018).
36. D. W. Cescon, S. V. Bratman, S. M. Chan, L. L. Siu, Circulating tumor DNA and liquid biopsy in oncology. *Nat. Cancer.* **1**, 276–290 (2020).
37. A. A. Chaudhuri, J. J. Chabon, A. F. Lovejoy, A. M. Newman, H. Stehr, T. D. Azad, M. S. Khodadoust, M. S. Esfahani, C. L. Liu, L. Zhou, F. Scherer, D. M. Kurtz, C. Say, J. N. Carter, D. J. Merriott, J. C. Dudley, M. S. Binkley, L. Modlin, S. K. Padda, M. F. Gensheimer, R. B. West, J. B. Shrager, J. W. Neal, H. A. Wakelee, B. W. Loo Jr., A. A. Alizadeh, M. Diehn, Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov.* **7**, 1394–1403 (2017).
38. NCI Seer, Surveillance, Epidemiology, and End Results (SEER) Research Data 1973–2015–ASCII Text Data (2018); <https://seer.cancer.gov>.
39. A. P. Makohon-Moore, K. Matsukuma, M. Zhang, J. G. Reiter, J. M. Gerold, Y. Jiao, L. Sikkema, M. A. S. Yachida, C. Sandone, R. H. Hruban, D. S. Klimstra, N. Papadopoulos, M. A. Nowak, K. W. Kinzler, B. Vogelstein, C. A. Iacobuzio-Donahue, Precancerous neoplastic cells can move through the pancreatic ductal system. *Nature* **561**, 201–205 (2018).
40. I. Martincorena, P. J. Campbell, Somatic mutation in cancer and normal cells. *Science* **349**, 1483–1489 (2015).
41. A. McWilliams, M. C. Tammemagi, J. R. Mayo, H. Roberts, G. Liu, K. Soghrati, K. Yasufuku, S. Martel, F. Laberge, M. Gingras, S. Atkar-Khattra, C. D. Berg, K. Evans, R. Finley, J. Yee, J. English, P. Nasute, J. Goffin, S. Puksa, L. Stewart, S. Tsai, M. R. Johnston, D. Manos, G. Nicholas, G. D. Goss, J. M. Seely, K. Amjadi, A. Tremblay, P. Burrows, P. MacEachern, R. Bhatia, M. S. Tsao, S. Lam, Probability of cancer in pulmonary nodules detected on first screening CT. *N. Engl. J. Med.* **369**, 910–919 (2013).
42. The National Lung Screening Trial Research Team, Results of initial low-dose computed tomographic screening for lung cancer. *N. Engl. J. Med.* **368**, 1980–1991 (2013).
43. C. A. Clarke, E. Hubbell, A. W. Kurian, G. A. Colditz, A.-R. Hartman, S. L. Gomez, Projected reductions in absolute cancer-related deaths from diagnosing cancers before metastasis, 2006–2015. *Cancer Epidemiol. Biomarkers. Prev.* **29**, 895–902 (2020).
44. C. Douville, J. D. Cohen, J. Ptak, M. Popoli, J. Schaefer, N. Silliman, L. Dobbyn, R. E. Schoen, J. Tie, P. Gibbs, M. Goggins, C. L. Wolfgang, T.-L. Wang, I. M. Shih, R. Karchin, A. M. Lennon, R. H. Hruban, C. Tomasetti, C. Bettgeowda, K. W. Kinzler, N. Papadopoulos, B. Vogelstein, Assessing aneuploidy with repetitive element sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4858–4863 (2020).
45. The National Lung Screening Trial Research Team, Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* **365**, 395–409 (2011).
46. H. J. de Koning, C. M. van der Aalst, P. A. de Jong, E. T. Scholten, K. Nackaerts, M. A. Heuvelmans, J.-W. J. Lammers, C. Weenink, U. Yousaf-Khan, N. Horeweg, S. van 't Westeinde, M. Prokop, W. P. Mali, F. A. A. Mohamed Hoesin, P. M. A. van Ooijen, J. G. J. V. Aerts, M. A. den Bakker, E. Thunnissen, J. Verschakelen, R. Vliegthart, J. E. Walter, K. ten Haaf, H. J. M. Groen, M. Oudkerk, Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N. Engl. J. Med.* **382**, 503–513 (2020).
47. A. P. Makohon-Moore, C. A. Iacobuzio-Donahue, Pancreatic cancer biology and genetics from an evolutionary perspective. *Nat. Rev. Cancer* **16**, 553–565 (2016).
48. B. M. Lang, J. Kuipers, B. Misselwitz, N. Beerenwinkel, Predicting colorectal cancer risk from adenoma detection via a two-type branching process model. *PLOS Comput. Biol.* **16**, e1007552 (2020).
49. H. J. de Koning, R. Meza, S. K. Plevritis, K. ten Haaf, V. N. Munshi, J. Jeon, S. A. Erdogan, C. Y. Kong, S. S. Han, J. van Rosmalen, S. E. Choi, P. F. Pinsky, A. B. de Gonzalez, C. D. Berg, W. C. Black, M. C. Tammemagi, W. D. Hazelton, E. J. Feuer, P. M. McMahon, Benefits and harms of computed tomography lung cancer screening strategies: A comparative modeling study for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* **160**, 311–320 (2014).
50. K. Curtius, A. Dewanji, W. D. Hazelton, J. H. Rubenstein, E. G. Luebeck, Optimal timing for cancer screening and adaptive surveillance using mathematical modeling. *bioRxiv*, 2020.02.11.927475 (2020).
51. S. S. Hori, A. M. Lutz, R. Paulmurugan, S. S. Gambhir, A model-based personalized cancer screening strategy for detecting early-stage tumors using blood-borne biomarkers. *Cancer Res.* **77**, 2570–2584 (2017).
52. K. B. Athreya, P. E. Ney, *Branching Processes* (Springer-Verlag, Berlin Heidelberg, 1972).
53. L. Allen, *An Introduction to Stochastic Processes with Applications to Biology* (Boca Raton, FL: Chapman & Hall/CRC, ed. 2, 2011).
54. V. G. Kulkarni, *Modeling and Analysis of Stochastic Systems* (Chapman and Hall/CRC, 2016).
55. D. Cheek, T. Antal, Mutation frequencies in a birth-death branching process. *Ann. Appl. Probab.* **28**, 3922–3947 (2018).
56. D. J. Daley, D. Vere-Jones, *Introduction to the Theory of Point Processes* (Springer New York, 2006).
57. T. Antal, P. L. Krapivsky, Exact solution of a two-type branching process: Clone size distribution in cell division kinetics. *J. Stat. Mech. Theory Exp.* **2010**, P07028 (2010).
58. T. Antal, P. L. Krapivsky, Exact solution of a two-type branching process: Models of tumor progression. *J. Stat. Mech. Theory Exp.* **08**, P08018 (2011).
59. A. Ronveaux, *Heun's Differential Equations* (Oxford Univ. Press, 1995).
60. O. V. Motygin, in *2018 Days on Diffraction (DD)* (IEEE, 2018), pp. 223–229.
61. M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Applied mathematics series, Dover Publications, 1964).
62. G. L. Choudhury, D. M. Lucantoni, W. Whitt, Multidimensional transform inversion with applications to the transient M/G/1 queue. *Ann. Appl. Probab.* **4**, 719–740 (1994).

**Acknowledgments:** We thank T. Antal, D. Cheek, and M. Nicholson for valuable discussions and feedback. Some of the computations for this project were 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 NIH grant 4R00CA22999102 (J.G.R.) and the Department of Radiology in the School of Medicine at Stanford University. **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., E.J.M., A.A.A., M.D., and J.G.R. analyzed the data. All authors discussed and interpreted the data. J.G.R. and S.A. wrote the original manuscript. All authors reviewed and edited the manuscript. **Competing interests:** S.A., J.J.C., S.S.H., S.S.G., and J.G.R. declare that they have no competing interests. D.M.K. is a consultant for Roche Molecular Diagnostics and has ownership equity in Foresight Diagnostics. D.M.K. has patents pending related to methods for analysis of cell-free nucleic acids and methods for treatment selection based on statistical frameworks of clinical outcome. E.J.M. has served as a paid consultant for DeciBio. A.A.A. and M.D. are co-inventors on patent applications related to CAPP-Seq. A.A.A. has equity in Forty Seven and CiberMed and has served as a consultant for Roche, Genentech, Chugai Pharma, and Pharmacyclics. M.D. has equity in CiberMed, has received research funding from Varian Medical Systems, and has served as a paid consultant for Roche, AstraZeneca, BioNTech, Illumina, and Reflexion and as an unpaid consultant for Genentech. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data associated with this study have been generated using the presented analytical results or follow from simulating the described branching process. The code for the calculations and simulations is available as Python package at <https://github.com/reiterlab/ctdna>. Data to estimate the ctDNA shedding rate were provided in the original studies (14, 22, 23). Additional data related to this paper may be requested from the authors.

Submitted 24 April 2020

Accepted 29 October 2020

Published 11 December 2020

10.1126/sciadv.abc4308

**Citation:** S. Avanzini, D. M. Kurtz, J. J. Chabon, E. J. Moding, S. S. Hori, S. S. Gambhir, A. A. Alizadeh, M. Diehn, J. G. Reiter, A mathematical model of ctDNA shedding predicts tumor detection size. *Sci. Adv.* **6**, eabc4308 (2020).