RATIONAL CANCER DIAGNOSTICS USING RARE MUTATION DETECTION TECHNOLOGY WITH MASSIVELY PARALLEL SEQUENCING

by
Isaac Kinde

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March 2015

Abstract

Worldwide, cancers remain a leading cause of death. The judicious use of cancer diagnostics – broadly defined as tests for cancer – has great potential to reduce disease morbidity and mortality. Impeding this potential is the difficulty of creating effective new tests, as the techniques successful for one type of cancer frequently cannot be generalized to another. Although the ability to detect cancer-specific DNA mutations at the low levels commonly encountered in clinical specimens would yield a promising, broadly applicable diagnostic strategy, existing technologies have been unacceptably limited in throughput or accuracy. Here we describe the development and application of a scalable, generalizable DNA sequence-based technology for the reliable detection of mutations. By drastically reducing artifacts introduced through sample preparation and massively parallel sequencing, rare mutations arising from cancer cells – when present – can be confidently discriminated from a large excess of non-mutant DNA. The technology can be directed to virtually any genomic region, affording rational test design. When applied to routinely collected Pap specimens, our approach detected cancer-specific mutations in 41% (9 of 22) and 100% (24 of 24) of women harboring various stages of ovarian and endometrial cancers, respectively. Our approach was highly specific, as no false positives were detected in a cohort of Pap specimens collected from women without gynecologic cancer. We also demonstrate how the urine of patients with urothelial carcinoma can be utilized to predict disease recurrences. Eighty-eight percent (7 of 8) of patients with a detectable mutation had recurrences while none were detected in the six patients without recurrent disease (P <0.001). Finally we present data suggesting that a wide range of cancers shed mutant DNA into blood and that these mutations are sensitive and specific markers for disease. Taken together, our results demonstrate the potential and feasibility of improved diagnostics for several cancers using a

variety of clinical specimens obtainable in a minimally invasive fashion. Larger studies are underway as a prelude to implementing these tests in the clinic – a critical step in addressing the many unmet clinical needs of patients with cancer.

Advisor: Bert Vogelstein, M.D.

Reader: Nickolas Papadopoulos, Ph.D.

Acknowledgments

I am truly thankful for the extraordinary mentorship I received as a graduate student in the Ludwig Center for Cancer Genetics and Therapeutics at the Sidney Kimmel Cancer Center of Johns Hopkins University School of Medicine. The lab is relatively unique in that a handful of principal investigators work together to advance science and train scientists. The insights that Luis Diaz, M.D., a practicing oncologist, imported from taking care of cancer patients brought realism to my thinking. Questioning by Shibin Zhou, M.D., Ph.D., made my thinking sharper. My most important experiments were performed on massively parallel sequencing instruments, for which I relied on the deep expertise that Nickolas Papadopoulos, Ph.D. has gathered from both industry and academia. When it came time to analyze the resulting data, the bioinformatics expertise of Ken Kinzler, Ph.D., who determined the sequences of some of the first cancer genomes, became invaluable for my own analyses. Over the years I worked most closely with Bert Vogelstein, M.D., who somehow managed to know everything about everything. His knowledge was only matched by his dedication – to progressing the lab's research, to his trainees, to his colleagues, and to his family. Considering the remarkable trainees and staff they have assembled, I cannot imagine what learning from a better group of mentors and colleagues would be like.

Johns Hopkins is a big place, but the training programs here helped me navigate it. For this I would like to thank my graduate program, the Program in Cellular and Molecular Medicine, directed sequentially by Pierre Coulombe, Ph.D, then Rajini Rao, Ph.D., during my time as a student. I am particularly thankful for Colleen Graham and Leslie Lichter, who guided me through the many important but nonobvious requirements of completing graduate school. The Medical Scientist Training Program – co-directed by Robert Siliciano, M.D., Ph.D., and

Andrea Cox, M.D., Ph.D. – was the glue that connected my experiences in graduate school with medical school. Kudos goes to Sharon Welling, Bernadine Harper, and Martha Buntin, for making the Program feel like a home within the institution.

Most importantly, I would not be here today were it not for the love and support of my family. Immigrants from Ethiopia, my parents brought with them a strong work ethic that they instilled in my brother and me. Early on, my father, Hailu, a veterinary pathologist, piqued my curiosity by showing me a wide variety of disease manifestations in animals, inspiring my eventual pursuit of studying human diseases. After focusing on raising my brother and me, my mother, Haragewen, a mathematics professor, obtained her doctoral degree – the latter being a much simpler task than the former. My closest friend and brother, Benyam, and I have formed an uncommon bond in our joint quest to become medical scientists.

I recognize that being surrounded by such wonderful people – in the lab, at my institution, and at home – is indeed a rare privilege. In the forthcoming chapters, I describe how appreciating a different kind of rarity proved fundamental to my doctoral thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Mutations are a driving force

Genetic mutations underlie many aspects of life and death – through evolution and disease, respectively.

Accordingly, their measurement is critical to several fields of research. Luria and Delbrück's classic fluctuation analysis is a prototypic example of the insights into biological processes that can be gained simply by counting the number of mutations in carefully controlled experiments [1]. Counting de novo mutations in humans, not present in their parents, have similarly led to new insights into the rate at which our species can evolve [2,3]. Similarly, counting genetic or epigenetic changes in tumors can inform fundamental issues in cancer biology [4]. Mutations lie at the core of current problems in managing patients with viral diseases such as AIDS and hepatitis by virtue of the drug-resistance they can cause [5,6]. Detection of such mutations, particularly at a stage prior to their becoming dominant in the population, will likely be essential to optimize therapy.

Exploiting mutations for rational cancer diagnostics

In comprehensive cancer genome sequence determination, the unbiased analyses considered a tour de force a decade ago [7,8] have now become routine [9]. As a result, the genomic sequences of the most common cancers have been deciphered. As all cancers are caused by mutations, all cancers can theoretically be identified through their mutations.

This insight may potentially revolutionize cancer diagnostics, which are broadly defined as tests for the detection of cancers. Goals of these tests include the early detection of disease, to identify asymptomatic patients that may undergo curative therapy with high rates of success; precision medicine, where disease treatment is tailored to the vulnerabilities specific to a particular patient's disease; and prognosis and surveillance, which guide physician and patient expectations for the probability of eradicating existing disease and developing a recurrence. By considering the mutation spectrum of a particular cancer, sequence-based diagnostics can be rationally designed for virtually any cancer.

The most challenging technical barrier to utilizing mutation detection for cancer diagnostics is the relatively low prevalence of mutations seen in some clinical specimens – sometimes as low as 0.01% [10]. When cancers shed mutant DNA into readily accessible diagnostic compartments such as blood and stool, their mutant DNA must be discriminated from a large excess of wild-type DNA. At the low proportions commonly observed, reliably discriminating these mutations from technical errors may be impossible. The importance of accuracy is paramount as results of these tests may influence the decision to use invasive follow-up testing.

In the ensuing chapters, I present studies highlighting the feasibility of exploiting mutation detection for cancer diagnostics from clinical specimens. First, I describe a generalizable, massively parallel sequencing-based method named the Safe-Sequencing System ("Safe-SeqS"), capable of reliably detecting mutations arising from small populations of cancer cells. In the following two chapters, I provide concrete examples of how Safe-SeqS can power new diagnostics by applying it to cancers of the gynecologic and urinary tract. Finally I close

with a survey of the amount of mutant DNA shed into readily accessible patient specimens by a variety of human cancers, which optimistically suggests that the management of several cancers can be improved by applying Safe-SeqS or related mutation detection technologies.

Acknowledgments

The section "Mutations are a driving force" first appeared in the Introduction of 10.1073/pnas.1105422108[11].

Chapter 2: Detection and Quantification of Rare Mutations With Massively Parallel Sequencing

Introduction

In neoplastic diseases, which are all driven by somatic mutations, the applications of rare mutant detection are manifold; they can be used to help identify residual disease at surgical margins or in lymph nodes, to follow the course of therapy when assessed in plasma, and perhaps to identify patients with early, surgically curable disease when evaluated in stool, sputum, plasma, and other bodily fluids [10,12,13].

These examples highlight the importance of identifying rare mutations for both basic and clinical research. Accordingly, innovative ways to assess them have been devised over the years. The first methods involved biologic assays based on prototrophy, resistance to viral infection or drugs, or biochemical assays [1,14-20]. Molecular cloning and sequencing provided a new dimension to the field, as it allowed the type of mutation, rather than simply its presence, to be identified [21-26]. Some of the most powerful of these newer methods are based on Digital PCR, in which individual molecules are assessed one-by-one [27]. Digital PCR is conceptually identical to the analysis of individual clones of bacteria, cells, or virus, but is performed entirely in vitro with defined, inanimate reagents. Several implementations of Digital PCR have been described, including the analysis of molecules arrayed in multi-well plates, in polonies, in microfluidic devices, and in water-in-oil emulsions [27-32]. In each of

these technologies, mutant templates are identified through their binding to oligonucleotides specific for the potentially mutant base.

Massively parallel sequencing represents a particularly powerful form of Digital PCR in that hundreds of millions of template molecules can be analyzed one-by-one. It has the advantage over conventional Digital PCR methods in that multiple bases can be queried sequentially and easily in an automated fashion. However, massively parallel sequencing cannot generally be used to detect rare variants because of the high error rate associated with the sequencing process. For example, with the commonly used Illumina sequencing instruments, this error rate varies from ~1%[33,34] to ~0.05% [35,36], depending on factors such as the read length [37], use of improved base calling algorithms [38-40] and the type of variants detected [41]. Some of these errors presumably result from mutations introduced during template preparation, during the pre-amplification steps required for library preparation and during further solid-phase amplification on the instrument itself. Other errors are due to base mis-incorporation during sequencing and base-calling errors. Advances in base-calling can enhance confidence (e.g., [38-41]), but instrument-based errors are still limiting, particularly in clinical samples wherein the mutation prevalence can be 0.01% or less [10].

Our approach, called "Safe-SeqS" for Safe-Sequencing System, involves two basic steps (Fig. 1). The first is the assignment of a unique identifier (UID) to each DNA template molecule to be analyzed. The second is the amplification of each uniquely tagged template, so that many daughter molecules with the identical sequence are generated (defined as a UID-family). If a mutation pre-existed in the template molecule used for amplification, that mutation should be present in every daughter molecule containing that UID (barring any

subsequent replication or sequencing errors). A UID-family in which every family member has the identical mutation is called a "super-mutant". Mutations not occurring in the original templates, such as those occurring during the amplification steps or through errors in base-calling, should not give rise to super-mutants. Conceptual and practical issues related to UID assignment and super-mutants are discussed in detail in the SI text.

Results

Endogenous UIDs. UIDs, sometimes called barcodes or indexes, can be assigned to nucleic acid fragments in many ways. These include the introduction of exogenous sequences through PCR [42,43] or ligation [44,45]. Even more simply, randomly sheared genomic DNA inherently contains UIDs consisting of the sequences of the two ends of each sheared fragment (Fig. 2 and Fig. S1). Paired-end sequencing of these fragments yields UID-families that can be analyzed as described above. To employ such endogenous UIDs in Safe-SeqS, we used two separate approaches: one designed to evaluate many genes simultaneously and the other designed to evaluate a single gene fragment in depth (Fig. 2 and Fig. S1, respectively).

For the evaluation of multiple genes, we ligated standard Illumina sequencing adapters to the ends of sheared DNA fragments to produce a standard sequencing library, then captured genes of interest on a solid phase [46]. In this experiment, a library made from the DNA of ~15,000 normal cells was used, and 2,594 bp from six genes were targeted for capture. After excluding known single nucleotide polymorphisms, 25,563 apparent mutations, corresponding to $2.4 \times 10^{-4} \pm$ mutations/bp, were also identified (Table 1). Based on previous analyses of mutation rates in human cells, at least 90% of these apparent mutations

were likely to represent mutations introduced during template and library preparation or base-calling errors. Note that the error rate determined here (2.4×10^{-4} mutations/bp) is considerably lower than usually reported in experiments using the Illumina instrument because we used very stringent criteria for base calling (see SI Materials and Methods).

With Safe-SeqS analysis of the same data, we determined that 69,505 original template molecules were assessed in this experiment (i.e., 69,505 UID-families, with an average of 40 members per family, were identified, Table 1). All of the polymorphic variants identified by conventional analysis were also identified by Safe-SeqS. However, only 8 super-mutants were observed among these families, corresponding to 3.5×10^{-6} mutations/bp. Thus Safe-SeqS decreased the presumptive sequencing errors by at least 70-fold.

A strategy employing endogenous UIDs was also used to reduce false positive mutations upon deep sequencing of a single region of interest. In this case, a library prepared as described above from ~1,750 normal cells was used as template for inverse PCR employing primers complementary to a gene of interest, so the PCR products could be directly used for sequencing (Fig. S1). With conventional analysis, an average of 2.3×10^{-4} mutations/bp were observed, similar to that observed in the capture experiment (Table 1). Given that only 1,057 independent molecules from normal cells were assessed in this experiment, as determined through Safe-SeqS analysis, all mutations observed with conventional analysis likely represented false positives (Table 1). With Safe-SeqS analysis of the same data, no super-mutants were identified at any position.

Exogenous UIDs. Though the results described above show that Safe-SeqS can increase the reliability of massively parallel sequencing, the number of different molecules that can be examined using endogenous UIDs is limited. For fragments sheared to an average size of 150 bp (range 125-175), 36 base paired-end sequencing can evaluate a maximum of ~7,200 different molecules containing a specific mutation (2 reads x 2 orientations x 36 bases/read x 50 base variation on either end of the fragment). In practice, the actual number of UIDs is smaller because the shearing process is not entirely random.

To make more efficient use of the original templates, we developed a Safe-SeqS strategy that employed a minimum number of enzymatic steps. This strategy also permitted the use of degraded or damaged DNA, such as found in clinical specimens or after bisulfite-treatment for the examination of cytosine methylation [47]. As depicted in Fig. 3, this strategy employs two sets of PCR primers. The first set is synthesized with standard phosphoramidite precursors and contained sequences complementary to the gene of interest on the 3' end and different tails at the 5' ends of both the forward and reverse primers. The different tails allowed universal amplification in the next step. Finally, there was a stretch of 12 to 14 random nucleotides between the tail and the sequence-specific nucleotides in the forward primer [42]. The random nucleotides form the UIDs. An equivalent way to assign UIDs to fragments, not used in this study, would employ 10,000 forward primers and 10,000 reverse primers synthesized on a microarray. Each of these 20,000 primers would have gene-specific primers at their 3'-ends and one of 10,000 specific, predetermined, non-overlapping UID sequences at their 5'-ends, allowing for 10^8 (i.e., $[10^4]^2$) possible UID combinations. In either case, two cycles of PCR are performed with the primers and a high-fidelity polymerase, producing a uniquely tagged, double-stranded DNA fragment from each of the two strands

of each original template molecule (Fig. 3). The residual, unused UID assignment primers are removed by digestion with a single-strand specific exonuclease, without further purification, and two new primers are added. The new primers, complementary to the tails introduced in the UID assignment cycles, contain grafting sequences at their 5' ends, permitting solid-phase amplification on the Illumina instrument, and phosphorothioate residues at their 3' ends to make them resistant to any remaining exonuclease. Following 25 additional cycles of PCR, the products are loaded on the Illumina instrument. As shown below, this strategy allowed us to evaluate the majority of input fragments and was used for several illustrative experiments.

Analysis of DNA polymerase fidelity. Measurement of the error rates of DNA polymerases is essential for their characterization and dictates the situations in which these enzymes can be used. We chose to measure the error rate of Phusion polymerase, as this polymerase has one of the lowest reported error frequencies of any commercially available enzyme and therefore poses a particular challenge for an in vitro-based approach. We first amplified a single human DNA template molecule, comprising a segment of an arbitrarily chosen human gene, through 19 rounds of PCR. The PCR products from these amplifications, in their entirety, were used as templates for Safe-SeqS as described in Fig. 3. In seven independent experiments of this type, the number of UID-families identified by sequencing was 624,678 ± 421,274, which is consistent with an amplification efficiency of 92 ± 9.6% per round of PCR.

The error rate of Phusion polymerase, estimated through cloning of PCR products encoding β-galactosidase in plasmid vectors and transformation into bacteria, is reported by the

manufacturer to be 4.4 x 10^{-7}errors/bp/PCR cycle. Even with very high stringency base-calling, conventional analysis of the Illumina sequencing data revealed an apparent error rate of 9.1 x 10^{-6} errors/bp/PCR cycle, more than an order of magnitude higher than the reported Phusion polymerase error rate (Table 2A). In contrast, Safe-SeqS of the same data revealed an error rate of 4.5 x 10^{-7}errors/bp/PCR cycle, nearly identical to that measured for Phusion polymerase in biological assays (Table 2A). The vast majority (>99%) of these errors were single base substitutions (Table S1A), consistent with previous data on the mutation spectra created by other prokaryotic DNA polymerases [17,48,49].

Safe-SeqS also allowed a determination of the total number of distinct mutational events and an estimation of PCR cycle in which the mutation occurred. There were 19 cycles of PCR performed in wells containing a single template molecule in these experiments. If a polymerase error occurred in cycle 19, there would be only one super-mutant produced (from the strand containing the mutation). If the error occurred in cycle 18 there should be two super-mutants (derived from the mutant strands produced in cycle 19), etc. Accordingly, the cycle in which the error occurred is related to the number of super-mutants containing that error. The data from seven independent experiments demonstrate a relatively consistent number of observed total polymerase errors (2.2 ± 1.1 x 10^{-6} distinct mutations/bp), in good agreement with the expected number of observations from simulations (1.5 ± 0.21 x 10^{-6} distinct mutations/bp, as detailed in SI text). The data also show a highly variable timing of occurrence of polymerase errors among experiments (Table S2), as predicted from classic fluctuation analysis [1]. This kind of information is difficult to derive using conventional analysis of the same next-generation sequencing data, in part because of the prohibitively high apparent mutation rate noted above.

Analysis of oligonucleotide composition. A small number of mistakes during the synthesis of oligonucleotides from phoshoramidite precursors are tolerable for most applications, such as routine PCR or cloning. However, for synthetic biology, wherein many oligonucleotides must be joined together, such mistakes present a major obstacle to success. Clever strategies for making the gene construction process more efficient have been devised [50,51], but all such strategies would benefit from more accurate synthesis of the oligonucleotides themselves. Determining the number of errors in synthesized oligonucleotides is difficult because the fraction of oligonucleotides containing errors can be lower than the sensitivity of conventional next-generation sequencing analyses.

To determine whether Safe-SeqS could be used for this determination, we used standard phosphoramidite chemistry to synthesize an oligonucleotide containing 31 bases that were designed to be identical to that analyzed in the polymerase fidelity experiment described above. In the synthetic oligonucleotide, the 31 bases were surrounded by sequences complementary to primers that could be used for the UID assignment steps of Safe-SeqS (Fig. 3). By performing Safe-SeqS on ~300,000 oligonucleotides, we found that there were $8.9 \pm 0.28 \times 10^{-4}$ super-mutants/bp and that these errors occurred throughout the oligonucleotides (Fig. S2A). The oligonucleotides contained a large number of insertion and deletion errors, representing $8.2 \pm 0.63\%$ and $25 \pm 1.5\%$ of the total super-mutants, respectively. Importantly, both the position and nature of the errors were highly reproducible among seven independent replicates of this experiment performed on the same batch of oligonucleotides (Fig. S2A). This nature and distribution of errors had little in common with that of the errors produced by Phusion polymerase (Fig. S2B and Table S3),

which were distributed in the expected stochastic pattern among replicate experiments. The number of errors in the oligonucleotides synthesized with *phosphoramidites was ~60 times higher than in the equivalent products* synthesized by Phusion polymerase. These data, in toto, indicate that the vast majority of errors in the former were generated during their synthesis rather than during the Safe-SeqS procedure.

Does Safe-SeqS preserve the ratio of mutant:normal sequences in the original templates? To address this question, we synthesized two 31-base oligonucleotides of identical sequence with the exception of nt 15 (50:50 C/G instead of T) and mixed them at nominal mutant/normal fractions of 3.3% and 0.33%. Through Safe-SeqS analysis of the oligonucleotide mixtures, we found that the ratios were 2.8% and 0.27%, respectively. We conclude that the UID assignment and amplification procedures used in Safe-SeqS do not greatly alter the proportion of variant sequences and thereby provide a reliable estimate of that proportion when unknown. This conclusion is also supported by the reproducibility of variant fractions when analyzed in independent Safe-SeqS experiments (Fig. S2A).

Analysis of DNA sequences from normal human cells. The exogenous UID strategy (Fig. 3) was then used to determine the prevalence of rare mutations in a small region of the *CTNNB1* gene from ~100,000 normal human cells from three unrelated individuals. Through comparison with the number of UID-families obtained in the Safe-SeqS experiments (Table 2B), we calculated that the majority (78 ± 9.8 %) of the input fragments were converted into UID-families. There was an average of 68 members/UID-family, easily fulfilling the required redundancy for Safe-SeqS (Fig. S3). Conventional analysis of the Illumina sequencing data revealed an average of 118,488 ± 11,357 mutations among the

~560 Mb of sequence analyzed per sample, corresponding to an apparent mutation prevalence of 2.1 ± 0.16 x 10^{-4} mutations/bp (Table 2B). Only an average of 99 ± 78 super-mutants were observed in the Safe-SeqS analysis. The vast majority (>99%) of super-mutants were single base substitutions and the calculated mutation rate was 9.0 ± 3.1 x 10^{-6} mutations/bp (Table S1B). Safe-SeqS thereby reduced the apparent frequency of mutations in genomic DNA by at least 24-fold (Fig. 4).

We applied the identical strategy to a short segment of mitochondrial DNA in ~1,000 cells from each of seven unrelated individuals. Conventional analysis of the Illumina sequencing libraries produced with the Safe-SeqS procedure (Fig. 3) revealed an average of 30,599 ± 12,970 mutations among the ~150 Mb of sequence analyzed per sample, corresponding to an apparent mutation prevalence of 2.1 ± 0.94 x 10^{-4} mutations/bp (Table 2C). Only 135 ± 61 super-mutants were observed in the Safe-SeqS analysis. As with the *CTNNB1* gene, the vast majority of mutations were single base substitutions, though occasional single base deletions were also observed (Table S1C). The calculated mutation rate in the analyzed segment of mtDNA was 1.4 ± 0.68 x 10^{-5} mutations/bp (Table 2C). Thus, Safe-SeqS thereby reduced the apparent frequency of mutations in genomic DNA by at least 15-fold.

Discussion

The results described above demonstrate that the Safe-SeqS approach can substantially improve the accuracy of massively parallel sequencing (Tables 1 and 2). It can be implemented through either endogenous or exogenously introduced UIDs and can be applied to virtually any sample preparation workflow or sequencing platform. As demonstrated here, the approach can easily be used to identify rare mutants in a population

of DNA templates, to measure polymerase error rates, and to judge the reliability of oligonucleotide syntheses. One of the advantages of the strategy is that it yields the number of templates analyzed as well as the fraction of templates containing variant bases. Previously described in vitro methods for the detection of small numbers of template molecules (e.g., [31,52]) allow the fraction of mutant templates to be determined but cannot determine the number of mutant and normal templates in the original sample.

It is of interest to compare Safe-SeqS to other approaches for reducing errors in next-generation sequencing. As mentioned in the Introduction, sophisticated algorithms to increase the accuracy of base-calling have been developed (e.g., [38-41]). These can certainly reduce false positive calls, but their sensitivity is still limited by artifactual mutations occurring during the PCR steps required for library preparation as well as by (a reduced number of) base-calling errors. For example, the algorithm employed in the current study used very stringent criteria for base-calling and was applied to short read-lengths, but was still unable to reduce the error rate to less than an average of 2.0×10^{-4} errors/bp. This error frequency is at least as low as those reported with other algorithms. To improve sensitivity further, these base-calling improvements can be used together with Safe-SeqS. Travers *et al.* have described another powerful strategy for reducing errors [53]. With this technology, both strands of each template molecule are sequenced redundantly after a number of preparative enzymatic steps. However, this approach can only be performed on a specific instrument. Moreover, for many clinical applications, there are relatively few template molecules in the initial sample and evaluation of nearly all of them is required to obtain the requisite sensitivity. The approach described here with exogenously introduced UIDs (Fig. 3) fulfills this requirement by coupling the UID assignment step with a subsequent amplification in

which few molecules are lost. Our endogenous UID approaches (Fig. 2 and Fig. S1) and the one described by Travers *et al.* are not well-suited for this purpose because of the inevitable losses of template molecules during the ligation and other preparative steps.

How do we know that the mutations identified by conventional analyses in the current study represent artifacts rather than true mutations in the original templates? Strong evidence supporting this is provided by the observation that the mutation prevalence in all but one experiment was similar - 2.0×10^{-4} to 2.4×10^{-4} mutations/bp (Tables 1 and 2). The exception was the experiment with oligonucleotides synthesized from phosphoramidites, in which the error of the synthetic process was apparently higher than the error rate of conventional Illumina analysis when used with stringent base-calling criteria. In contrast, the mutation prevalence of Safe-SeqS varied much more, from 0.0 to 1.4×10^{-5} mutations/bp, depending on the template and experiment. Moreover, the mutation prevalence measured by Safe-SeqS in the most controlled experiment, in which polymerase fidelity was measured (Table 2A), was almost identical to that predicted from previous experiments in which polymerase fidelity was measured by biological assays. Our measurements of mutation prevalence in the DNA from normal cells are consistent with some previous experimental data. However, estimates of these prevalences vary widely and may depend on cell type and sequence analyzed (see SI text). We therefore cannot be certain that the few mutations revealed by Safe-SeqS represented errors occurring during the sequencing process rather than true mutations present in the original DNA templates. Potential sources of error in the Safe-SeqS process are described in the SI text.

Like all techniques, Safe-SeqS has limitations. For example, we have demonstrated that the exogenous UIDs strategy can be used to analyze a single amplicon in depth. This technology may not be applicable to situations wherein multiple amplicons must be analyzed from a sample containing a limited number of templates. Multiplexing in the UID assignment cycles (Fig. 3) may provide a solution to this challenge. A second limitation is that the efficiency of amplification in the UID assignment cycles is critical for the success of the method. Clinical samples can contain inhibitors that reduce the efficiency of this step. This problem can presumably be overcome by performing more than two cycles in the UID assignment PCR step (Fig. 3), though this would complicate the determination of the number of templates analyzed. The specificity of Safe-SeqS is currently limited by the fidelity of the polymerase used in the UID assignment PCR step, i.e., 8.8×10^{-7} mutations/bp in its current implementation with two cycles. Increasing the number of cycles in the UID assignment PCR step to five would decrease the overall specificity to $\sim 2 \times 10^{-6}$ mutations/bp. However, this specificity can be increased by requiring more than one super-mutant for mutation identification - the probability of introducing the same artifactual mutation twice or three times would be exceedingly low ($[2 \times 10^{-6}]^{2}$ or $[2 \times 10^{-6}]^{3}$, respectively). In sum, there are several simple ways to vary the Safe-SeqS procedure and analysis to realize the needs of specific experiments.

Luria and Delbrück, in their classic paper in 1943, wrote that their "prediction cannot be verified directly, because what we observe, when we count the number of resistant bacteria in a culture, is not the number of mutations which have occurred but the number of resistant bacteria which have arisen by multiplication of those which mutated, the amount of multiplication depending on how far back the mutation occurred." The Safe-SeqS procedure

described here can verify such predictions because the number as well as the time of occurrence of each mutation can be estimated from the data, as noted in the experiments on polymerase fidelity. In addition to templates generated by polymerases in vitro, the same approach can be applied to DNA from bacteria, viruses, and mammalian cells. We therefore expect that this strategy will provide definitive answers to a variety of important biomedical questions.

Materials and Methods

Endogenous UIDs. To create endogenous UIDs, DNA was fragmented to an average size of ~200 bp by acoustic shearing (Covaris), then end-repaired, A-tailed, and ligated to Y-shaped adapters according to standard Illumina protocols. DNA was captured [46] with a filter containing 2,594 nt corresponding to six cancer genes. For the inverse PCR experiments (Fig. S1), we ligated custom adapters (IDT, Table S4) instead of standard Y-shaped Illumina adapters to sheared cellular DNA. Inverse PCR was performed using *KRAS* forward and reverse primers (Table S4) and 1U of Phusion polymerase. The *KRAS*-specific primers both contained grafting sequences for hybridization to the Illumina GA IIx flow cell (Table S4). Further details are provided in SI Materials and Methods.

Exogenous UIDs. Each strand of each template molecule was encoded with a 12 or 14 base UID using two cycles of amplicon-specific PCR, as described in the text and Fig. 3. The amplicon-specific primers both contained universal tag sequences at their 5' ends for a later amplification step. The UIDs constituted 12 or 14 random nucleotide sequences appended to the 5' end of the forward amplicon-specific primers (Table S4). Following 2 cycles of PCR for UID assignment, the products were digested with a single strand DNA specific nuclease.

Primers complementary to the introduced universal tags and containing 3' terminal phosphorothioates (Table S4) were added and 25 additional cycles of PCR were performed. Further details are provided in SI Materials and Methods.

Sequencing. Sequencing of all the libraries described above was performed using an Illumina GA IIx instrument as specified by the manufacturer. High quality reads were grouped into UID-families based on their endogenous or exogenous UIDs. Only UID-families with two or more members were considered, as described in detail in the SI Materials and Methods.

Supplementary Materials and Methods

Endogenous UIDs. Genomic DNA from human pancreas or cultured lymphoblastoid cells was prepared using Qiagen kits. The pancreas DNA was used for the capture experiment and the lymphoblastoid cells were used for the inverse PCR experiment. DNA was quantified by optical absorbance and with qPCR. DNA was fragmented to an average size of ~200 bp by acoustic shearing (Covaris), then end-repaired, A-tailed, and ligated to Y-shaped adapters according to standard Illumina protocols. The ends of each template molecule provide endogenous UIDs corresponding to their chromosomal positions. After PCR-mediated amplification of the libraries with primer sequences within the adapters, DNA was captured [46] with a filter containing 2,594 nt corresponding to six cancer genes. After capture, 18 cycles of PCR were performed to ensure sufficient amounts of template for sequencing on an Illumina GA IIx instrument.

For the inverse PCR experiments (Fig. S1), we ligated custom adapters (IDT, Table S4) instead of standard Y-shaped Illumina adapters to sheared cellular DNA. These adapters retained the region complementary to the universal sequencing primer but lacked the grafting sequences required for hybridization to the Illumina GA IIx flow cell. The ligated DNA was diluted into 96 wells and the DNA in each column of 8 wells was amplified with a unique forward primer containing one of 12 index sequences at its 5' end plus a standard reverse primer (Table S4). Amplifications were performed using Phusion HotStart I (NEB) in 50 uL reactions containing 1X Phusion HF buffer, 0.5 mM dNTPs, 0.5 uM each forward and reverse primer (both 5'-phosphorylated), and 1U of Phusion polymerase. The following cycling conditions were used: one cycle of 98^0C for 30s; and 16 cycles of 98^0C for 10s, 65^0C for 30s, and 72^0C for 30s. All 96 reactions were pooled and then purified using a Qiagen MinElute PCR Purification Kit (cat. no. 28004) and a QIAquick Gel Extraction kit (cat. no. 28704). To prepare the circular templates necessary for inverse PCR, DNA was diluted to ~1 ng/uL and ligated with T4 DNA Ligase (Enzymatics) for 30min at room temperature in a 600uL reaction containing 1X T4 DNA Ligation Buffer and 18,000U of T4 DNA Ligase. The ligation reaction was purified using a Qiagen MinElute kit. Inverse PCR was performed using Phusion Hot Start I on 90 ng of circular template distributed in twelve 50 uL reactions, each containing 1X Phusion HF Buffer, 0.25mM dNTPs, 0.5uM each of *KRAS* forward and reverse primers (Table S4) and 1U of Phusion polymerase. The *KRAS*-specific primers both contained grafting sequences for hybridization to the Illumina GA IIx flow cell (Table S4). The following cycling conditions were used: one cycle of 98^0C for 2 min; and 37 cycles of 98^0C for 10s, 61^0C for 15s, and 72^0C for 10s. The final purification was performed with a NucleoSpin Extract II kit (Macherey-Nagel) and eluted in 20uL NE Buffer. The resulting DNA fragments contained UIDs composed of three sequences: two endogenous ones,

represented by the two ends of the original sheared fragments plus the exogenous sequence introduced during the indexing amplification. As 12 exogenous sequences were used, this increased the number of distinct UIDs by 12-fold over that obtained without exogenous UIDs. This number could easily be increased by using a greater number of distinct primers.

Exogenous UIDs. Genomic DNA from normal human colonic mucosae or blood lymphocytes was prepared using Qiagen kits. The DNA from colonic mucosae was used for the experiments on *CTNNB1* and mitochondrial DNA, while the lymphocyte DNA was used for the experiments on *CTNNB1* and on polymerase fidelity. DNA was quantified with Digital PCR [27] using primers that amplified single-copy genes from human cells (Analysis of Polymerase Fidelity and *CTNNB1*), qPCR (mitochondrial DNA), or by optical absorbance (oligonucleotides). Each strand of each template molecule was encoded with a 12 or 14 base UID using two cycles of amplicon-specific PCR, as described in the text and Fig. 3. The amplicon-specific primers both contained universal tag sequences at their 5' ends for a later amplification step. The UIDs constituted 12 or 14 random nucleotide sequences appended to the 5' end of the forward amplicon-specific primers (Table S4). These primers can generate 16.8 and 268 million distinct UIDs, respectively. It is important that the number of distinct UIDs greatly exceed the number of original template molecules to minimize the probability that two different original templates acquired the same UID. The UID assignment PCR cycles included Phusion Hot Start II (NEB) in a 45 uL reaction containing 1X Phusion HF buffer, 0.25mM dNTPs, 0.5 uM each forward (containing 12-14 Ns) and reverse primers, and 2U of Phusion polymerase. To keep the final template concentrations <1.5 ng/uL, multiple wells were used to create some libraries. The following cycling conditions were employed: one cycle of 98^0C for 30s; and two cycles of 98^0C for 10

s, 61^{0}C for 120 s, and 72^{0}C for 10 s. To ensure complete removal of the first round primers, each well was digested with 60 U of a single strand DNA specific nuclease (Exonuclease-I; Enzymatics) at 37^{0}C for 1hr. After a 5 min heat-inactivation at 98^{0}C, primers complementary to the introduced universal tags (Table S4) were added to a final concentration of 0.5uM each. These primers contained two terminal phosphorothioates to make them resistant to any residual Exonuclease-I activity. They also contained 5' grafting sequences necessary for hybridization to the Illumina GA IIx flow cell. Finally, they contained an index sequence between the grafting sequence and the universal tag sequence. This index sequence enables the PCR products from multiple different individuals to be simultaneously analyzed in the same flow cell compartment of the sequencer. The following cycling conditions were used for the subsequent 25 cycles of PCR: 98^{0}C for 10s and 72^{0}C for 15s. No intermediate purification steps were performed in an effort to reduce the losses of template molecules.

After the second round of amplification, wells were consolidated and purified using a Qiagen QIAquick PCR Purification Kit (cat. no. 28104) and eluted in 50 uL EB Buffer (Qiagen). Fragments of the expected size were purified after agarose (mtDNA libraries) or polyacrylamide (all other libraries) gel electrophoresis. For agarose gel purification, the eight 6-uL aliquots were loaded into wells of a 2% Size Select Gel (Invitrogen) and bands of the expected size were collected in EB Buffer as specified by the manufacturer. For polyacrylamide gel purification, ten 5-uL aliquots were loaded into wells of a 10% TBE Polyacrylamide Gel (Invitrogen). Gel slices containing the fragments of interest were excised, crushed, and eluted essentially as described [54].

Analysis of Phusion polymerase fidelity. Amplification of a fragment of human genomic DNA within the BMX (RefSeq Accession NM_203281.2) gene was first performed using the PCR conditions described above. The template was diluted so that an average of one template molecule was present in every 10 wells of a 96-well PCR plate. Fifty uL PCR reactions were then performed in 1X Phusion HF buffer, 0.25mM dNTPs, 0.5uM each forward and reverse primers (Table S4), and 2U of Phusion polymerase. The cycling conditions were one cycle of 98^0C for 30s; and 19 cycles of 98^0C for 10 s, 61^0C for 120 s, and 72^0C for 10s. The primers were removed by digestion with 60 U of Exonuclease-I at 37^0C for 1hr followed by a 5 min heat-inactivation at 98^0C. No purification of the PCR product was performed, either before or after Exonuclease-I digestion. The entire contents of each well were then used as templates for the exogenous UIDs strategy described above.

Sequencing. Sequencing of all the libraries described above was performed using an Illumina GA IIx instrument as specified by the manufacturer. The total length of the reads used for each experiment varied from 36 to 73 bases. Base-calling and sequence alignment was performed with the Eland pipeline (Illumina). Only high quality reads meeting the following criteria were used for subsequent analysis: (i) the first 25 bases passed the standard Illumina chastity filter; (ii) every base in the read had a quality score $\geq$20; and (iii) $\leq$ 3 mismatches to expected sequences. For the exogenous UID libraries, we additionally required the UIDs to have a quality score $\geq$30. We noticed a relatively high frequency of errors at the ends of the reads in the endogenous UID libraries prepared with the standard Illumina protocol, presumably introduced during shearing or end-repair, so the first and last three bases of these tags were excluded from analysis .

Safe-SeqS analysis. High quality reads were grouped into UID-families based on their endogenous or exogenous UIDs. Only UID-families with two or more members were considered. Such UID-families included the vast majority (≥99%) of the sequencing reads. To ensure that the same data was used for both conventional and Safe-SeqS analysis, we also excluded UID-families containing only one member from conventional analysis. Furthermore, we only identified a base as "mutant" in conventional sequencing analysis if the same variant was identified in at least two members of at least one UID-family (i.e., two mutations) when comparing conventional analysis to that of Safe-SeqS with exogenous UIDs. For comparison with Safe-SeqS with endogenous UIDs, we required at least two members of each of two UID-families (i.e., four mutations) to identify a position as "mutant" in conventional analysis. With either endogenous or exogenous UIDs, a super-mutant was defined as a UID-family in which ≥95% of members shared the identical mutation. Thus, UID-families with <20 members had to be 100% identical at the mutant position, while a 5% combined replication and sequencing error rate was permitted in UID-families with more members. To determine polymerase fidelity using Safe-SeqS, and to compare the results with previous analyses of Phusion polymerase fidelity, it was necessary to realize that the previous analyses would only detect mutations present in both strands of the PCR products [14]. This would be equivalent to analyzing PCR products generated with one less cycle with Safe-SeqS, and the appropriate correction was made in Table 2A. Unless otherwise specified, all values listed in the text and Tables represent means and standard deviations.

Error-generating processes. Apparent mutations, defined as any base call that varies from the expected base at a defined position, can result from a variety of processes.

1. Mutations present in the template DNA. For templates derived from normal human cells, these include mutations that were present in the zygote, occurred later during embryonic and adult development, or were present in a contaminant inadvertently introduced into the sample. These mutations are expected to be present in both strands of the relevant templates. If the mutation occurred only in the last cell-cycle of a cell whose DNA was used as template, the mutation would be present in only one strand of the template.
2. Chemically-modified bases present in the templates. It has been estimated that there are many thousands of oxidized bases present in every human cell [55]. When such DNA is amplified by Phusion polymerase, an apparent mutation in one strand may result.
3. Errors introduced during the shearing process required to generate small fragments for sequencing. Acoustic shearing generates short-lived, high temperatures that can damage DNA.
4. Errors introduced during end-repair of the sheared fragments. The source of these errors can be polymerase infidelity or through incorporation of chemically-modified bases in the dNTPs used for polymerization.
5. Errors introduced by other enzymatic steps, particularly if the enzymes are impure and contaminated with nucleases, polymerases, or ligases.
6. Errors introduced during PCR amplification to prepare the libraries for capturing or for inverse PCR.
7. Errors during PCR after capturing or during inverse PCR amplification.
8. Errors introduced into the UID assignment cycles of Safe-SeqS (Fig. 3).
9. Errors introduced into the library amplification cycles of Safe-SeqS performed with exogenous UIDs. Note that if UID assignment primers from process #8 are not completely removed, they could potentially amplify DNA fragments containing errors introduced during these cycles, creating a new super-mutant.

10. Errors introduced into the first bridge-PCR cycle on the Illumina flow cell. If amplification is inefficient, an error introduced into the second bridge-PCR cycle could also result in a cluster containing a mutation in most of its component molecules.
11. Errors in base-calling.

Achieving accuracy with Safe-SeqS. With conventional sequencing-by-synthesis approaches, all the error-producing processes described above are relevant, resulting in a relatively high number of false-positive mutation calls (Tables 1 and 2). Safe-SeqS minimizes the number of false-positive mutation calls in several ways. Safe-SeqS with exogenous UIDs results in the fewest false-positive mutation calls because it requires the fewest enzymatic steps. With exogenous UIDs, error-generating processes #3 to #7 are completely eliminated because these steps aren't performed. Safe-SeqS with exogenous UIDs also drastically reduces errors resulting from error-generating processes #10 and #11 because of the way the data is analyzed.

After Safe-SeqS with exogenous UIDs, the only false-positive errors remaining should be those introduced during the UID assignment PCR cycles (error-generating process #8) or residual UID-containing primers during the library amplification cycles (error-generating process #9). The errors from error-generating process #8 can theoretically be eliminated by requiring at least two super-mutants to identify a position as "mutant." This requirement is reasonable because every pre-existing mutation in a double stranded DNA template should give rise to two super-mutants, one from each strand. Furthermore, this requirement would eliminate error-generating process #2 (damaged bases in the original templates) because such bases, when copied, should give rise to only one super-mutant. Finally, errors generated

during the library amplification cycles (process #9) will not be amplified by residual UID-containing primers if those primers are completely removed, such as performed here with excess Exonuclease-I.

With endogenous UIDs, the mistakes introduced by processes #10 and #11 are drastically reduced because of the way in which the data is analyzed (as with exogenous UIDs). Errors introduced in processes #2 to #7 can be minimized by requiring that a mutation be observed in at least two UID-families, for the reasons stated in the paragraph above. With this requirement, few false-positive mutations, in theory, should be identified.

In practice, the situation is complicated by the fact that the various amplifications are not perfect, so every strand of every original template molecule is not recovered as a UID-family. This efficiency can vary from sample to sample, depending in part on the concentration of inhibitors present in clinical samples. Moreover, with exogenous UIDs, a polymerase error during the library amplification step can create a new UID-family that wasn't represented in the UID assignment step. If this error occurred in a mutant template, an additional, artificial super-mutant would be created.

These factors can be managed by incorporating various additional criteria into the analyses. For example, one might require UID-families to contain more than two, five or ten members. Another requirement could be that the exogenous UIDs of super-mutants not be related to any other UID in the library by a one-base difference. This would eliminate artificial super-mutants generated during the library amplification steps (noted in above paragraph). We routinely instituted this requirement in our Safe-SeqS analyses, but it made

little difference (<1%) in the number of super-mutants identified. Specificity for mutations can be further increased by requiring more than one super-mutant to identify a position as "mutant," as described above for endogenous UIDs. By requiring multiple super-mutants, the specificity can be even further increased by requiring that each strand of the original double stranded template contain the mutation or, when libraries are amplified using multiple wells, that rare mutations share an introduced sequence that identifies the well in which the mutations were amplified. Such decisions involve the usual trade-off between specificity and sensitivity. In our experiments with exogenous UIDs (Table 2), we required only one super-mutant to identify a position as "mutant" and included all UID-families with more than one member. As endogenous UIDs was associated with more error-generating processes than with exogenous UIDs, we required two super-mutants to identify a position as mutant in the experiments reported in Table 1 and also included all UID-families with more than one member.

Mutation prevalences in normal human tissues. The experiments reported in Tables 1 and 2, in which > 10,000 templates were assessed, show that mutations are present in the nuclear DNA of normal human cells at a frequency of 3.5×10^{-6} to 9.0×10^{-6} mutants/bp depending on the region analyzed. It is impossible to determine whether this low level represents genuine mutations present in the original templates or the sum of genuine mutations plus artifactual mutations from the error-generating processes described above. Mutation prevalences in human cells have not been widely investigated, in part because they are so infrequent. However, several clever techniques to identify rare mutants have been devised and can in principle be used for comparison. Unfortunately, estimates of human mutation prevalences vary widely, ranging from as many as 10^{-5} mutants/bp to as many as

10^{-8} mutants/bp [15,16,20,56-59]. In several of these studies, the estimates are complicated by the lack of data on the nature of the actual mutations - they could in some cases be caused by losses of whole chromosomes, in others by missense mutations, and in others mainly by nonsense mutations or small insertions or deletions. Additionally, these studies used various sources of normal cells and examined different genes, making direct comparisons difficult. Estimates of the prevalences and rates of mitochondrial DNA mutations similarly vary [22,60-65]. It will be of interest in future work to analyze the same DNA templates and genes with various technologies to determine the basis for these different estimates.

But let us assume that all of the mutations identified with Safe-SeqS represent genuine mutations present in the original DNA templates from normal cells. What does this tell us about the number of generations though which these cells have proceeded since the organism was conceived? There is a simple relationship between mutation rate and mutation prevalence: the mutation prevalence equals the product of the mutation rate and the number of generations that the cell has gone through since conception. The somatic mutation rate has been determined in previous studies to be ~ 10^{-9} mutants/bp/generation, though this estimate also varies from study to study for reasons related to those mentioned above with respect to mutation prevalence. Combining this literature-derived estimate of mutation rate with our estimates of mutation prevalence suggests that the normal cells analyzed (lymphocytes, lymphoblastoid cell lines or colonic mucosae) had proceeded through 3,500 to 8,900 generations, representing cells dividing every 3 to 7 days for the individuals examined in this study (average age 65 years).

Computer simulation of polymerase-introduced errors. The timing of mutations introduced by polymerases greatly alters the final number of mutations observed [1]. For example, two mutations would differ in prevalence by ~64-fold if introduced 6 cycles apart (2^6). Because polymerases introduce mutations in a stochastic manner, a simple Monte Carlo method was employed for the simulations. In these simulations, we used the manufacturer's estimate of the Phusion polymerase error rate with an appropriate adjustment for ability of Safe-SeqS to detect mutations in only one strand [14]. Note that errors introduced in cycle 19, as well as in the two UID assignment cycles, would result in changes in only one strand of the duplex - i.e., result in one super-mutant rather than two. In each experiment, we assumed that there was a constant efficiency of amplification given by the total number of templates obtained at the end of the experiment (i.e., if the number of UID-families was N, then we assumed that the number of templates increased by a factor of $N/2^{19}$ in each cycle). One-thousand simulations were performed for each of seven experiments, and the results reported in Table S2.

Acknowledgments

This chapter first appeared in 10.1073/pnas.1105422108[11].

Table 2-1. Safe-SeqS with Endogenous UIDs

Conventional Analysis	**Capture**	**Inverse PCR**
High quality bp	106,958,863	1,041,346,645
Mean high quality bp read depth	38,620×	2,085,600×
Mutations identified	25,563	234,352
Mutations/bp	**2.4E-04**	**2.3E-04**
Safe-SeqS Analysis		
High quality bp	106,958,863	1,041,346,645
Mean high quality bp read depth	38,620×	2,085,600×
UID-families	69,505	1,057
Average # of members/UID-family	40	21,688
Median # of members/UID-family	19	4
Super-mutants identified	8	0
Super-mutants/bp	**3.5E-06**	**0.0**

Table 2-2. Safe-SeqS with Exogenous UIDs

A. Polymerase Fidelity	Mean	Standard Deviation
Conventional analysis of 7 replicates		
High quality bp	996,855,791	64,030,757
Total mutations identified	198,638	22,515
Mutations/bp	**2.0E-04**	**1.7E-05**
Calculated Phusion Error Rate (errors/bp/cycle)	**9.1E-06**	**7.7E-07**
Safe-SeqS analysis of 7 replicates		
High quality bp	996,855,791	64,030,757
UID-families	624,678	421,274
Members/UID-family	107	122
Total super-mutants identified	197	143
Super-mutants/bp	**9.9E-06**	**2.3E-06**
Calculated Phusion Error Rate (errors/bp/cycle)	**4.5E-07**	**1.0E-07**

Table 2-2. Continued

B. *CTNNB1* mutations in DNA from normal human cells

Conventional analysis of 3 individuals		
High quality bp	559,334,774	66,600,749
Total mutations identified	118,488	11,357
Mutations/bp	**2.1E-04**	**1.6E-05**
Safe-SeqS analysis of 3 individuals		
High quality bp	559,334,774	66,600,749
UID-families	374,553	263,105
Members/UID-family	68	38
Total super-mutants identified	99	78
Super-mutants/bp	**9.0E-06**	**3.1E-06**

C. Mitochondrial mutations in DNA from normal human cells

Conventional analysis of 7 individuals		
High quality bp	147,673,456	54,308,546
Total mutations identified	30,599	12,970
Mutations/bp	**2.1E-04**	**9.4E-05**
Safe-SeqS analysis of 7 individuals		
High quality bp	147,673,456	54,308,546
UID-families	515,600	89,985
Members/UID-family	15	6
Total super-mutants identified	135	61
Super-mutants/bp	**1.4E-05**	**6.8E-06**

Table 2-S1. Fraction of Single Base Substitutions, Insertions, and Deletions with Exogenous UIDs

	Mean	SD
Polymerase fidelity		
Conventional analysis of seven replicates		
Total mutations identified	198,638	22,515
% mutations represented by single-base substitutions	98.90	0.16
% mutations represented by deletions	1.05	0.16
% mutations represented by insertions	0.05	0.01
Safe-SeqS analysis of seven replicates		
Total supermutants identified	197	143
% supermutants represented by single-base substitutions	99.10	2.06
% supermutants represented by deletions	0.90	2.06
% supermutants represented by insertions	0.00	0.00
CTNNB1 mutations in DNA from normal human cells		
Conventional analysis of three individuals		
Total mutations identified	118,488	11,357
% mutations represented by single-base substitutions	97.38	0.30
% mutations represented by deletions	2.62	0.30
% mutations represented by insertions	0.00	0.00
Safe-SeqS analysis of three individuals		
Total supermutants identified	99	78
% supermutants represented by single-base substitutions	99.52	0.84
% supermutants represented by deletions	0.48	0.84
% supermutants represented by insertions	0.00	0.00
Mitochondrial mutations in DNA from normal human cells		
Conventional analysis of seven individuals		
Total mutations identified	30,599	12,970
% mutations represented by single-base substitutions	98.42	0.64
% mutations represented by deletions	1.58	0.64
% mutations represented by insertions	0.00	0.00
Safe-SeqS analysis of seven individuals		
Total supermutants identified	135	61
% supermutants represented by single-base substitutions	98.66	1.39
% supermutants represented by deletions	1.34	1.39
% supermutants represented by insertions	0.00	0.00

Table 2-S2. Observed and Expected Number of Errors Generated by Phusion Polymerase

	Observed	Expected (mean ± SD)*
Experiment 1		
Mutations represented by 1 supermutant	10	19 ± 3.7
Mutations represented by 2 supermutants	8	5.8 ± 2.3
Mutations represented by 3 supermutants	4	1.3 ± 1.1
Mutations represented by 4 supermutants	4	1.8 ± 1.3
Mutations represented by 5 supermutants	2	0.61 ± 0.75
Mutations represented by 6 supermutants	2	0.22 ± 0.44
Mutations represented by 7 supermutants	0	0.01 ± 0.10
Mutations represented by 8 supermutants	0	0.87 ± 0.86
Mutations represented by 9 supermutants	2	0.28 ± 0.51
Mutations represented by 10 supermutants	0	0.14 ± 0.38
Mutations represented by >10 supermutants	3	1.5 ± 2.7
Distinct mutations	35	32 ± 4.2
Experiment 2		
Mutations represented by 1 supermutant	19	23 ± 4.1
Mutations represented by 2 supermutants	5	9.5 ± 2.8
Mutations represented by 3 supermutants	4	2.7 ± 1.6
Mutations represented by 4 supermutants	7	2.7 ± 1.7
Mutations represented by 5 supermutants	2	0.88 ± 0.94
Mutations represented by 6 supermutants	1	0.40 ± 0.60
Mutations represented by 7 supermutants	3	0.16 ± 0.42
Mutations represented by 8 supermutants	1	0.99 ± 1.0
Mutations represented by 9 supermutants	1	0.39 ± 0.68
Mutations represented by 10 supermutants	0	0.17 ± 0.43
Mutations represented by >10 supermutants	9	1.8 ± 3.4
Distinct mutations	52	43 ± 5.1
Experiment 3		
Mutations represented by 1 supermutant	7	17 ± 3.4
Mutations represented by 2 supermutants	9	5.4 ± 2.0
Mutations represented by 3 supermutants	4	1.2 ± 1.1
Mutations represented by 4 supermutants	4	1.7 ± 1.4
Mutations represented by 5 supermutants	2	0.50 ± 0.70
Mutations represented by 6 supermutants	0	0.17 ± 0.45
Mutations represented by 7 supermutants	1	0.03 ± 0.17
Mutations represented by 8 supermutants	0	0.59 ± 0.74
Mutations represented by 9 supermutants	0	0.24 ± 0.50
Mutations represented by 10 supermutants	1	0.07 ± 0.29
Mutations represented by >10 supermutants	5	1.5 ± 2.6
Distinct mutations	33	28 ± 3.7
Experiment 4		
Mutations represented by 1 supermutant	7	15 ± 3.7
Mutations represented by 2 supermutants	8	4.1 ± 1.7
Mutations represented by 3 supermutants	2	0.70 ± 0.74
Mutations represented by 4 supermutants	1	1.5 ± 1.3
Mutations represented by 5 supermutants	3	0.21 ± 0.52
Mutations represented by 6 supermutants	2	0.08 ± 0.27
Mutations represented by 7 supermutants	1	0.0 ± 0.0
Mutations represented by 8 supermutants	2	0.65 ± 0.77
Mutations represented by 9 supermutants	2	0.17 ± 0.43
Mutations represented by 10 supermutants	0	0.05 ± 0.22
Mutations represented by >10 supermutants	1	0.92 ± 2.1
Distinct mutations	29	23 ± 3.2
Experiment 5		
Mutations represented by 1 supermutant	9	23 ± 4.1
Mutations represented by 2 supermutants	6	9.5 ± 2.8
Mutations represented by 3 supermutants	5	2.7 ± 1.6
Mutations represented by 4 supermutants	3	2.7 ± 1.7
Mutations represented by 5 supermutants	6	0.88 ± 0.94
Mutations represented by 6 supermutants	2	0.40 ± 0.60
Mutations represented by 7 supermutants	1	0.16 ± 0.42
Mutations represented by 8 supermutants	2	0.99 ± 1.0
Mutations represented by 9 supermutants	2	0.39 ± 0.68
Mutations represented by 10 supermutants	3	0.17 ± 0.43

Table 2-S2. Continued

	Observed	Expected (mean ± SD)*
Mutations represented by >10 supermutants	7	1.8 ± 3.4
Distinct mutations	46	43 ± 5.1
Experiment 6		
Mutations represented by 1 supermutant	4	6.7 ± 2.8
Mutations represented by 2 supermutants	7	1.5 ± 1.2
Mutations represented by 3 supermutants	1	0.10 ± 0.33
Mutations represented by 4 supermutants	2	0.60 ± 0.82
Mutations represented by 5 supermutants	0	0.07 ± 0.26
Mutations represented by 6 supermutants	0	0.01 ± 0.10
Mutations represented by 7 supermutants	1	0.0 ± 0.0
Mutations represented by 8 supermutants	1	0.39 ± 0.60
Mutations represented by 9 supermutants	0	0.01 ± 0.10
Mutations represented by 10 supermutants	0	0.0 ± 0.0
Mutations represented by >10 supermutants	2	0.50 ± 1.1
Distinct mutations	18	9.9 ± 1.4
Experiment 7		
Mutations represented by 1 supermutant	8	2.9 ± 1.6
Mutations represented by 2 supermutants	2	0.61 ± 0.79
Mutations represented by 3 supermutants	0	0.04 ± 0.24
Mutations represented by 4 supermutants	0	0.41 ± 0.59
Mutations represented by 5 supermutants	1	0.01 ± 0.10
Mutations represented by 6 supermutants	0	0.0 ± 0.0
Mutations represented by 7 supermutants	0	0.0 ± 0.0
Mutations represented by 8 supermutants	0	0.14 ± 0.35
Mutations represented by 9 supermutants	0	0.01 ± 0.10
Mutations represented by 10 supermutants	0	0.0 ± 0.0
Mutations represented by >10 supermutants	0	0.32 ± 0.93
Distinct mutations	11	4.5 ± 0.62

*See Supplementary Information for details of the simulations

Table 2-S3. Phosphoramidite- vs Phusion-Synthesized DNA: Transitions vs Transversions Comparison

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7	Average	SD
Phosphoramidites									
Transition supermutants	496	509	471	396	323	273	470	420	92
Transversion supermutants	1,494	1,499	1,521	1,154	944	907	1,626	1,306	298
P value*	3.4*E*-05								
Phusion									
Transition supermutants	63	275	127	5	87	182	103	120	87
Transversion supermutants	14	124	77	12	57	191	63	77	63
P value*	0.08								

**P* values were calculated using a two-tailed paired *t* test. Exp., experiment.

Table 2-S4. Oligonucleotides Used in this Study

	Sequence
Endogenous UIDs	
Capture	
Adapter, strand 1	/5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter, strand 2	ACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, rev	CAAGCAGAAGACGGCATACGAGATCTCGGCATTCCTGCTGAACCGCTCTTCCGAT*C*T
Postcapture amplification, for	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Postcapture amplification, rev	CAAGCAGAAGACGGCATACGAGATCTCGGCATTCCTGCTGAACCGCTCTTCCGAT*C*T
Sequencing primer, read 1 (Illumina)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Sequencing primer, read 2 (Illumina)	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
Inverse PCR	
Adapter, strand 1	/5Phos/GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
Adapter, strand 2	ACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-1	/5Phos/CGTGATACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-2	/5Phos/ACATCGACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-3	/5Phos/GCCTAAACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-4	/5Phos/TGGTCAACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-5	/5Phos/CACTGTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-6	/5Phos/ATTGGCACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-7	/5Phos/GATCTGACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-8	/5Phos/TCAAGTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-9	/5Phos/CTGATCACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-10	/5Phos/AAGCTAACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-11	/5Phos/GTAGCCACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, for-12	/5Phos/TACAAGACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
Whole-genome amplification, rev	/5Phos/CTCGGCATTCCTGCTGAACCGCTCTTCCGAT*C*T
Inverse PCR, antisense	AATGATACGGCGACCACCGAGATCTACACCAGCAGGCCTTATAATAAAAATAATGA
Inverse PCR, for	CAAGCAGAAGACGGCATACGAGATTGACTGAATATAAACTTGTGGTAGTTG
Sequencing primer 1 (to read internal sequences)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Sequencing primer 2 (to read internal sequences)	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
Index primer 1 (to read experiment indexes)	CGGAAGAGCGTCGTGTAGGGAAAGAGTGT
Index primer 2 (to read experiment indexes)	CGGAAGAGCGGTTCAGCAGGAATGCCGAG
Exogenous UIDs	
Polymerase fidelity	
Digital PCR amplification, for	GGTTACAGGCTCATGATGTAACC
Digital PCR amplification, rev	GATACCAGCTTGGTAATGGCA
UID assignment amplification, for	CGACGTAAAACGACGGCCAGTNNNNNNNNNNNNGGTTACAGGCTCATGATGTAACC
UID assignment amplification, rev	CACACAGGAAACAGCTATGACCATGGATACCAGCTTGGTAATGGCA
Library amplification, for-1	AATGATACGGCGACCACCGAGATCTACACCGTGATCGACGTAAAACGACGGCCA*G*T
Library amplification, for-2	AATGATACGGCGACCACCGAGATCTACACACATCGCGACGTAAAACGACGGCCA*G*T
Library amplification, for-3	AATGATACGGCGACCACCGAGATCTACACGCCTAACGACGTAAAACGACGGCCA*G*T
Library amplification, for-4	AATGATACGGCGACCACCGAGATCTACACTGGTCACGACGTAAAACGACGGCCA*G*T
Library amplification, for-5	AATGATACGGCGACCACCGAGATCTACACCACTGTCGACGTAAAACGACGGCCA*G*T
Library amplification, for-6	AATGATACGGCGACCACCGAGATCTACACATTGGCCGACGTAAAACGACGGCCA*G*T
Library amplification, for-7	AATGATACGGCGACCACCGAGATCTACACGATCTGCGACGTAAAACGACGGCCA*G*T
Library amplification, for-8	AATGATACGGCGACCACCGAGATCTACACTCAAGTCGACGTAAAACGACGGCCA*G*T
Library amplification, for-9	AATGATACGGCGACCACCGAGATCTACACCTGATCCGACGTAAAACGACGGCCA*G*T
Library amplification, for-10	AATGATACGGCGACCACCGAGATCTACACAAGCTACGACGTAAAACGACGGCCA*G*T
Library amplification, rev	CAAGCAGAAGACGGCATACGAGATCACACAGGAAACAGCTATGACCA*T*G
Sequencing primer (to read UID and internal sequences)	CGACGTAAAACGACGGCCAGT
Index primer (to read experiment indexes)	ACTGGCCGTCGTTTTACGTCG

Table 2-S4. Continued

	Sequence
CTNNB1 mutations in DNA from normal human cells	
UID assignment amplification, for	CGACGTAAAACGACGGCCAGTNNNNNNNNNNNNNNGCAGCAACAGTCTTACCTGGACT
UID assignment amplification, rev	CACACAGGAAACAGCTATGACCATGTCCACATCCTCTTCCTCAGGATT
Library amplification, for	AATGATACGGCGACCACCGAGATCTACACCGACGTAAAACGACGGCCA*G*T
Library amplification, rev-1	CAAGCAGAAGACGGCATACGAGATATCACGCACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-2	CAAGCAGAAGACGGCATACGAGATCGATGTCACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-3	CAAGCAGAAGACGGCATACGAGATTGACCACACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-4	CAAGCAGAAGACGGCATACGAGATGCCAATCACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-5	CAAGCAGAAGACGGCATACGAGATCAGATCCACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-6	CAAGCAGAAGACGGCATACGAGATACTTGACACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-7	CAAGCAGAAGACGGCATACGAGATGATCAGCACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-8	CAAGCAGAAGACGGCATACGAGATTAGCTTCACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-9	CAAGCAGAAGACGGCATACGAGATGGCTACCACACAGGAAACAGCTATGACCA*T*G
Library amplification, rev-10	CAAGCAGAAGACGGCATACGAGATCTTGTACACACAGGAAACAGCTATGACCA*T*G
Sequencing primer (to read UID and internal sequences)	CGACGTAAAACGACGGCCAGT
Index primer (to read experiment indexes)	CATGGTCATAGCTGTTTCCTGTGTG
Mitochondrial mutations in DNA from normal human cells	
UID assignment amplification, for	CGACGTAAAACGACGGCCAGTNNNNNNNNNNNNNNTTACCGAGAAAGCTCACAAGAA
UID assignment amplification, rev	CACACAGGAAACAGCTATGACCATGATGCTAAGGCGAGGATGAAA
Library amplification, for-1	AATGATACGGCGACCACCGAGATCTACACACATCGCGACGTAAAACGACGGCCA*G*T
Library amplification, for-2	AATGATACGGCGACCACCGAGATCTACACGCCTAACGACGTAAAACGACGGCCA*G*T
Library amplification, for-3	AATGATACGGCGACCACCGAGATCTACACTGGTCACGACGTAAAACGACGGCCA*G*T
Library amplification, for-4	AATGATACGGCGACCACCGAGATCTACACATTGGCCGACGTAAAACGACGGCCA*G*T
Library amplification, for-5	AATGATACGGCGACCACCGAGATCTACACGATCTGCGACGTAAAACGACGGCCA*G*T
Library amplification, for-6	AATGATACGGCGACCACCGAGATCTACACTCAAGTCGACGTAAAACGACGGCCA*G*T
Library amplification, for-7	AATGATACGGCGACCACCGAGATCTACACCTGATCCGACGTAAAACGACGGCCA*G*T
Library amplification, rev	CAAGCAGAAGACGGCATACGAGATCACACAGGAAACAGCTATGACCA*T*G
Sequencing primer 1 (to read UIDs)	CGACGTAAAACGACGGCCAGT
Sequencing primer 2 (to read internal sequences)	CCTAATTCCCCCCATCCTTAC
Index primer (to read experiment indexes)	ACTGGCCGTCGTTTTACGTCG
Analysis of phosphoramidite oligonucleotide composition	
Synthesized template, WT	GGTTACAGGCTCATGATGTAACCTCTGTGTCTTGGTGTAACTTTAAAACATATTTTTGCCATTACCAAGCTGGTATC
Synthesized template, mut (*S* = 50/50 mix of C and G)	GGTTACAGGCTCATGATGTAACCTCTGTGTCTTGGTGSAACTTTAAAACATATTTTTGCCATTACCAAGCTGGTATC
UID assignment amplification, for	ACACTCTTTCCCTACACGACGCTCNNNNNNNNNNNNGGTGAGTCTGTGCAGGCAT
UID assignment amplification, rev	CTCGAGCACTGTCCTGACTGAGACGATACCAGCTTGGTAATGGCA
Library amplification, for	AATGATACGGCGACCACCGAGATCTACACCGTGATACACTCTTTCCCTACACGACGC*T*C
Library amplification, rev	CAAGCAGAAGACGGCATACGAGATCTCGAGCACTGTCCTGACTGAG*A*C
Sequencing primer (to read UID and internal sequences)	ACACTCTTTCCCTACACGACGCTC

Colors: blue, region complementary to templates; green, template-specific UID sequence; black, universal sequence; purple, experiment-specific index sequence; red, Illumina grafting primers (for hybridization to flow cell). /5Phos/, 5′ phosphate; *, phosphorothioate linkage; for, forward; rev, reverse.

Figure 2-1. Essential Elements of Safe-SeqS.

In the first step, each fragment to be analyzed is assigned a unique identification (UID) DNA sequence (green or blue bars). In the second step, the uniquely tagged fragments are amplified, producing UID-families, each member of which has the same UID. A super-mutant is defined as a UID-family in which ≥95% of family members have the same mutation.

Figure 2-1. Continued.

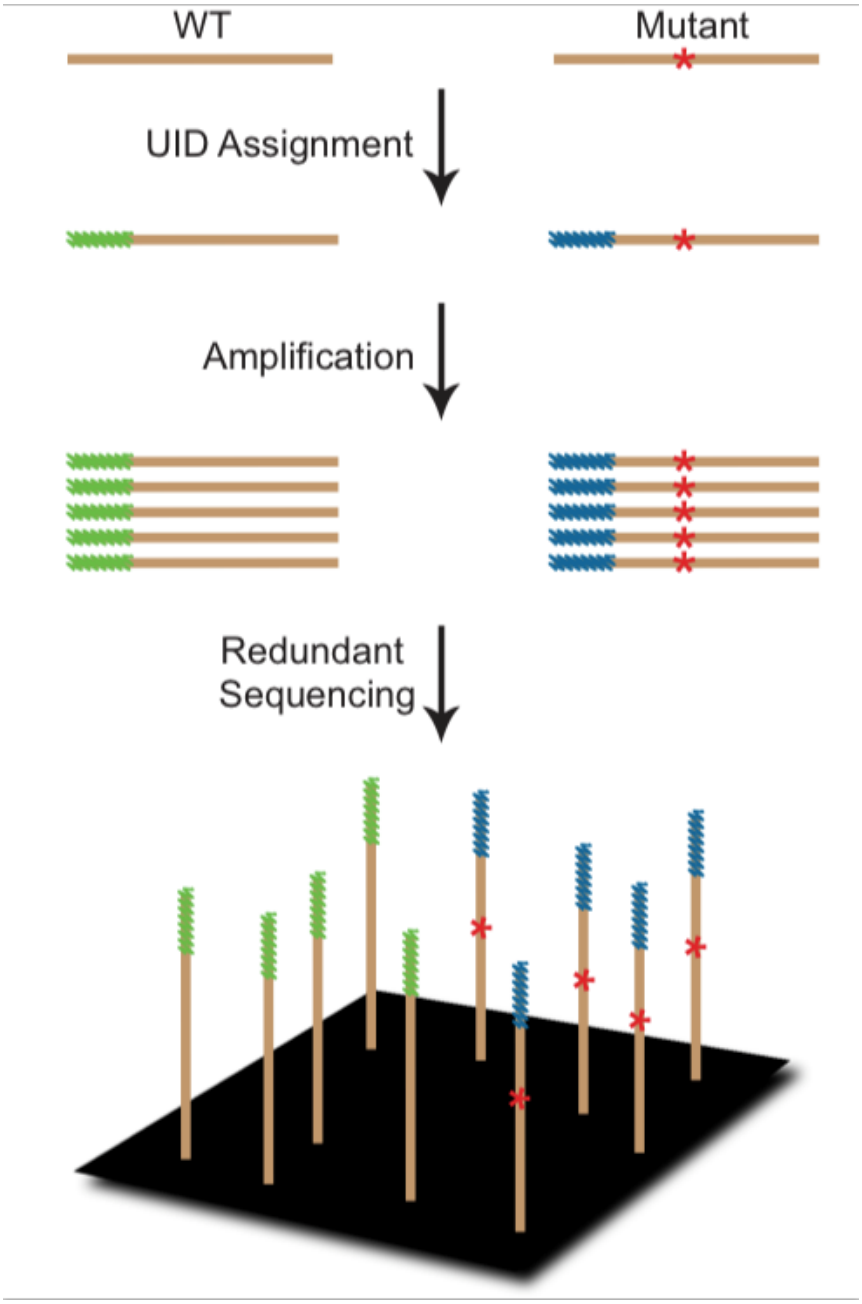


Figure 2-2. Safe-SeqS with Endogenous UIDs Plus Capture.

The sequences of the ends of each fragment produced by random shearing (variously colored bars) serve as the unique identifiers (UIDs). These fragments are ligated to adapters (yellow and orange bars) so they can subsequently be amplified by PCR. One uniquely identifiable fragment is produced from each strand of the double-stranded template; only one strand is shown. Fragments of interest are captured on a solid phase containing oligonucleotides complementary to the sequences of interest. Following PCR amplification to produce UID-families with primers containing 5' "grafting" sequences (black and red bars), sequencing is performed and super-mutants are defined as in Fig. 1.

Figure 2-2. Continued.

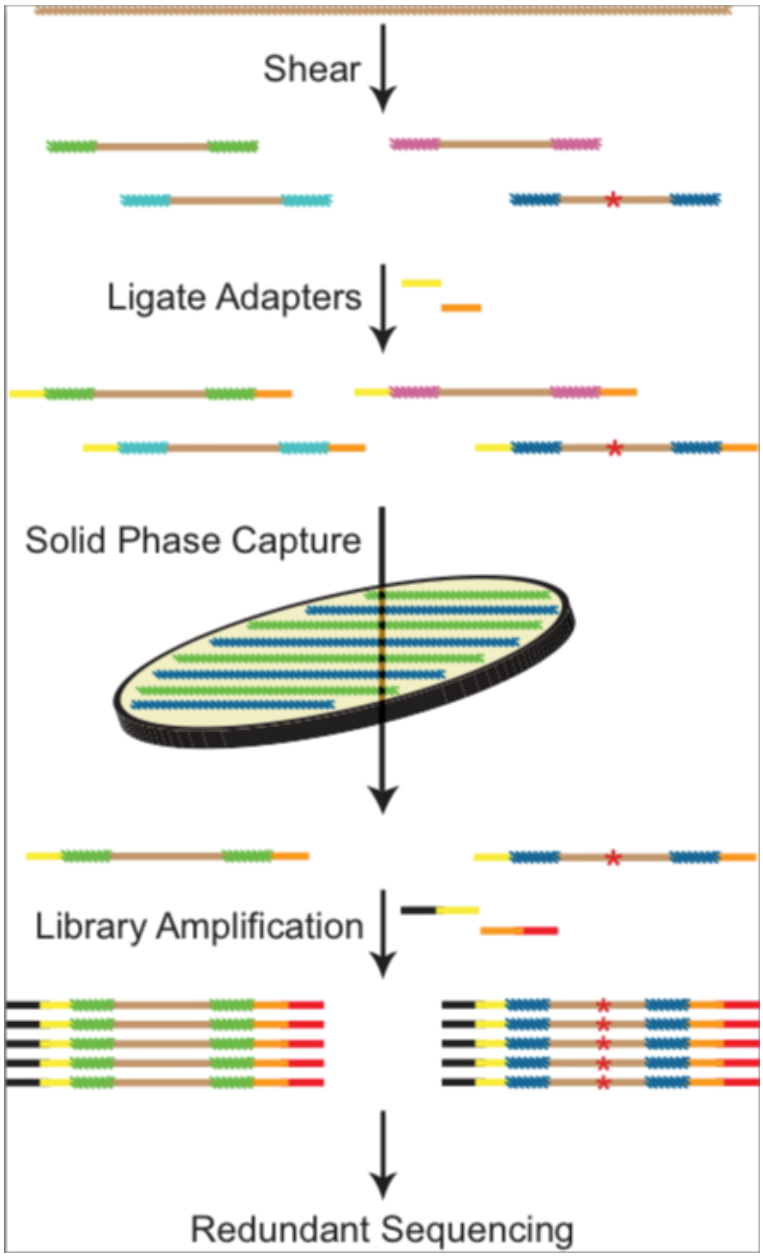


Figure 2-3. Safe-SeqS with Exogenous UIDs.

DNA (sheared or unsheared) is amplified with a set of gene-specific primers. One of the primers has a random DNA sequence (e.g., a set of 14 N's) that forms the unique identifier (UID; variously colored bars), located 5' to its gene-specific sequence, and both have sequences that permit universal amplification in the next step (yellow and orange bars). Two UID assignment cycles produce two fragments - each with a different UID - from each double-stranded template molecule, as shown. Subsequent PCR with universal primers, which also contain "grafting" sequences (black and red bars), produces UID-families which are directly sequenced. Super-mutants are defined as in the legend to Fig. 1.

Figure 2-3. Continued.

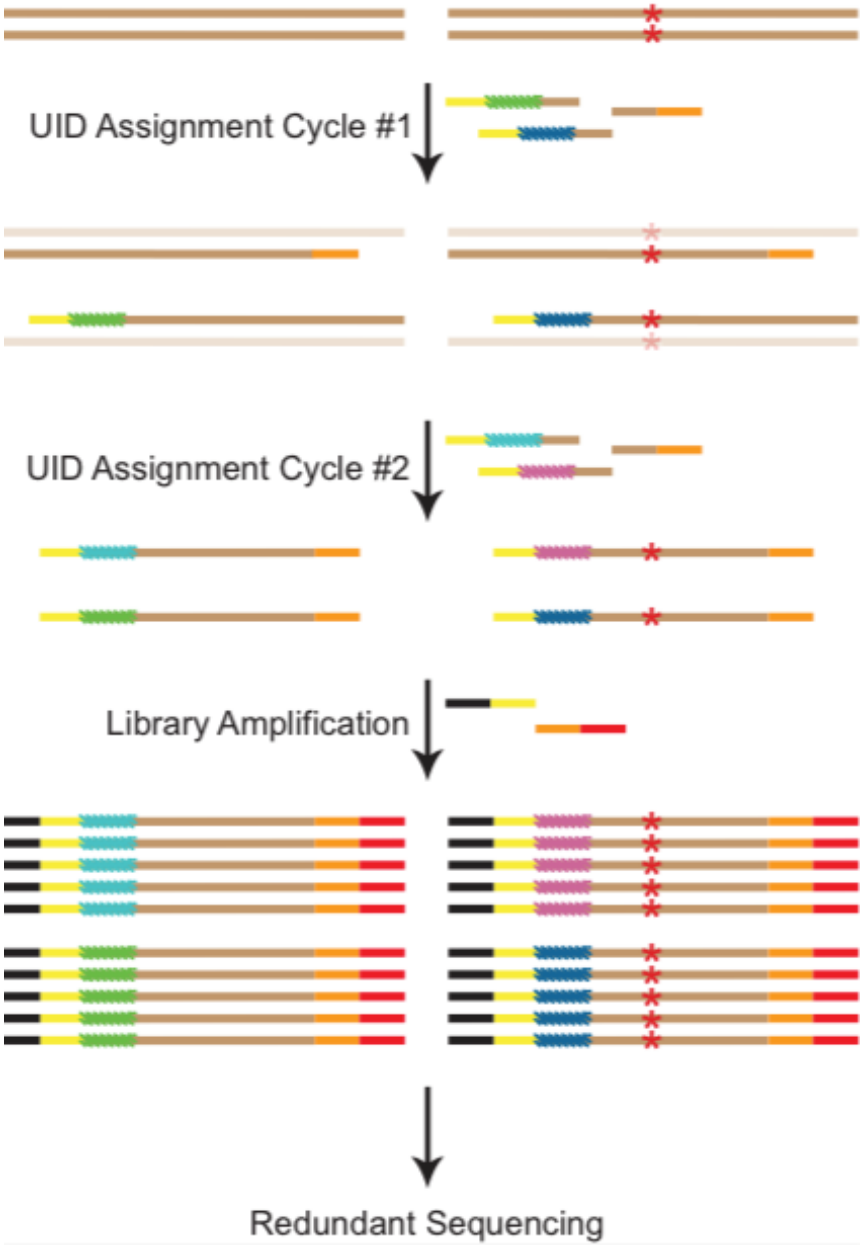


Figure 2-4. Single Base Substitutions Identified by Conventional and Safe-SeqS Analysis.

The exogenous UID strategy depicted in Fig. 3 was used to produce PCR fragments from the *CTNNB1* gene of three normal, unrelated individuals. Each position represents one of 87 possible single base substitutions (3 possible substitutions/base x 29 bases analyzed). These fragments were sequenced on an Illumina GA IIx instrument and analyzed in the conventional manner (A) or with Safe-SeqS (B). Safe-SeqS results are displayed on the same scale as conventional analysis for direct comparison; the inset is a magnified view. Note that most of the variants identified by conventional analysis are likely to represent sequencing errors, as indicated by their high frequency relative to Safe-SeqS and their consistency among unrelated samples.

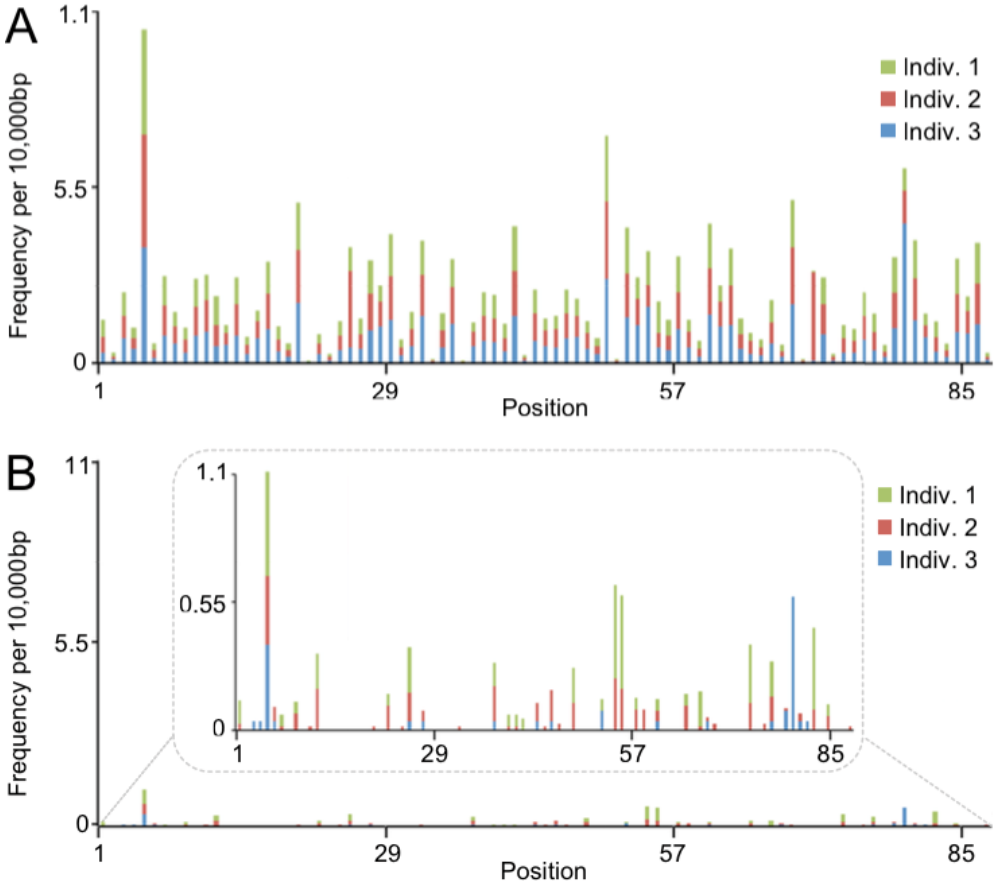


Figure 2-S1. Safe-SeqS with endogenous UIDs plus inverse PCR.

The sequence of the ends of each fragment produced by random shearing serve as unique identifiers (UIDs; variously colored bars). These fragments are ligated to adapters (yellow and orange bars) as in a standard Illumina library preparation. One uniquely tagged fragment is produced from each strand of the double-stranded template; only one strand is shown. Following circularization with a ligase, inverse PCR is performed with gene-specific primers that also contain 5' "grafting" sequences (black and red bars). This PCR produces UID-families which are directly sequenced. Super-mutants are defined as in Fig. 1.

Figure 2-S1. Continued.

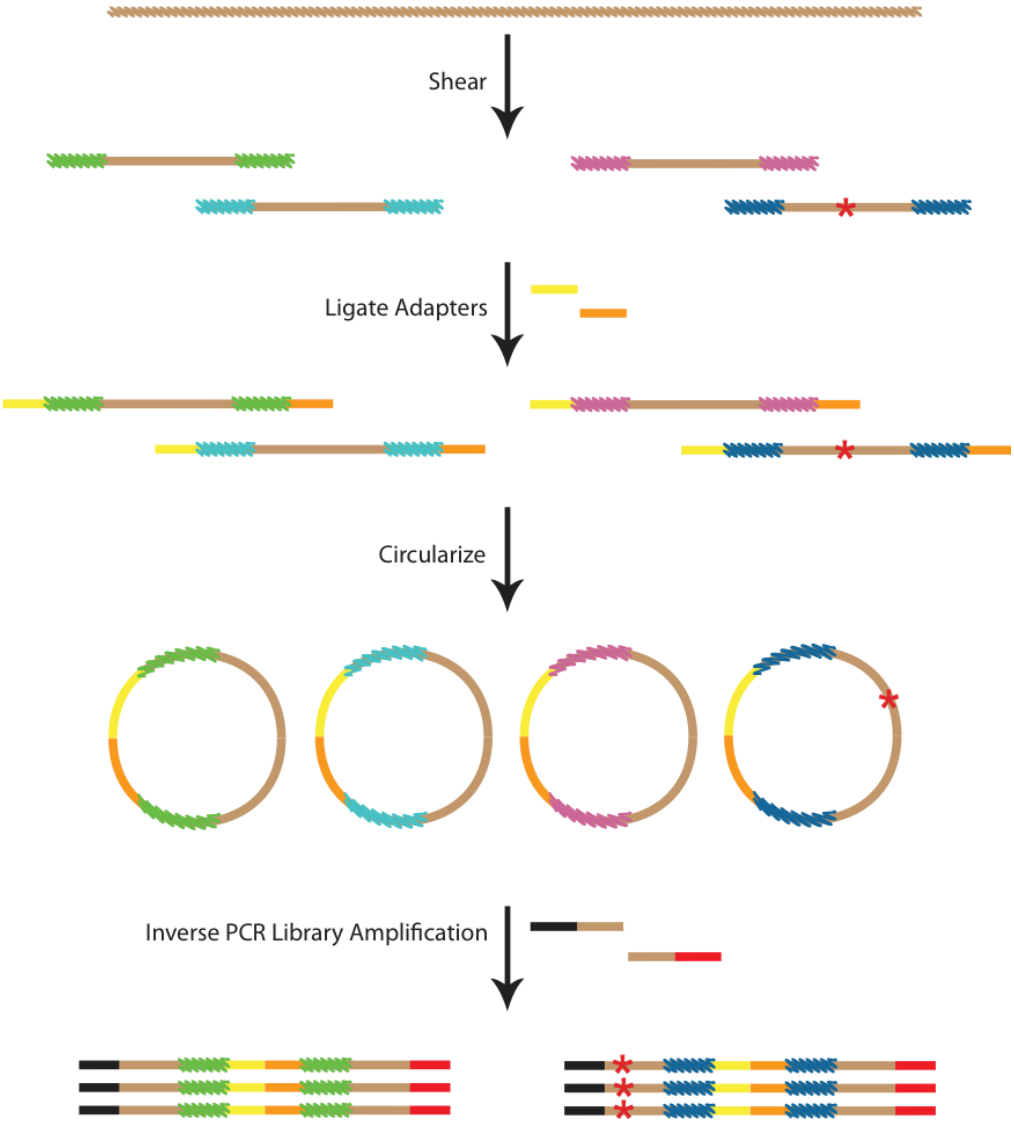


Figure 2-S2. Single base substitutions position vs. error frequency in oligonucleotides synthesized with phosphoramidites and Phusion

A representative portion of the same 31-base DNA fragment synthesized with phosphoramidites (A) or Phusion polymerase (B) was analyzed by Safe-SeqS. The means and standard deviations for seven independent experiments of each type are plotted. There was an average of 1,721 ± 383 and 196 ± 143 SBS super-mutants identified in the phosphoramidite-synthesized and Phusion-generated fragments, respectively. The y-axis indicates the fraction of the total errors at the indicated position. Note that the errors in the phosphoramidite-synthesized DNA fragment were consistent among the seven replicates, as would be expected if the errors were systematically introduced during the synthesis itself. In contrast, the errors in the Phusion-generated fragments appeared to be heterogeneous among samples, as expected from a stochastic process [1].

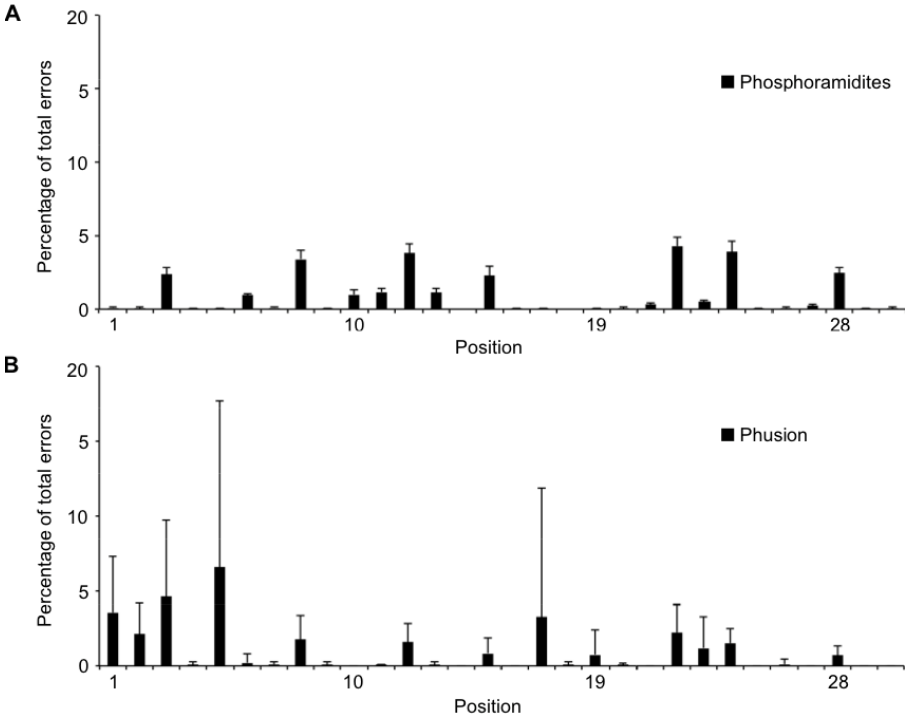
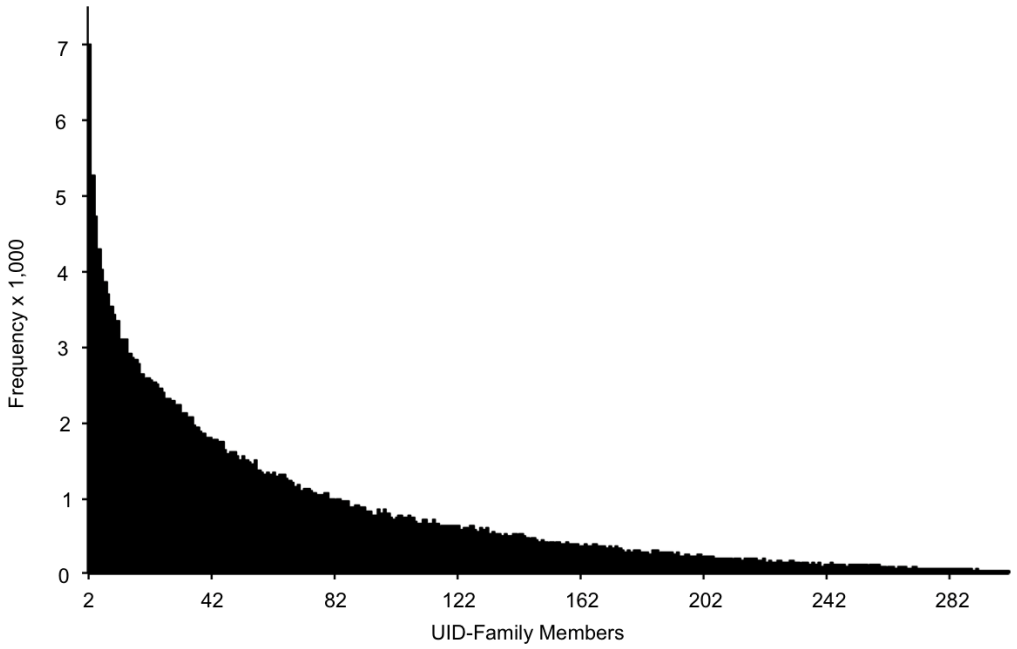


Figure 2-S3. UID-family member distribution.

The exogenous UID strategy depicted in Fig. 3 was used to produce PCR fragments from a region of *CTNNB1* from three normal, unrelated individuals (Table 2B); a representative example of the UID-families with ≤ 300 members (99% of total UID-families) generated from one individual is shown. The y-axis indicates the number of different UID-families that contained the number of family members shown on the x-axis.



Chapter 3: Evaluation of DNA From the Papanicolaou Test to Detect Ovarian and Endometrial Cancers

Introduction

Since the introduction of the Papanicolaou test, the incidence and mortality of cervical cancer in screened populations has been reduced by more than 75% [66,67]. In contrast, deaths from ovarian and endometrial cancers have not substantially decreased during that same time period. As a result, more than 69,000 women in the U.S. are estimated to be diagnosed with ovarian or endometrial cancer in 2012 [68]. Although endometrial cancer is more common than ovarian cancer, the latter is more lethal. In the U.S., approximately 15,000 and 8,000 women are expected to die each year from ovarian and endometrial cancers, respectively [68]. World-wide, over 200,000 deaths from these tumors are expected this year alone [69,70].

In an effort to replicate the success of cervical cancer screening, several approaches for the early detection of endometrial and ovarian cancers have been proposed. For endometrial cancers, efforts have focused on cytology and transvaginal ultrasound (TVS). Cytology can indeed indicate a neoplasm within the uterus in some cases, albeit with low specificity [71]. TVS is used to measure the thickness of the endometrium, because it is known that endometria harboring a cancer are thicker than normal endometria [72]. As with cytology, screening measurement of the endometrial thickness with TVS lacks sufficient specificity

because benign lesions, such as polyps, can also result in a thickened endometrium. Accordingly, neither cytology nor TVS fulfills the requirements for a screening test [71,73].

Even greater efforts have been made to develop a screening test for ovarian cancer, including the assessment of serum CA-125 levels in conjunction with TVS. CA-125 is a high molecular weight transmembrane glycoprotein expressed by coelomic- and Müllerian-derived epithelia that is elevated in a subset of ovarian cancer patients with early stage disease and in some cases prior to clinical diagnosis [74,75]. The specificity of CA-125 is limited by the fact that it is also elevated in a variety of benign conditions, such as pelvic inflammatory disease, endometriosis and ovarian cysts [76]. Although TVS can visualize the ovary, it can only detect large tumors and cannot definitively distinguish benign from malignant tumors. Several clinical screening trials with serum CA-125 and TVS have been conducted but none have shown a survival benefit. In fact, some have shown an increase in morbidity compared to controls because false positive tests elicit further evaluation by laparoscopy or exploratory laparotomy [77-79].

Accordingly, the U.S. Preventative Services Task Force, the American Cancer Society, the American Congress of Obstetricians and Gynecologists, and the National Comprehensive Cancer Network do not recommend routine screening for endometrial or ovarian cancers in the general population. In fact, these organizations warn that "the potential harms outweigh the potential benefits" [80-83]. An exception to this recommendation has been made for patients with a hereditary predisposition to ovarian cancer, such as those with germline mutations in a *BRCA* gene or those with Lynch syndrome. It is recommended that *BRCA* mutation carriers be screened every 6 months with TVS and serum CA-125, starting at a

relatively early age. Screening guidelines for women with Lynch syndrome include annual endometrial sampling and TVS beginning between age 30 and 35 [82,84].

The mortality associated with undetected gynecologic malignancies has made the development of an effective screening tool a high priority. An important observation that inspired the current study is that asymptomatic women occasionally present with abnormal glandular cells (AGCs) detected in a cytology specimen as part of their routine cervical cancer screening procedure. Although AGCs are associated with premalignant or malignant disease in some cases [85-89], it is often difficult to distinguish the AGCs arising from endocervical, endometrial, or ovarian cancer from one another and from more benign conditions.

We reasoned that more sophisticated molecular methods might be able to detect the presence of cancer cells in endocervical specimens at higher sensitivities and specificities than possible with conventional methods. In particular, we hypothesized that somatic mutations characteristic of endometrial and ovarian cancers would be found in the DNA purified from routine liquid-based Pap specimens (henceforth denoted as "Pap specimens"; Fig. 1). Unlike cytologically abnormal cells, such oncogenic DNA mutations are specific, clonal markers of neoplasia that should be absent in non-neoplastic cells. The experiments described here were carried out to test this hypothesis.

Results

There were four components to this study: I. Establishing the somatic mutations typically present in endometrial and ovarian cancers; II. Identifying at least one mutation in each

tumor from 46 patients with these cancers; III. Determining whether the mutations identified in these tumors could also be detected in Pap specimens from the same patients; and IV. Developing a technology that could directly assess cells from Pap specimens for mutations commonly found in endometrial or ovarian cancers.

Prevalence of somatically mutated genes in endometrial and ovarian cancers. There are five major histopathologic subtypes of epithelial ovarian cancers (Table 1). The most prevalent subtype is high-grade serous (60% of total), followed by endometrioid (15%), clear cell (10%), low-grade serous carcinoma (8%), mucinous (2%), and transitional cell carcinoma (2%) [90-92]. The majority of these cases were found at an advanced stage and the combined 5-year survival [68,93] for these malignancies is approximately 27% (Table 1). Genome-wide studies [94-96] have identified commonly mutated genes among the most prevalent ovarian cancer subtypes (Table 2).

Such comprehensive studies have not yet been reported for the endometrioid and mucinous subtypes, which collectively represent ~20% of ovarian cancer cases (Table 1). However, commonly mutated genes in the endometrioid and mucinous subtypes have been reported [97]. In aggregate, the most commonly mutated gene in epithelial ovarian cancers was *TP53*, which was mutated in 69% of these cancers (Table 2). Other highly mutated genes included *ARID1A*, *BRAF*, *CTNNB1*, *KRAS*, *PIK3CA*, and *PPP2R1A* (Table 2).

Among endometrial cancers, the endometrioid subtype is by far the most common, representing 85% of the total (Table 1). Because cancers of this subtype are so frequent and have not been analyzed at a genome-wide level, we evaluated them through whole-exome

sequencing. The DNA purified from 22 sporadic endometrioid carcinomas, as well as from matched non-neoplastic tissues, was used to generate 44 libraries suitable for massively parallel sequencing. The clinical aspects of the patients and histopathologic features of the tumors are listed in table S1. Though the examination of 22 cancers cannot provide a comprehensive genome landscape of a tumor type, it is adequate for diagnostic purposes, as these only require the identification of the most frequently mutated genes.

Among the 44 libraries, the average coverage of each base in the targeted region was 149.1 with 88.4% of targeted bases represented by at least ten reads. After applying stringent criteria for the identification of somatic mutations (as described in Materials and Methods), the sequencing data clearly demarcated the tumors into two groups: ten cancers (termed the N Group, for *n*on-highly mutated) harbored <100 somatic mutations per tumor (median 32, range 7 to 50), and 12 cancers (termed the H Group, for *h*ighly mutated) harbored >100 somatic mutations per tumor (median 674, range 164 to 4,629) (table S1).

The high number of mutations in the Group H tumors was consistent with a deficiency in DNA repair. Eight of the 12 Group H tumors had microsatellite instability (MSI-H, table S1), supporting this conjecture. Moreover, six of the Group H tumors contained somatic mutations in the mismatch repair genes *MSH2* or *MSH6*, but none of the Group N cancers contained mutations in mismatch repair genes (table S2). Mismatch repair deficiency is known to be common among endometrial cancers and these tumors occur in 19-71% of women with inherited mutations of mismatch repair genes (i.e., patients with Hereditary Nonpolyposis Colorectal Cancer) [98].

A complete list of the 12,795 somatic mutations identified in the 22 cancers is provided in table S2. The most commonly mutated genes included the PI3K pathway genes *PTEN* and *PIK3CA* [99], the APC pathway genes *APC* and *CTNNB1*, the fibroblast growth factor receptor *FGFR2*, the adapter protein *FBXW7*, and the chromatin-modifying genes *ARID1A* and *MLL2* (Table 2). Genes in these pathways were mutated in both Group N and H tumors. Our results are consistent with prior studies of endometrioid endometrial cancer that had evaluated small numbers of genes, though mutations in *FBXW7*, *MLL2* and *APC* had not been appreciated to occur as frequently as we found them. It was also interesting that few*TP53* mutations (5%) were found in these endometrial cancers (Table 2 and table S2), a finding also consistent with prior studies.

Papillary serous carcinomas of the endometrium account for 10-15% of endometrial cancers, and a recent genome-wide sequencing study of this tumor subtype has been published [100]. The most common mutations in this subtype are listed in Table 2. The least common subtype of endometrial cancers is clear cell carcinoma [101], which occur in <5% of cases. Genes reported to be mutated in these cancers were garnered from the literature (Table 2).

Identification of mutations in tumor tissues. We acquired tumors from 46 cancer patients for whom Pap specimens were available. These included 24 patients with endometrial cancers and 22 with ovarian cancers; their clinical, demographic and histopathologic features are listed in table S3.

Somatic mutations in the 46 tumors were identified through whole-exome sequencing as described above (table S2) or through targeted sequencing of genes frequently mutated in the

most common subtypes of ovarian or endometrial cancer (Table 2). Enrichment for these genes was achieved with a custom solid phase capture assay comprised of oligonucleotides ("capture probes") complementary to a panel of gene regions of interest. For the oncogenes, we only targeted their commonly mutated exons, whereas we targeted the entire coding regions of the tumor suppressor genes.

DNA sequencing libraries were generated from tumors and their matched non-neoplastic tissues, then captured with the assay described above. After amplification by PCR, four to eight captured DNA libraries were sequenced per lane on an Illumina GA IIx instrument. In each of the 46 cases, we identified at least one somatic mutation (table S3) that was confirmed by an independent assay, as described below.

Identification of somatic mutations in Pap specimens. In the liquid-based Pap smear technique in routine use today, the clinician inserts a small brush into the endocervical canal during a pelvic exam and rotates the brush so that it dislodges and adheres to loosely attached cells or cell fragments. The brush is then placed in a vial of fixative solution (e.g., ThinPrep). Some of the liquid from the vial is used to prepare a slide for cytological analysis or for purification of HPV DNA. In our study, an aliquot of the DNA purified from the liquid was assessed for the presence of DNA from the cancers of the 46 patients described above. Preliminary studies showed that the fixed cells or cell fragments in the liquid contained >95% of the total DNA in the vial. We therefore purified DNA from the cell pellets when the amount of available liquid was greater than 3 mL (as occurs with some liquid-based Pap smear kits) and, for convenience, purified DNA from both the liquid and cells when smaller amounts of liquid were in the kit. In all cases, the purified DNA was of

relatively high molecular weight (95% >5 kb). The average amount of DNA recovered from the 46 Pap specimens was 9.9 ± 14.8 μg (table S3).

We anticipated that, if present at all, the amount of DNA derived from neoplastic cells in the Pap smear fluid would be relatively small compared to the DNA derived from normal cells brushed from the endocervical canal. This necessitated the use of an analytic technique that could reliably identify a rare population of mutant alleles among a great excess of wild-type alleles. A modification of one of the Safe-SeqS (Safe-Sequencing System) procedures described in [11], in which DNA templates are amplified with modified gene-specific primers, was designed for this purpose (Fig. 2).

In brief, a limited number of PCR cycles was performed with a set of gene-specific primers. One of the primers contained 14 degenerate "N" bases (i.e., equal probability of being an "A", "C", "G", or "T") located 5' to its gene-specific sequence, and both primers contained sequences that permitted universal amplification in the next step. The 14 "N" bases formed unique identifiers (UID) for each original template molecule. Subsequent PCR products generated with universal primers were purified and sequenced. If a mutation preexisted in a template molecule, that mutation should be present in every daughter molecule containing that UID, and such mutations are called "supermutants" [11]. Mutations not occurring in the original templates, such as those occurring during the amplification steps or through errors in base calling, should not give rise to supermutants. The Safe-SeqS approach used here is capable of detecting 1 mutant template among 5,000 to 1,000,000 wild-type templates, depending on the amplicon and the position within the amplicon that is queried [11].

We designed Safe-SeqS primers (table S4) to detect at least one mutation from each of the 46 patients described in table S3. In the 24 Pap specimens from patients with endometrial cancers, the mutation present in the tumor was identified in every case (100%). The median fraction of mutant alleles was 3%, and ranged from 0.01% to 80% (Fig. 3 and table S3). Amplifications of DNA from non-neoplastic tissues were used as negative controls in these experiments to define the detection limits of each queried mutation. In all cases, the fraction of mutant alleles was significantly different from the background mutation levels determined from the negative controls (P <0.001, binomial test). There was no obvious correlation between the fraction of mutant alleles and the histopathologic subtype or the stage of the cancer (Fig. 3 and table S3).

In endometrial cancer cases PAP 041 and PAP 083, two mutations found in the tumor DNA were evaluated in the Pap specimens (table S3). In both cases, the mutations were identified in DNA from the Pap smear (table S3). Moreover, the ratios between the mutant allele fractions of the two mutations in the Pap specimens were correlated with those of the corresponding tumor samples. For example, in the Pap smear of case PAP 083 the mutant allele fractions for the *CTNNB1* and *PIK3CA* mutations were 0.143% and 0.064%, respectively - a ratio of 2.2 (0.14% to 0.064%). In the primary tumor from PAP 083, the corresponding ratio was 2.0 (79.5% to 39.5%).

Similar analysis of Pap smear DNA from ovarian cancer patients revealed detectable mutations in nine of the 22 patients (41%). The fraction of mutant alleles was smaller than in endometrial cancers (median of 0.49%, range 0.021% to 5.9%; see Fig. 3 and table S3). All but one of the cases with detectable mutations were epithelial tumors; the exception was a

dysgerminoma, a malignant germ cell tumor of the ovary (table S3). As with endometrial cancers, there was no statistically significant correlation between the fraction of mutant alleles and histopathologic criteria. However, most ovarian cancers are detected only at an advanced stage, and this was reflected in the patients available in our cohort.

A genetic test for screening purposes. The results described above document that mutant DNA molecules from most endometrial cancers and some ovarian cancers can be found in routinely collected Pap specimens. However, in all 46 cases depicted in Fig. 3, a specific mutation was known to occur in the tumor, and an assay was subsequently designed to determine whether that mutation was also present in the corresponding Pap specimens. In a screening setting, the presence and genotype of tumors would obviously not be known prior to evaluation by such a test. We therefore designed a prototype test based on Safe-SeqS that could assess several genes and could be used in a screening setting (Fig. 2).

This multiplexed approach included 50 primer pairs that amplified segments of 241 to 296 bp containing frequently mutated regions of DNA. The regions to be amplified were chosen from the results described in Section I and included exons from *APC*, *AKT1*, *BRAF*, *CTNNB1*, *EGFR*, *FBXW7*, *KRAS*, *NRAS*, *PIK3CA*, *PPP2R1A*, *PTEN*, *and TP53*. In control experiments, 46 of the 50 amplicons were shown to provide information on a minimum of 2,500 templates as the number of templates sequenced can be determined directly from sequencing with Safe-SeqS (Fig. 2). Given the accuracy of Safe-SeqS, this number was adequate to comfortably detect mutations existing in >0.1% of template molecules [11]. The regions covered by these 46 amplicons (table S5), encompassing 10,257 bp, were predicted to be able to detect at least one mutation in >90% of either endometrial or ovarian cancers.

This test was applied to Pap specimens of 14 cases - twelve endometrial and two ovarian - as well as 14 Pap specimens collected from normal women. The two ovarian cancers used were stage IA and IV. The endometrial cancers were stage I (n=10), stage II (n=1), and stage IV (n=1). The 14 cancer cases were arbitrarily chosen from those which had mutant allele fractions >0.1% (table S3) and therefore above the detection limit of the multiplexed assay. In all 14 Pap specimens from women with cancer, the mutation expected to be present (table S3) was identified (Fig. 4 and table S6). The fraction of mutant alleles in the multiplexed test was similar to that observed in the original analysis of the same samples where only one Safe-SeqS primer pair per amplicon was employed (table S3 and table S6). Importantly, no mutations were detected in the 14 Pap specimens from women without cancer (Fig. 4).

Discussion

Georgios Papanicolaou published his seminal work, entitled "Diagnosis of Uterine Cancer by the Vaginal Smear," in 1943 [102]. At that time, he suggested that endocervical sampling could in theory be used to detect not only cervical cancers but also other cancers arising in the female reproductive tract, including endometrial carcinomas. The research reported here moves us much closer to that goal. In honor of Papanicolaou's pioneering contribution to the field of early cancer detection, we have named the approach described herein as the "PapGene" test.

One of the most important developments over the last several years is the recognition that all human cancers are the result of mutations in a limited set of genes and an even more limited set of pathways through which these genes act [103]. The whole-exome sequencing

data we present, combined with previous genome-wide studies, provide a striking example of the common genetic features of cancer (Table 2). Through the analysis of particular regions of only 12 genes (table S5), we could detect at least one driver mutation in the majority of nine different gynecologic cancers (Table 1). Though several of these 12 genes were tumor suppressors, and therefore difficult to therapeutically target, knowledge of their mutational patterns provides actionable opportunities for cancer diagnostics.

The most important finding in this paper is that diagnostically useful amounts of cells or cell fragments from endometrial and ovarian cancers are present in the cervix and can be detected through molecular genetic approaches. Detection of malignant cells from endometrial and ovarian carcinomas in cervical cytology specimens is relatively uncommon. Microscopic examination cannot always distinguish them from one another, from cervical carcinomas, or from more benign conditions. In our study, 100% of endometrial cancers (n=24), even those of low grade, and 41% of ovarian cancers (n=22), shed cells into the cervix that could be detected from specimens collected as part of routine Pap specimens. This finding, in conjunction with technical advances allowing the reliable detection of mutations present in only a very small fraction of DNA templates, provided the foundation for the PapGene test.

This study provides proof-of-principle for endocervical DNA testing for gynecologic cancers, but there are important limitations that need to be addressed before this approach can be used in the clinic. The test, even in its current format, appears to be promising as a screening tool for endometrial cancer, as the data in Fig. 3 show that even the lowest stage endometrial cancers could be detected through the analysis of DNA in Pap specimens.

However, only 41% of ovarian cancers could be detected in Pap specimens, even when the mutations in their tumors were known. In eight of the nine Pap specimens from ovarian cancer patients that contained detectable mutations, the mutant allele fractions were >0.1% and therefore within the range currently detectable by PapGene testing (table S3). Further improvements in the technology could increase the technical sensitivity of the PapGene test and allow it to detect more ovarian cancers. One improvement would involve an increase in the number of potential gene targets assessed by the PapGene test. Development of an improved method of collection may be more important to improve sensitivity. The current liquid specimen is designed for the detection of cervical cancer and as such employs a brush that collects cells from the ectocervix and only minimally penetrates the endocervical canal. A small cannula introduced into the endometrial cavity, similar to the Pipelle endometrial biopsy instrument, could theoretically be used to obtain a more highly enriched sample of cells coming from the endometrium, fallopian tube and ovary [104]. Specificity must also be addressed further in the future. Although a greater number of healthy controls need to be evaluated, it is encouraging that none evaluated so far had detectable mutations. This result is consistent with the idea that mutation-based screening should be exquisitely specific because mutations should not be found in normal cells. As noted in the Introduction, specificity is a major limitation of current screening tests in general, and for ovarian cancer in particular.

The quantitative nature of the PapGene test also opens the possibility of using it to monitor the response to hormonal agents (e.g., progestins) when treating young women with low risk endometrial cancers. Some of these women choose to preserve fertility, undergoing medical

therapy rather than hysterectomy [105], and PapGene testing could be performed at regular intervals to monitor them for local cancer recurrence or progression.

Even if the tumors detected were advanced, detection of pre-symptomatic ovarian cancers could also be of benefit. It has been demonstrated that one of the most important prognostic indicators for ovarian cancer is the amount of residual disease after surgical debulking. Initially, debulking was considered optimal if the residual tumor was less than 2 cm. Subsequently, the threshold was reduced to 1 cm, and now surgeons attempt to remove any visible tumor. With each improvement in surgical debulking, survival has lengthened [106]. The earlier these advanced stage ovarian cancer are diagnosed, the lower the overall tumor burden and the better the chance at optimal debulking. Furthermore, it is possible that a small volume of tumor is likely to be more sensitive to cytotoxic chemotherapy than the large, bulky disease typical of symptomatic high-grade serous carcinoma.

An essential aspect of any screening approach is that it should be relatively inexpensive and easily incorporated into standard medical practice. Evaluation of HPV DNA is already part of routine Pap smear testing because HPV analysis increases the test's sensitivity [107,108]. The DNA purification component of the PapGene test is identical to that used for HPV, so this component is clearly feasible. The preparation of DNA, multiplex amplification, and sequencing constituting the PapGene test can be performed at a cost comparable to a routine HPV test in the U.S. today. Note that the increased sensitivity provided by the Safe-SeqS component of the PapGene test can be implemented on any next-generation sequencing instrument, not just those used in this study. With the reduction in the cost of next-generation sequencing expected in the future, PapGene testing should become even less expensive.

There are millions of Pap smear tests performed annually in the U.S. Could PapGene testing be performed on such a large number of specimens? We believe so, because the entire DNA purification and amplification process can be automated, just as it is for HPV testing. Though it may now seem unrealistic to have millions of these sophisticated sequence-based tests performed every year, it would undoubtedly have seemed unrealistic to have widespread, conventional Pap smear testing performed when Papanicolaou published his original paper in 1943 [102]. Even today, when many cervical cytology specimens are screened with automated technologies, at least two to eight percent of samples require evaluation by a skilled cytopathologist [109]. In contrast, the analysis of PapGene testing is done completely *in silico* and the read-out of the test is objective and quantitative.

In sum, these data highlight the high specificity of mutation-based diagnostics paired with the sensitivity of interrogating local-regional bodily secretions for tumor-derived DNA. PapGene testing has the capacity to increase the utility of conventional cytology screening through the unambiguous detection of DNA from endometrial and ovarian carcinomas, and lays the foundation for a new generation of screening tests.

Materials and Methods

Patient Samples. All samples for this study were obtained according to protocols approved by the Institutional Review Boards of The Johns Hopkins Medical Institutions (Baltimore, MD), Memorial Sloan Kettering Cancer Center (New York, NY), University of Sao Paulo (Sao Paulo, Brazil), and ILSbio, LLC (Chestertown, MD). Demographic, clinical and pathologic staging data were collected for each case. All histopathology was centrally re-

reviewed by board-certified pathologists. Staging was based on 2009 FIGO criteria [110]. Purified DNA from tumor and normal tissue as well as liquid-based Pap smears were quantified in all cases with qPCR, employing the primers and conditions previously described [111]. Unless otherwise indicated, all patient-related values are reported as mean ± 1 standard deviation. Additional details are provided in Supplementary Materials and Methods.

Microsatellite instability testing. Tumor samples were designated as follows: MSI-high if two or more mononucleotides varied in length compared to the germline DNA; MSI-low if only one locus varied; and microsatellite stable (MSS) if there was no variation compared to the germline. Pentanucleotide loci confirmed identity in all cases. Additional details are provided in Supplementary Materials and Methods.

Preparation and sequencing of captured Illumina DNA libraries. Preparation of Illumina genomic DNA libraries and selection for exomic DNA were performed according to the manufacturer's recommendations. Exomic capture was performed with the SureSelect Human Exome Kit V 4.0 (Agilent), while the custom solid phase capture assay was performed by modifications of previously described methods [112,113]. Paired-end sequencing with an Illumina GA IIx Genome Analyzer provided 2 x 75 base reads from each fragment. Known polymorphisms recorded in dbSNP Build 130 [114] in the sequence tags that passed filtering were removed from the analysis. Identification of high confidence mutations was performed as described previously [95]. Additional details are provided in Supplementary Materials and Methods.

Assessment of low-frequency mutations. Primers were designed as described previously [11] with Primer3 [115]. Sixty-six ng of templates were prepared for sequencing as described previously [11], with modifications that facilitated the amplification of multiple gene regions in a single well of a 96-well PCR plate. With the primers described in table S4, 66 ng of templates were amplified in two rounds of PCR (Fig. 2) for the single amplicon assays. The multiplexed assays were performed in similar fashion utilizing six independent amplifications – each containing 66 ng of DNA (i.e., ~400 ng total) – per sample with the primers described in table S5. High quality sequence reads were analyzed as previously described [11] by utilizing the quality scores generated by default, which indicate the probability that an individual base call was made in error [116]. The template-specific portion of the reads was matched to a reference sequence set with a custom script (available from the authors upon request). Additional details are provided in Supplementary Materials and Methods.

Supplementary Materials and Methods

Patient Samples. All samples for this study were obtained according to protocols approved by the Institutional Review Boards of The Johns Hopkins Medical Institutions (Baltimore, MD), Memorial Sloan Kettering Cancer Center (New York, NY), University of Sao Paulo (Sao Paulo, Brazil), and ILSbio, LLC (Chestertown, MD). Demographic, clinical and pathologic staging data were collected for each case. All histopathology was centrally re-reviewed by board-certified pathologists. Staging was based on 2009 FIGO criteria [110].

Fresh-frozen tissue specimens of surgically resected neoplasms of the ovary and endometrium were assessed for neoplastic cellularity by a board-certified pathologist. Serial frozen sections were used to guide the trimming of Optimal Cutting Temperature (OCT)

compound embedded frozen tissue blocks to enrich the fraction of neoplastic cells for DNA extraction.

Formalin-fixed paraffin embedded (FFPE) tissue samples were assessed by a board-certified pathologist (Propath LLC, Dallas, TX) for tumor cellularity and to demarcate area of high tumor cellularity. Tumor tissues from serial 10 micron sections on slides from the original tumor block were macrodissected with a razorblade to enrich the fraction of neoplastic cells for DNA extraction.

The source of normal DNA was matched whole blood or non-neoplastic normal adjacent tissue.

Liquid-based Pap smear specimens were collected with cervical brushes and transport medium from Digene HC2 DNA Collection Device (Qiagen) or ThinPrep 2000 System (Hologic) and stored according to the manufacturer's recommendations.

Unless otherwise indicated, all patient-related values are reported as mean ± 1 standard deviation.

DNA Extraction. DNA was purified from tumor and normal tissue as well as liquid-based Pap smears with an AllPrep kit (Qiagen) according to the manufacturer's instructions. DNA was purified from tumor tissue by adding 3 mL RLTM buffer (Qiagen) and then binding to an AllPrep DNA column (Qiagen) according to the manufacturer's protocol. DNA was purified from Pap smear liquids by adding five volumes of RLTM buffer when the amount

of liquid was less than 3 mL. When the amount of liquid was >3 mL, the cells and cell fragments were pelleted at 1,000 x g for five minutes and the pellets were dissolved in 3 mL RLTM buffer. DNA was quantified in all cases with qPCR, employing the primers and conditions previously described [111].

Microsatellite instability testing. Microsatellite instability was detected with the MSI Analysis System (Promega), containing five mononucleotide repeats (BAT-25, BAT-26, NR-21, NR-24 and MONO-27) and two pentanucleotide repeat loci, per the manufacturer's instructions. After amplification, the fluorescent PCR products were sized on an Applied Biosystems 3130 capillary electrophoresis instrument (Invitrogen). Tumor samples were designated as follows: MSI-high if two or more mononucleotides varied in length compared to the germline DNA; MSI-low if only one locus varied; and microsatellite stable (MSS) if there was no variation compared to the germline. Pentanucleotide loci confirmed identity in all cases.

Preparation of Illumina DNA libraries and capture for exomic sequencing. Preparation of Illumina genomic DNA libraries for exomic and targeted DNA captures was performed according to the manufacturer's recommendations. Briefly, 1-3 μg of genomic DNA was used for library preparation with the TruSeqDNA Sample Preparation Kit (Illumina). The DNA was acoustically sheared (Covaris) to a target size of ~200 bp. The fragments were subsequently end-repaired to convert overhangs into blunt ends. A single "A" nucleotide was then added to the 3' ends of blunt fragments to later prevent them from self-ligation; a corresponding "T" on the 3' end of adaptor molecules provided the complementary overhang. After ligation to adaptors, the library was amplified with 8-14

cycles of PCR to ensure yields of 0.5 and 4 μg for exomic and targeted gene captures, respectively.

Exomic capture was performed with the SureSelect Human Exome Kit V 4.0 (Agilent) according to the manufacturer's protocol, with the addition of TruSeq index-specific blocks in the hybridization mixture (AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-XXXXXX-ATCTCGTATGCCGTCTTCTGCTTGT, where the six base "XXXXXX" region denotes one of 12 sample-specific indexes).

Targeted gene enrichment. Targeted gene enrichment was performed by modifications of previously described methods [112,113]. In brief, targeted regions of selected oncogenes and tumor suppressor genes were synthesized as oligonucleotide probes by Agilent Technologies. Probes of 36 bases were designed to capture both the plus and the minus strand of the DNA and had a 33-base overlap. The oligonucleotides were cleaved from the chip by incubating with 3 mL of 35% ammonium hydroxide at room temperature for five hours. The solution was transferred to two 2 mL tubes, dried under vacuum, and redissolved in 400 μL of ribonuclease (RNase) – and deoxyribonuclease (DNase) –free water. Five microliters of the solution was used for PCR amplification with primers complementary to the 12-base sequence common to all probes: 5'-TGATCCCGCGACGA*C-3' and 5'-GACCGCGACTCCAG*C-3', with * indicating a phosphorothioate bond. The PCR products were purified with a MinElute Purification Column (Qiagen), end-repaired with End-IT DNA End-Repair Kit (Epicentre), and then purified with a MinElute Purification Column. The PCR products were ligated to form concatamers as described [112].

The major differences between the protocol previously described [112,113] and the one used in the present study involved the amplification of the ligated PCR products and the solid phase capture method as noted below. Biotinylated capture probes were prepared by amplifying 50 ng of ligated PCR products with the REPLI-g Midi Kit (Qiagen) supplemented with 2.5 nmol Biotin-16-dUTP (Roche) in a 27.5 μL reaction incubated at 30°C for 16 hours. After inactivating the polymerase by incubating at 65°C for three mins, the biotinylated capture probes were purified with QIAquick PCR Purification Columns. For capture, 4-5 μg of library DNA was incubated with 1 μg of the prepared probes in a hybridization mixture as previously described [112]. The biotinylated capture probes and captured library sequences were subsequently purified with 500 μg Dynabeads MyOne Streptavidin C1 (Invitrogen). After washing as per the manufacturer's recommendations, the captured sequences were eluted with 0.1 M NaOH and then neutralized with 1 M Tris-HCl (pH 7.5). Neutralized DNA was desalted and concentrated with a MinElute Purification Column in 20 μL. The eluate was amplified in a 100 μL Phusion Hot Start II (Thermo Scientific) reaction containing 1X Phusion HF buffer, 0.25 mM dNTPs, 0.5 μM each forward and reverse TruSeq primers, and 2 U polymerase with the following cycling conditions: 98°C for 30 s; 14 cycles of 98°C for 10s, 60°C for 30 s, and 72°C for 30 s; and 72°C for 5 min. The amplified pool containing enriched target sequences was purified with an Agencourt AMPure XP system (Beckman) and quantified with a 2100 Bioanalyzer (Agilent).

Next-generation sequencing and somatic mutation identification. After capture of targeted sequences, paired-end sequencing with an Illumina GA IIx Genome Analyzer provided 2 x 75 base reads from each fragment. The sequence tags that passed filtering were aligned to the human genome reference sequence (hg18) and subsequent variant-calling

analysis was performed with the ELANDv2 algorithm in the CASAVA 1.6 software (Illumina). Known polymorphisms recorded in dbSNP Build 130 [114] were removed from the analysis. Identification of high confidence mutations was performed as described previously [95].

Assessment of low-frequency mutations

Primer Design. We attempted to design primer pairs to detect mutations in the 46 cancers described in the text. Primers were designed as described previously [11] with Primer3 [115]. Sixty percent of the primer pairs amplified the expected fragments; in the other 40%, a second or third set of primer pairs had to be designed to reduce primer dimers or non-specific amplification.

Sequencing Library Preparation. Templates were prepared for sequencing as described previously [11], with modifications noted below that facilitated the amplification of multiple gene regions in a single well of a 96-well PCR plate. In brief, each strand of each template molecule was encoded with a 14 base unique identifier (UID) – comprised of degenerate "N" bases (equal probability of being an "A," "C," "G," or "T") - with two to four cycles of amplicon-specific PCR ("UID assignment PCR cycles," see Fig. 2). Both forward and reverse gene-specific primers contained universal tag sequences at their 5' ends, providing the primer binding sites for the second-round amplification, but only the forward primer contained the UID, which was positioned between the 5' universal tag and the 3' gene-specific sequences. Four "N" bases were additionally included in the reverse primer to facilitate sequencing analysis of paired-end libraries (table S4). The UID assignment PCR cycles were performed on 66 ng of DNA in a 50 μL reaction containing 1X Phusion HF

buffer, 0.25 mM dNTPs, 0.5 μM each of forward (containing 14 "N" bases) and reverse primers, and 2 U of Phusion Hot Start II Polymerase (Thermo Scientific). Carryover of residual UID-containing primers to the second-round amplification, which can complicate template quantification [11], was minimized through a 15 s exonuclease digestion at 37^0C to degrade unincorporated primers. In Kinde *et al.* [11], Exonuclease-I (Enzymatics) was chosen to eliminate the residual UID-containing primers, however we found that a different exonuclease – $RecJ_f$ (New England Biolabs) – followed by purification with AMPure XP beads (Beckman) and elution in 10 μL of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), more extensively removed the UID-containing primers and yielded more robust amplification products. The eluted templates were amplified in a second-round PCR with primers containing the grafting sequences necessary for hybridization to the Illumina GA IIx flow cell at their 5' ends (Fig. 2). The reverse amplification primer additionally contained an index sequence between the 5 'grafting and 3' universal tag sequences to enable the PCR products from multiple individuals to be simultaneously analyzed in the same flow cell compartment of the sequencer [11]. The second-round amplification reactions contained 1X Phusion HF buffer, 0.25 mM dNTPs, 0.5 μM each of forward and reverse primers, and 2 U of Phusion Hot Start II Polymerase in a total of 50 μL. After an initial heat activation step at 98^0 C for 2 minutes, twenty-three cycles of PCR were performed with the following cycling conditions: 98^0C for 10 s, 65^0C for 15 s, and 72^0C for 15 s. The multiplexed assay was performed in similar fashion utilizing six independent amplifications – each containing 66 ng of DNA (i.e.,~400 ng total) – per sample with the primers described in table S5. The PCR products were purified with AMPure XP beads and used directly for sequencing on either Illumina MiSeq or GA IIx instruments, with equivalent results.

Data Analysis. High quality sequence reads were analyzed as previously described [11]. Briefly, we selected reads that contained high quality basecalls in their UID region (i.e., the first 14 cycles) by utilizing the quality scores generated by default, which indicate the probability that an individual base call was made in error [116]. Reads in which each of the 14 bases comprising the UID (representing one original template strand; see Fig. 2) had a quality score ≥15 were grouped by their UIDs and only the UIDs supported by more than one read were retained for further analysis. The template-specific portion of the reads that contained the sequence of an expected amplification primer was matched to a reference sequence set with a custom script (available from the authors upon request). Artifactual mutations – introduced during the sample preparation or sequencing steps – were eliminated by requiring that >50% of reads sharing the same UID contained the identical mutation (a "supermutant," Fig. 2). For the 46 assays querying a single amplicon, we required that the fraction of mutant alleles was significantly different from the background mutation levels determined from a negative control (P <0.001, binomial test). As mutations are not known *a priori* in a screening environment, we used a more agnostic metric to detect mutations in the multiplexed assay. A threshold supermutant frequency was defined for each sample as equaling the mean frequency of all supermutants plus six standard deviations of the mean. Only supermutants exceeding this threshold were designated as mutations and reported in Fig. 4 and table S6.

Acknowledgments

This chapter first appeared in 10.1126/scitranslmed.3004952[117].

Table 3-1. Epidemiology of Ovarian and Endometrial Tumors in the United States.

Tissue	Type	Subtype	Subtype distribution	Estimated new cases, 2012	5-year survival
Ovarian	Epithelial	High-grade serous	60%	13,368	9%
		Endometrioid	15%	3,342	71%
		Clear cell	10%	2,228	62%
		Low-grade serous	8%	1,782	40%
		Mucinous	2%	446	65%
		Transitional cell	2%	446	57%
		Other	3%	668	N/A
Endometrial	Type I: Endometrioid	Endometrioid	85%	40,060	91%
	Type II: Non-Endometrioid	Papillary serous	10%	4,713	45%
		Clear cell	5%	2,357	68%

Table 3-2. Genetic characteristics of ovarian and endometrial cancers.

Tissue	Type	Subtype	Somatically mutated genes (frequency)
Ovarian	Epithelial	High-grade serous	*TP53* (96%)
			CSMD3 (6%)
			FAT3 (6%)
			BRCA1 (3%)
			BRCA2 (3%)
		Endometrioid	*TP53* (68%)
			ARID1A (30%)
			CTNNB1 (26%)
			PTEN (17%)
			PIK3CA (15%)
			KRAS (10%)
			PPP2R1A (11%)
			CDKN2A (12%)
			BRAF (8%)
		Clear cell	*ARID1A* (57%)
			PIK3CA (40%)
			PPP2R1A (7%)
			KRAS (4.7%)
		Low-grade serous	*BRAF* (38%)
			KRAS (19%)
		Mucinous	*TP53* (56%)
			KRAS (40%)
			PPP2R1A (33%)
			CDKN2A (16%)
			PTEN (11%)

Table 3-2. Continued

Tissue	Type	Subtype	Somatically mutated genes (frequency)
Endometrial	Type I: Endometriod	Endometrioid	*PTEN* (64%)
			PIK3CA (59%)
			ARID1A (55%)
			CTNNB1 (32%)
			MLL2 (32%)
			FBXW7 (27%)
			RNF43 (27%)
			APC (23%)
			FGFR2 (18%)
			KRAS (9%)
			PIK3R1 (9%)
			EGFR (14%)
			AKT1 (5%)
			NRAS (5%)
			TP53 (5%)
	Type II: Non-Endometrioid	Papillary serous	*TP53* (82%)
			PIK3CA (24%)
			FBXW7 (20%)
			PPP2R1A (18%)
		Clear Cell	*TP53* (45%)
			PPP2R1A (33%)
			PIK3CA (29%)
			PTEN (13%)
			PIK3R1 (9%)
			KRAS (5%)

Table 3-S1. Summary Characteristics of Endometrial Cancers (Endometrioid Subtype) Studied by Whole-Exome Sequencing.

Case #	Age	Stage (FIGO)	Pathologic stage (TNM class)	Number of mutations	Microsatellite stability status*
PAP 003	53	IB	T1bN0M0	847	MSS
PAP 010	73	IB	T1bN0M1	29	MSS
PAP 011	58	IB	T1bN0M2	579	MSI-H
PAP 024	56	IA	T1aN0	7	MSS
PAP 026	86	IA	T1a	769	MSI-H
PAP 030	73	IA	T1aNx	49	MSS
PAP 031	61	IA	T1aNx	41	MSS
PAP 032	82	IA	T1aNx	9	MSS
PAP 033	68	IA	T1aNX	34	MSS
PAP 034	55	IA	T1aN0	454	MSI-H
PAP 043	55	IB	T1BN0MX	26	MSS
PAP 045	57	IB	T1BN0MX	4,629	MSS
PAP 046	44	IIA	T2ANXMX	40	MSS
PAP 047	53	IA	T1AN0MX	1,767	MSS
PAP 048	62	IIIC	T2AN1MX	394	MSI-H
PAP 049	45	IIB	T2BN0MX	20	MSS
PAP 050	39	IB	T1BN0MX	50	MSS
PAP 052	70	IVB	T1AN1M1	164	MSI-H
PAP 053	66	IB	T1BN0MX	1,102	MSI-H
PAP 054	73	IA	T1ANXMX	413	MSI-H
PAP 055	61	IA	T1AN0MX	1,195	MSS
PAP 057	59	IIB	T2BN0MX	176	MSI-H

* MSI-H: micosatellite unstable; MSS: microsatellite stable. See Materials and Methods for additional details

Table 3-S2. Mutations Identified by Whole-Exome Sequencing in 22 Endometrioid Endometrial Cancers.

[Too large to display]

Full table available at http://stm.sciencemag.org/content/5/167/167ra4/suppl/DC1

Table 3-S3. Clinical Characteristics and Mutations Assessed in Pap Specimens

[Too large to display]

Full table available at http://stm.sciencemag.org/content/5/167/167ra4/suppl/DC1

Table 3-S4. Primers Used to Assess Individual Mutations in Pap Specimens.

Case #	Gene	Mutation	Amplicon length (bp)	Forward primer sequence	Reverse primer sequence
PAP 001	*PIK3CA*	p.H1047R	82	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGATCCAATCCATTTTTGTTGTCCAG	cacacaggaaacagctatgaccatgTGAGCAAGAGGCTTTGGAGT
PAP 002	*TP53*	p.V147D	93	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGGCCAAGACCTGCCCTG	cacacaggaaacagctatgaccatgTGCTGTGACTGCTTGTAGATGG
PAP 003	*APC*	p.R1450X	93	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCCACCTCCTCAAACAGCTCAA	cacacaggaaacagctatgaccatgTGCAGCTTGCTTAGGTCCACT
PAP 004	*TP53*	p.Q167fs*(1 base deletion of C)	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGGCCATCTACAAGCAGTCA	cacacaggaaacagctatgaccatgNNNNTCACCATCGCTATCTGAGCA
PAP 005	*NF1*	p.L414fs*(1 base insertion of C)	74	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCATTGGTGATGATTCGATGG	cacacaggaaacagctatgaccatgCTGCCTGGCTCAGAATTCAC
PAP 006	*TP53*	p.R280K	90	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCCCTTTCTTGCGGAGATTC	cacacaggaaacagctatgaccatgCTACTGGGACGGAACAGCTT
PAP 007	*TP53*	p.E294X	87	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGGAAGAGAATCTCCGCAAGA	cacacaggaaacagctatgaccatgGCTTCTTGTCCTGCTTGCTT
PAP 010	*FBXW7*	p.R479X	82	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGGCCTGTCTCAATATCCCAAA	cacacaggaaacagctatgaccatgTTGTTTTTCTGTTTCTCCCTCTG
PAP 011	*KRAS*	p.G12A	92	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCAAGGCACTCTTGCCTACG	cacacaggaaacagctatgaccatgCATTTTCATTATTTTTATTATAAGGCCTG
PAP 024	*CTNNB1*	p.G34V	72	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCTGTGGTAGTGGCACCAGAA	cacacaggaaacagctatgaccatgNNNNAAGCGGCTGTTAGTCACTGG
PAP 025	*TP53*	p.P98fs*(2 base deletion of CC)	91	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGGCCCCTGTCATCTTCTGT	cacacaggaaacagctatgaccatgGACTTGGCTGTCCCAGAATG
PAP 026	*PTEN*	p.R308H	99	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCAAGAAATCGATAGCATTTGCAG	cacacaggaaacagctatgaccatgTTTATTTGCTTTGTCAAGATCATTTTT
PAP 027	*SETD2*	p.Q269X	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNAGGAAATATCTGCTTGCTCATTCAA	cacacaggaaacagctatgaccatgGAAGCAGATACTAAGCAGGACACTATATC
PAP 030	*MSH6*	p.S718N	76	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTTCCCTTGGATTCTGACACA	cacacaggaaacagctatgaccatgAGCACCATTCGTTGATAGGC
PAP 031	*CTNNB1*	p.S37Y	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCACTGGCAGCAACAGTCTTACC	cacacaggaaacagctatgaccatgGATTGCCTTTACCACTCAGAGAAG
PAP 032	*PTEN*	p.R130G	71	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNttgcagcaattcactgtaaagc	cacacaggaaacagctatgaccatgccgatgtaataaatatgcacatatcattac
PAP 033	*KRAS*	p.G12V	92	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCAAGGCACTCTTGCCTACG	cacacaggaaacagctatgaccatgCATTTTCATTATTTTTATTATAAGGCCTG
PAP 034	*CTNNB1*	p.G34R	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCACTGGCAGCAACAGTCTTACC	cacacaggaaacagctatgaccatgGATTGCCTTTACCACTCAGAGAAG
PAP 035	*KRAS*	p.G12D	92	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCAAGGCACTCTTGCCTACG	cacacaggaaacagctatgaccatgCATTTTCATTATTTTTATTATAAGGCCTG
PAP 036	*PIK3CA*	p.Q546L	84	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNAGCTCAAAGCAATTTCTACACGA	cacacaggaaacagctatgaccatgNNNNGCACTTACCTGTGACTCCATAGAA
PAP 037	*PIK3CA*	p.T1025A	70	cgacgtaaaacgacggccagtNNNNNNNNNNNNNtcttttgatgacattgcatacattc	cacacaggaaacagctatgaccatgNNNNactccaaagcctcttGCTCA
PAP 038	*TP53*	p.Y220C	83	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNcagttgcaaaccagacctca	cacacaggaaacagctatgaccatgtgtggagtatttggatgacagaaa
PAP 039	*TP53*	p.I254fs*(1 base deletion of C)	76	cgacgtaaaacgacggccagtNNNNNNNNNNNNNgtggcaagtggctcctga	cacacaggaaacagctatgaccatgNNNNCATGGGCGGCATGAAC
PAP 040	*CTNNB1*	p.S45Y	90	cgacgtaaaacgacggccagtNNNNNNNNNNNNNactggcagcaacagtcttacct	cacacaggaaacagctatgaccatgNNNNCCTCAGGATTGCCTTTACCa
PAP 041	*FGFR2*	p.S252W	71	cgacgtaaaacgacggccagtNNNNNNNNNNNNNagtccggcttGGAGGATG	cacacaggaaacagctatgaccatgtccccactcctccttttcttc
	PTEN	p.C136R	78	cgacgtaaaacgacggccagtNNNNNNNNNNNNNggaaagGGACGAACTGGTG	cacacaggaaacagctatgaccatgNNNNtagggcctcttgtgcctttta
PAP 042	*TP53*	p.F113fs*(1 base insertion of T)	80	cgacgtaaaacgacggccagtNNNNNNNNNNNNNtcttctgtcccttcccagaa	cacacaggaaacagctatgaccatgNNNNgacttGGCTGTCCCAGAATG
PAP 058	*PIK3CA*	p.R108H	82	cgacgtaaaacgacggccagtNNNNNNNNNNNNNtcatACCAATTTCTCGATTGAGG	cacacaggaaacagctatgaccatgNNNNcggctttttcaaccctttt
PAP 060	*TP53*	p.R175H	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGGCCATCTACAAGCAGTCA	cacacaggaaacagctatgaccatgNNNNTCACCATCGCTATCTGAGCA
PAP 061	*TP53*	p.R273C	76	cgacgtaaaacgacggccagtNNNNNNNNNNNNNggGACGGAACAGCTTTGAG	cacacaggaaacagctatgaccatgNNNNgcggagattctcttcctctgt
PAP 062	*TP53*	p.M66fs*(4 base deletion of GCAT)	77	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCGGTGTAGGAGCTGCTGGT	cacacaggaaacagctatgaccatgNNNNACCCAGGTCCAGATGAAGC
PAP 063	*TP53*	p.R175H	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGGCCATCTACAAGCAGTCA	cacacaggaaacagctatgaccatgNNNNTCACCATCGCTATCTGAGCA
PAP 064	*TP53*	p.E258A	76	cgacgtaaaacgacggccagtNNNNNNNNNNNNNgtggcaagtggctcctga	cacacaggaaacagctatgaccatgNNNNCATGGGCGGCATGAAC
PAP 066	*TP53*	p.Y205C	74	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCGAAAAGTGTTTCTGTCATCCA	cacacaggaaacagctatgaccatgNNNNGCCCCTCCTCAGCATCTTAT
PAP 067	*TP53*	p.R213X	83	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNcagttgcaaaccagacctca	cacacaggaaacagctatgaccatgNNNNtgtggagtatttggatgacagaaa
PAP 068	*PIK3CA*	p.C420R	90	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTTATATTTCCCCATGCCAATG	cacacaggaaacagctatgaccatgNNNNggtgttttgaaatgtgttttataatttaga
PAP 069	*TP53*	p.R110C	80	cgacgtaaaacgacggccagtNNNNNNNNNNNNNtcttctgtcccttcccagaa	cacacaggaaacagctatgaccatgNNNNgacttGGCTGTCCCAGAATG
PAP 070	*PIK3CA*	p.R88Q	66	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGAAAAGCCGAAGGTCACAA	cacacaggaaacagctatgaccatgNNNNCTCAAGAAGCAGAAAGGGAAGA
PAP 071	*PIK3CA*	p.H1047R	82	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGATCCAATCCATTTTTGTTGTCCAG	cacacaggaaacagctatgaccatgNNNNTGAGCAAGAGGCTTTGGAGT
PAP 072	*TP53*	p.G245D	74	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGTGATGATGGTGAGGATGg	cacacaggaaacagctatgaccatgNNNNtccacTACAACTACATGTGTAACAGTTC
PAP 073	*TP53*	p.Y163C	84	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCTCCGTCATGTGCTGTGACT	cacacaggaaacagctatgaccatgNNNNcagctgtgggttgattcca
PAP 074	*TP53*	p.R175H	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGGCCATCTACAAGCAGTCA	cacacaggaaacagctatgaccatgNNNNTCACCATCGCTATCTGAGCA
PAP 075	*PIK3CA*	p.M1043I	79	cgacgtaaaacgacggccagtNNNNNNNNNNNNNccaatccatttttgttgtcca	cacacaggaaacagctatgaccatgNNNNTGAGCAAGAGGCTTTGGAGT
PAP 076	*TP53*	p.S127F	82	cgacgtaaaacgacggccagtNNNNNNNNNNNNNctgcacagggcaggtctt	cacacaggaaacagctatgaccatgNNNNctctgtctccttcctcttcctaca
PAP 078	*TP53*	p.R175H	85	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGGCCATCTACAAGCAGTCA	cacacaggaaacagctatgaccatgNNNNTCACCATCGCTATCTGAGCA
PAP 080	*PIK3CA*	p.A189fs*(1 base deletion of C)	78	cgacgtaaaacgacggccagtNNNNNNNNNNNNNcacgcaaATTTCCTTCCACT	cacacaggaaacagctatgaccatgNNNNggcctctgattcctcactgat
PAP 083	*PIK3CA*	p.H1047R	82	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGATCCAATCCATTTTTGTTGTCCAG	cacacaggaaacagctatgaccatgNNNNTGAGCAAGAGGCTTTGGAGT
	CTNNB1	p.S33Y	90	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNactggcagcaacagtcttacct	cacacaggaaacagctatgaccatgNNNNCCTCAGGATTGCCTTTACCa

Table 3-S5. Primers Used to Simultaneously Assess 12 Genes in Pap Specimens with the Multiplexed Safe-SeqS Strategy.

Set #	Gene (codons)	Amplicon length (bp)	Forward primer sequence	Reverse primer sequence
1	*CTNNB1* 5-64	251	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNAAAGTAACATTTCCAATCTACTAATGCT	cacacaggaaacagctatgaccatgNNNNtgagaaaatccctgttcccac
	EGFR 697-728	264	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGGAGCCTCTTACACCCAGT	cacacaggaaacagctatgaccatgNNNNAAAAACACTGGAGTTTCCCAAA
	PIK3CA 23-88	241	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNgaaaAAGCCGAAGGTCACAA	cacacaggaaacagctatgaccatgNNNNATGCCCCCAAGAATCCTAGT
	PTEN 1-26	257	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNcatccgtctactcccacgtt	cacacaggaaacagctatgaccatgNNNNatcagctaccgccaagtcc
	PTEN 85-123	261	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTTCGTCCCTTTCCAGCTTTA	cacacaggaaacagctatgaccatgNNNNggaatccagtgtttcttttaaatacct
	TP53 126-186	259	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNaccagccctgtcgtctctc	cacacaggaaacagctatgaccatgNNNNgccctgactttcaactctgtct
	TP53 308-331	259	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNgcctcagattcacttttatcacct	cacacaggaaacagctatgaccatgNNNNaccaggagccattgtctttg
2	*BRAF* 439-475	251	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTCACCACATTACATACTTACCATGC	cacacaggaaacagctatgaccatgNNNNaaggggatctcttcctgtatcc
	EGFR 738-761	279	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTCTGGATCCCAGAAGGTGAG	cacacaggaaacagctatgaccatgNNNNGGCCAGTGCTGTCTCTAAGG
	KRAS 1-21	281	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGTCCACAAAATGATTCTGAATTAGC	cacacaggaaacagctatgaccatgNNNNACGATACACGTCTGCAGTCAAC
	PIK3CA 62-117	269	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNACAGAAATATTTTAGAAAGGGACAACA	cacacaggaaacagctatgaccatgNNNNAGAAAATACCCCCTCCATCAAC
	PIK3CA 980-1049	272	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGATCCAATCCATTTTTGTTGTCCAG	cacacaggaaacagctatgaccatgNNNNTCCAAACTGACCAAACTGTTCTTA
	PTEN 100-164	294	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNgaaacccaaaatctgttttcca	cacacaggaaacagctatgaccatgNNNNGACCATAACCCACCACAGCTA
	TP53 186-224	251	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNctcctcccagagaccccagt	cacacaggaaacagctatgaccatgNNNNcatgagcgctgctcagatag
	TP53 332-366	263	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNcctaggaaggcaggggagt	cacacaggaaacagctatgaccatgNNNNtgcatgttgcttttgtaccg
3	*AKT1* 17-58	279	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNaccacccgcacgtctgtag	cacacaggaaacagctatgaccatgNNNNagccagtgcttgttgcttg
	EGFR 762-823	251	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNCACACTGACGTGCCTCTCC	cacacaggaaacagctatgaccatgNNNNttatctccccctcccgtatc
	NRAS 1-14	275	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGATTGTCAGTGCGCTTTTCC	cacacaggaaacagctatgaccatgNNNNgctaaggatgggggttgcta
	PIK3CA 272-347	279	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGACTTTACCTTATCAATGTCTCGAA	cacacaggaaacagctatgaccatgNNNNGCTCGCCCCCTTAATCTCT
	PPP2R1A 176-217	277	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGGTACTTCCGGAACCTGTGC	cacacaggaaacagctatgaccatgNNNNccgagtcctagggagaggag
	PTEN 56-70	250	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNttgttaatggtggctttttgttt	cacacaggaaacagctatgaccatgNNNNaaatgatctaacaatgctcttggac
	PTEN 165-211	265	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNcatggaaggatgagaatttcaag	cacacaggaaacagctatgaccatgNNNNtggctacgacccagttacca
	TP53 33-102	256	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNAAACCGTAGCTGCCCTGGTA	cacacaggaaacagctatgaccatgNNNNtgactgctcttttcacccatc
	TP53 225-260	257	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNtcatcttgggcctgtgttatc	cacacaggaaacagctatgaccatgNNNNgatgagaggtggatgggtagtagt
4	*EGFR* 830-875	258	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTTCAGGGCATGAACTACTTGG	cacacaggaaacagctatgaccatgNNNNatcctccccctgcatgtgtta
	KRAS 38-61	260	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTCCCTCATTGCACTGTACTCC	cacacaggaaacagctatgaccatgNNNNggtgcttagtggccatttgt
	NRAS 38-62	250	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNTGGTCTCTCATGGCACTGTACT	cacacaggaaacagctatgaccatgNNNNattagcaatttgagggacaaacc
	PTEN 71-84	250	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNtgtatctcactcgataatctggatgac	cacacaggaaacagctatgaccatgNNNNtgtcacattataaagattcaggcaat
	PTEN 213-267	271	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNagtttgacagttaaaggcatttcc	cacacaggaaacagctatgaccatgNNNNtgtccttatttttggatatttctcccc
	TP53 63-125	276	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGAAGACCCAGGTCCAGATGA	cacacaggaaacagctatgaccatgNNNNagaaatgcaggggatacg
	TP53 262-306	280	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNgaggcataactgcacccttg	cacacaggaaacagctatgaccatgNNNNgggagtagatggagcctggt
5	*APC* 1254-1328	269	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNtgctggatttggttctaggg	cacacaggaaacagctatgaccatgNNNNtgccacttgcaaagtttcttc
	APC 1403-1476	264	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNggaagaacctggaccctctg	cacacaggaaacagctatgaccatgNNNNttttgagagtcgttcgattgc
	APC 1537-1617	289	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNtgcaacctgttttgtgatgg	cacacaggaaacagctatgaccatgNNNNaggaaaatgacaatgggaatga
	FBXW7 474-548	296	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNtgattcatcaggagagcatttaag	cacacaggaaacagctatgaccatgNNNNttgtttttctgtttctccctctg
	FBXW7 558-618	289	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGATGGTATCCATGTGGTGAGTG	cacacaggaaacagctatgaccatgNNNNtttgtgatgctaaggctccat
	TP53 1-24	254	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNactgccttccgggtcact	cacacaggaaacagctatgaccatgNNNNagcccaacccttgtccttac
6	*TP53* 368-393	265	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNgaggctgtcagtggggaac	cacacaggaaacagctatgaccatgNNNNaacatatttgcatggggtgtg
	APC 1326-1405	283	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNccactgcatggttcactctg	cacacaggaaacagctatgaccatgNNNNatcctgtgagcgaagttcca
	APC 1475-1545	264	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNtgcctctttctcttggttttca	cacacaggaaacagctatgaccatgNNNNggacctaagcaagctgcagtaa
	FBXW7 243-287	257	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNacacccaatgaagaatgtaattgat	cacacaggaaacagctatgaccatgNNNNggttgtgtgtagatgtgagttttcc
	FBXW7 329-374	276	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNttctgttacattgtgcagagttca	cacacaggaaacagctatgaccatgNNNNtggttttgagcagagagatgg
	FBXW7 413-472	266	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNggaagaagtcccaaccatga	cacacaggaaacagctatgaccatgNNNNtcacttttccttctcacccaaaag
	FBXW7 549-593	259	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGAATCTGCATTCCCAGAGACA	cacacaggaaacagctatgaccatgNNNNcctgtttcccatcctcttcc
	KRAS 134-150	267	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNGACACAAAACAGGCTCAGGAC	cacacaggaaacagctatgaccatgNNNNagaaaccaaagccaaaagca
	TP53 26-32	253	cgacgtaaaacgacggccagtNNNNNNNNNNNNNNccatgggactgactttctgc	cacacaggaaacagctatgaccatgNNNNtcatctggacctgggtcttc

Table 3-S6. Mutations Identified in Pap Specimens through Simultaneous Assessment of 12 Genes.

Case #	Tumor type	Gene name	Gene ID	Genomic coordinate*	Transcript	Nucleotide	Amino acid change	Mutation type	Detected fraction
PAP 001	Endometrial	*KRAS*	ENSG00000133703	g.chr12:25289551C>G	CCDS8703.1	c.35G>C	p.G12A	Missense	13%
		PIK3CA	ENSG00000121879	g.chr3:180434779A>G	CCDS43171.1	c.3140A>G	p.H1047R	Missense	5.7%
PAP 003	Endometrial	*PIK3CA*	ENSG00000121879	g.chr3:180399570G>A	CCDS43171.1	c.263G>A	p.R88Q	Missense	12%
		APC	ENSG00000134982	g.chr5:112203538C>T	CCDS4107.1	c.4348C>T	p.R1450X	Missense	13%
		PTEN	ENSG00000171862	g.chr10:89614243A>C	CCDS31238.1	c.38A>C	p.K13T	Missense	12%
PAP 010	Endometrial	*FBXW7*	ENSG00000109670	g.chr4:153466817G>A	CCDS3777.1	c.1435C>T	p.R479X	Nonsense	20%
PAP 011	Endometrial	*KRAS*	ENSG00000133703	g.chr12:25289551C>G	CCDS8703.1	c.35G>C	p.G12A	Missense	3.2%
PAP 025	Endometrial	*TP53*	ENSG00000141510	g.chr17:7520119_7520120delGG	CCDS11118.1	c.292_293delCC	Frameshift	Indel	10%
PAP 033	Endometrial	*PIK3CA*	ENSG00000121879	g.chr3:180404243T>G	CCDS43171.1	c.1031T>G	p.V344G	Missense	1.2%
		KRAS	ENSG00000133703	g.chr12:25289551C>A	CCDS8703.1	c.35G>T	p.G12V	Missense	1.1%
		PTEN	ENSG00000171862	g.chr10:89707637insC	CCDS31238.1	c.682insC	Frameshift	Indel	0.87%
PAP 034	Endometrial	*PIK3CA*	ENSG00000121879	g.chr3:180399630G>A	CCDS43171.1	c.323G>A	p.R108H	Missense	23%
		CTNNB1	ENSG00000168036	g.chr3:41241107G>A	CCDS2694.1	c.100G>A	p. G34R	Missense	18%
		PTEN	ENSG00000171862	g.chr10:89707750delA	CCDS31238.1	c.795delA	Frameshift	Indel	13%
		PIK3CA	ENSG00000121879	g.chr3:180399554T>G	CCDS43171.1	c.247T>G	p.F83V	Missense	4.5%
		KRAS	ENSG00000133703	g.chr12:25289551C>T	CCDS8703.1	c.35G>A	p.G12D	Missense	0.92%
PAP 035	Endometrial	*KRAS*	ENSG00000133703	g.chr12:25289551C>T	CCDS8703.1	c.35G>A	p.G12D	Missense	5.8%
		PIK3CA	ENSG00000121879	g.chr3:180399548G>A	CCDS43171.1	c.241G>A	p.E81K	Missense	5.3%
		PTEN	ENSG00000171862	g.chr10:89675287T>C	CCDS31238.1	c.202T>C	p.Y68H	Missense	4.7%
PAP 039	Ovarian	*TP53*	ENSG00000141510	g.chr17:7518244delG	CCDS11118.1	c.871delC	Frameshift	Indel	0.73%
PAP 067	Endometrial	*TP53*	ENSG00000141510	g.chr17:7518937G>A	CCDS11118.1	c.637C>T	p.R213X	Nonsense	2.3%
PAP 069	Endometrial	*TP53*	ENSG00000141510	g.chr17:7520084G>A	CCDS11118.1	c.328C>T	p.R110C	Missense	19%
		TP53	ENSG00000141510	g.chr17:7517747G>A	CCDS11118.1	c.916C>T	p.R306X	Nonsense	14%
PAP 070	Endometrial	*PIK3CA*	ENSG00000121879	g.chr3:180399570G>A	CCDS43171.1	c.263G>A	p.R88Q	Missense	28%
PAP 071	Endometrial	*KRAS*	ENSG00000133703	g.chr12:25289551C>T	CCDS8703.1	c.35G>A	p.G12D	Missense	0.39%
		PIK3CA	ENSG00000121879	g.chr3:180434779A>G	CCDS43171.1	c.3140A>G	p.H1047R	Missense	0.31%
PAP 072	Ovarian	*TP53*	ENSG00000141510	g.chr17:7518272C>T	CCDS11118.1	c.734G>A	p.G245D	Missense	0.54%

*Coordinates refer to the human reference genome hg18 release (NCBI 36.1, March 2006).

Figure 3-1. Schematic of the PapGene test.

Tumor cells shed from ovarian or endometrial cancers are carried into the endocervical canal. These cells can be captured by the brush used for performing a routine Pap smear. The brush contents are transferred into a liquid fixative, from which DNA is isolated. By means of next-generation sequencing, this DNA is queried for mutations that indicate the presence of a malignancy in the female reproductive tract.

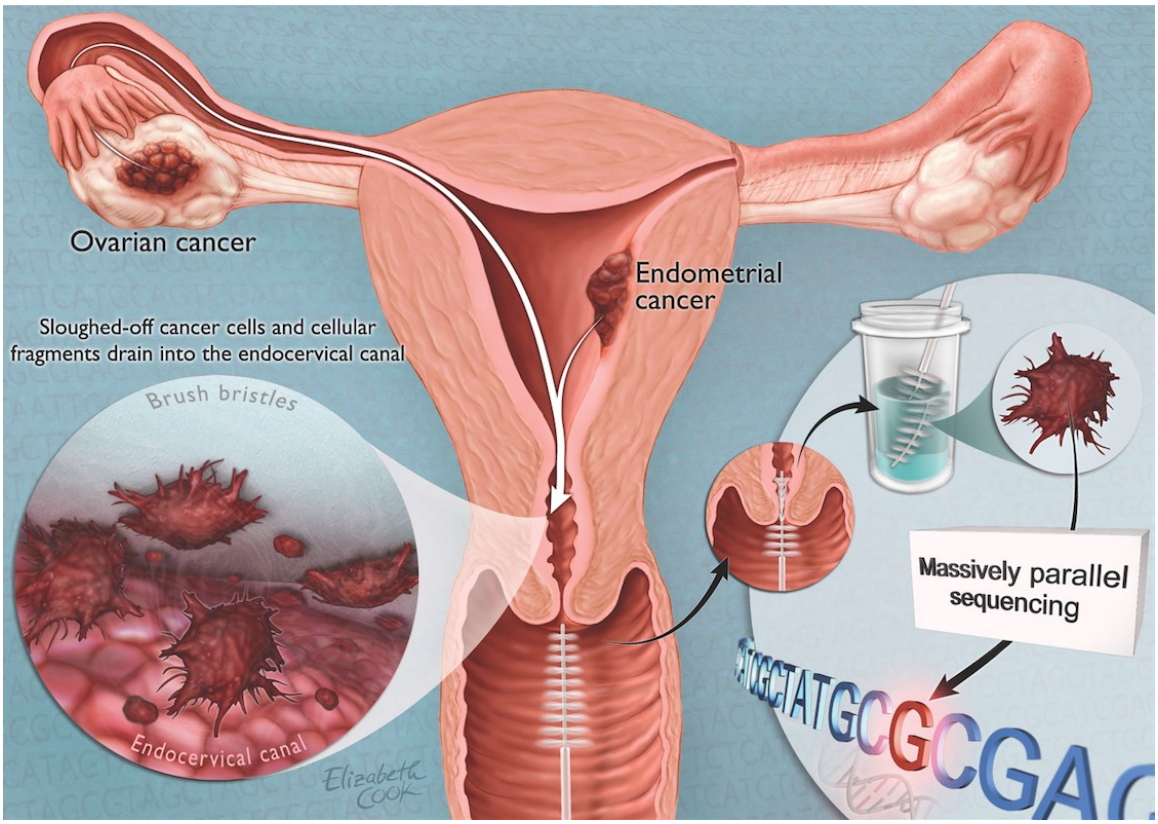


Figure 3-2. Diagram of the modified Safe-SeqS (Safe-Sequencing System) assay used allowing for the simultaneous detection of mutations in 12 different genes.

Top left: DNA templates from three exons of different genes (yellow, purple, and brown rectangles) to be queried for mutations. Note that only one of the templates contains a mutation (star) that exists before any sample preparatory steps or sequencing. Top right: Safe-SeqS primer pairs contain binding sites for universal primers ("UPS", blue), a unique identifier ("UID", red), and gene-specific sequences (colors match the targeted exon). Next, the templates and primers are combined into a single PCR compartment and a UID along, with UPS binding sites, are attached to each targeted template after a low number of PCR cycles ("UID assignment"). The Safe-SeqS primers are removed and subsequent PCR with universal primers, additionally containing the sequences required for attachment to the sequencing instrument ("GP", black), prepare the templates for next-generation sequencing. When mutations preexist in template DNA before sample preparation, all of the sequenced daughter molecules sharing the same UID will contain the same mutation (a "supermutant"). In contrast, artifactual mutations caused by sample preparation or sequencing are unlikely to be observed in most other daughter molecules sharing the same UID ("Artifact"). Note that only one of two DNA strands are depicted for clarity.

Figure 3-2. Continued.

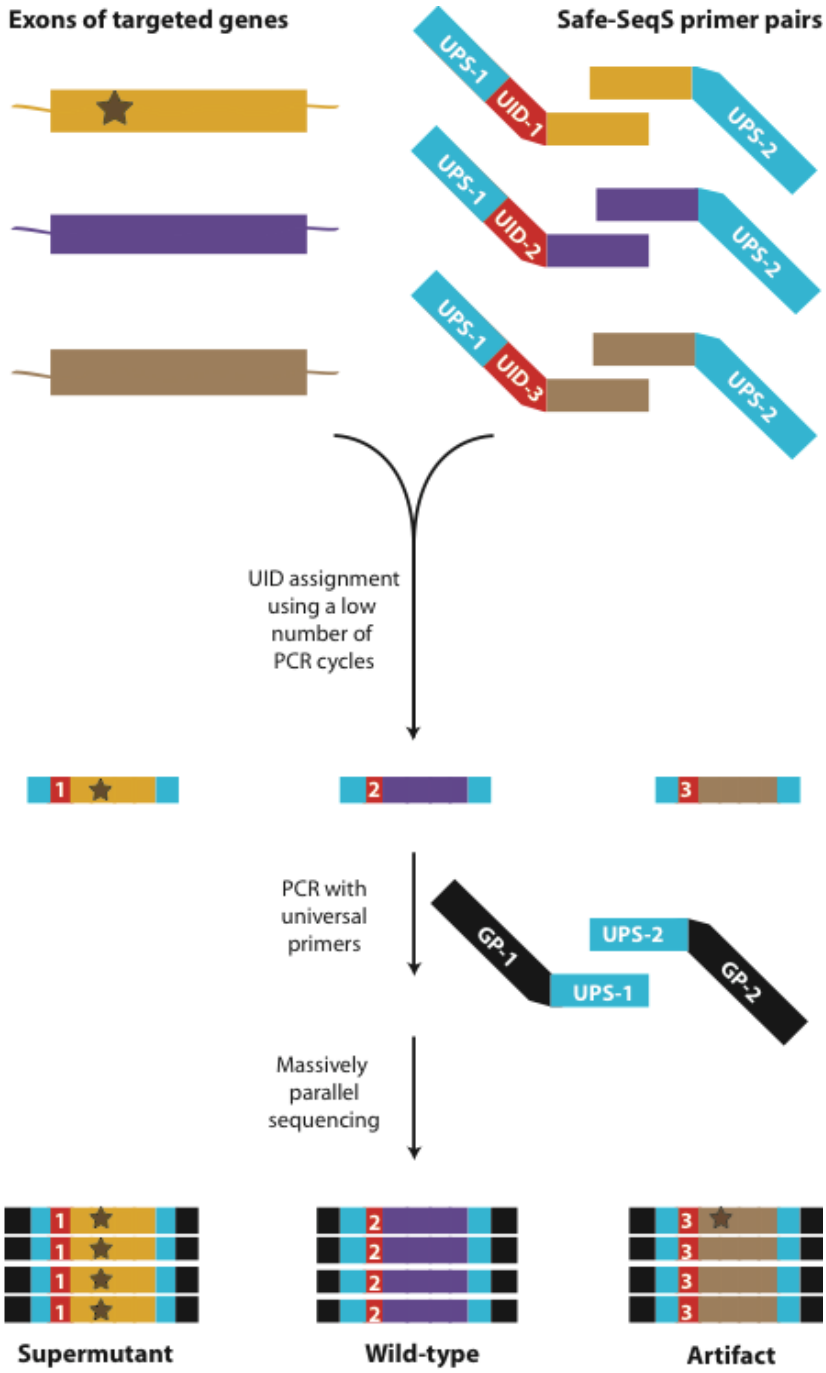


Figure 3-3. Mutant allele fractions in Pap smear fluids.

The fraction of mutant alleles from each of 46 Pap smear fluids is depicted. The stage of each tumor is listed on the Y-axis. The X-axis demonstrates the % mutant allele fraction (cut off <10%) as determined by traditional Safe-SeqS. Mutant allele frequencies are higher than 10% in some cases but are depicted at 10% in this figure for clarity. Precise mutation frequencies are reported in table S3 for all samples.

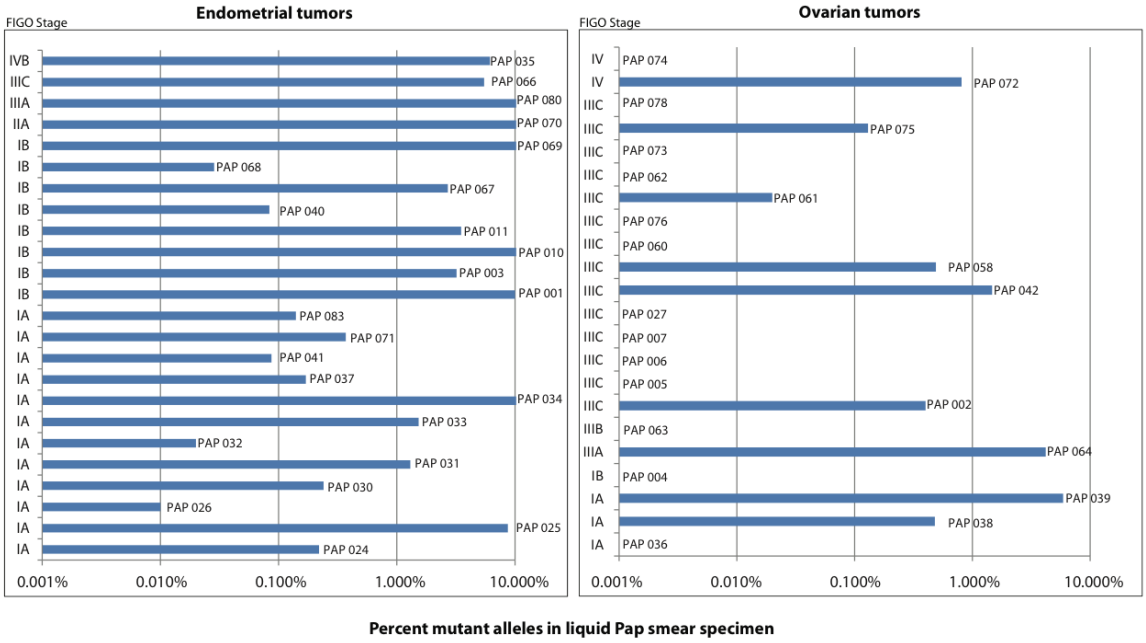
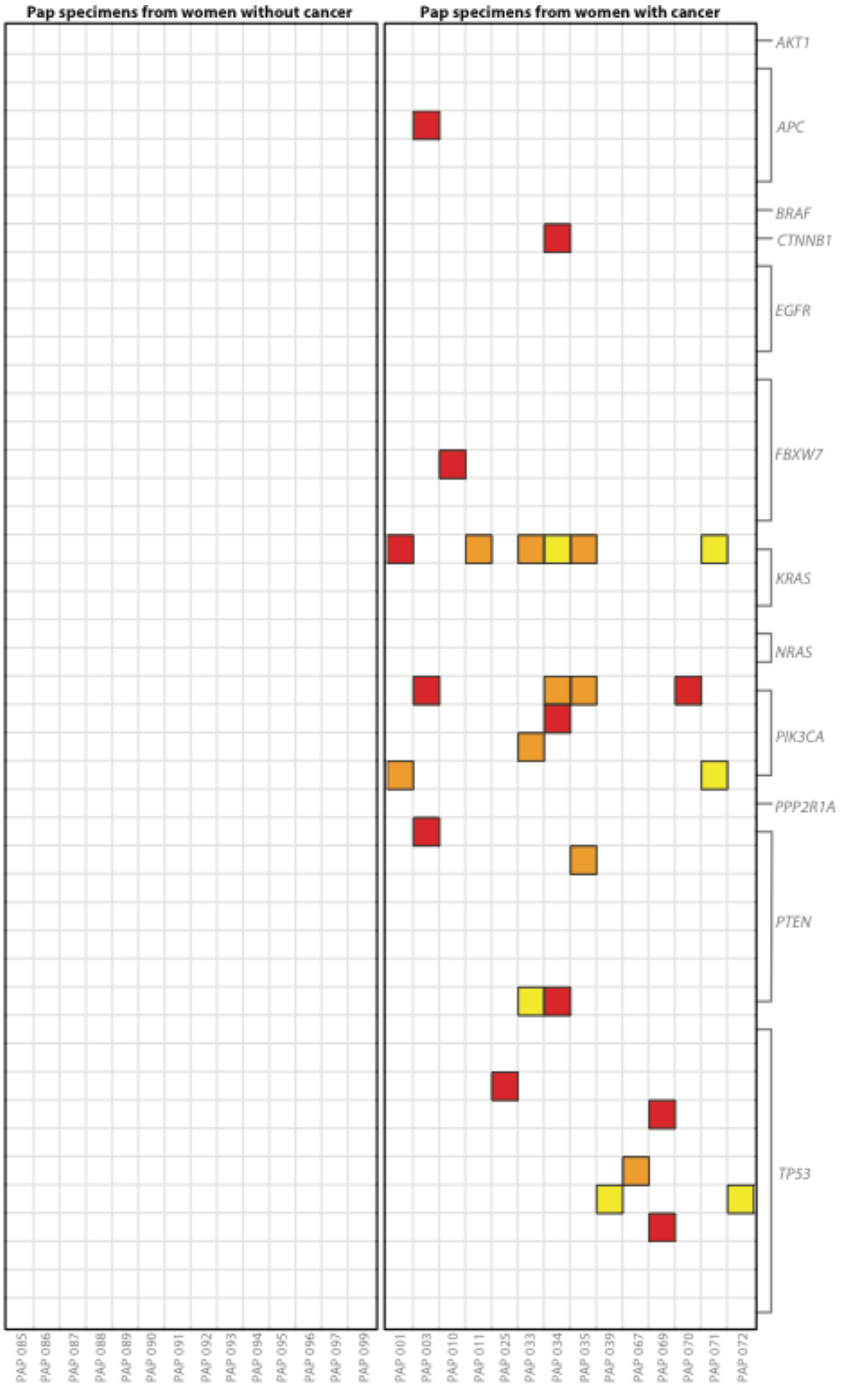


Figure 3-4. Heat map depicting the results of multiplex testing of 12 genes in Pap smear fluids.

The PapGene test interrogates 46 gene regions with each block on the Y-axis representing one region analyzed for the indicated gene. The 28 samples assessed (14 from healthy women without cancer, 14 from women with cancer) are indicated on the X-axis. Mutations are indicated as colored blocks, with white indicating no mutation, yellow indicating a mutant fraction of 0.1% to 1%, orange indicating a mutant fraction of 1% to 10%, and red indicating a mutant fraction of >10%.

Figure 3-4. Continued.



Chapter 4: TERT Promoter Mutations Occur Early in Urothelial Neoplasia and Are Biomarkers of Early Disease and Disease Recurrence in Urine

Introduction

Urothelial carcinoma of the bladder is the most common malignancy of the urinary tract with 73,000 new cases and 15,000 deaths expected in 2013 in the US alone [118]. These invasive carcinomas arise from histologically well-defined papillary and flat precursor lesions, providing a potential opportunity for early detection and treatment [119]. Although urine cytology enjoys a reasonable sensitivity and specificity for detecting high-grade neoplasms, its performance in detecting low-grade tumors is poor, with a sensitivity and specificity of 4% and 30%, respectively [120].

A number of urine-based markers have been developed to improve the accuracy of noninvasive screening and surveillance in bladder cancer. Among Food and Drug Administration (FDA) approved tests, the Immunocy test (Scimedx Corp, Danville, NJ), nuclear matrix protein 22 (NMP22) immunoassay test (Matritech, Cambridge, MA) and multitarget fluorescence in situ hybridization (FISH) (UroVysion; Abbott Park, IL) [121] have demonstrated an overall sensitivity of 70% and a specificity range of up to 89%. Performance inconsistencies, as a result of variability in pre-analytical and analytical specimen factors, have impeded their wide-spread clinical use.

Activating mutations in the promoter of the *telomerase reverse transcriptase* (*TERT*) gene lead to increased telomerase expression and, in doing so, allow some neoplasms to overcome the end-replication problem and avoid senescence. *TERT* promoter mutations were initially described in melanoma [122,123] and have subsequently been described in a discrete spectrum of cancer types, including 66% of muscle-invasive urothelial carcinomas of the bladder [122,124]. *TERT* is therefore the most frequently mutated gene in advanced forms of this disease, and the localization of these mutations to a small gene region in the *TERT* promoter provides an extraordinary opportunity for biomarker development [124].

For *TERT* promoter mutations to be a useful marker of early, curable disease, these mutations should be present in pre-invasive bladder tumors and shed into the urine. To this end, we have in this study evaluated the sequence of the *TERT* promoter in a large number of curable precursor neoplasms of the urinary bladder. We also determined the sequence of the *TERT* promoter in a separate group of superficial bladder cancers and corresponding follow-up urine samples to establish the feasibility of detecting *TERT* mutations in urine and their potential utility in predicting recurrence.

Materials and Methods

Patient Samples. This study was approved by the Institutional Review Board of Johns Hopkins University, School of Medicine. Two different sets of samples were analyzed in our study. The first sample set included 76 noninvasive papillary urothelial carcinomas and flat carcinoma *in situ* (CIS) lesions obtained by transurethral bladder resection (TURB) between 2000 and 2012. All specimens were rom the Surgical Pathology archives and were selected only on the basis of specimen availability. Pertinent patient demographics and clinical

information were obtained from electronic medical records. All sections were reviewed by three urological pathologists (EM, SFF and GJN) to confirm the original diagnoses. To enrich for neoplastic cells within the tissues, representative formalin-fixed paraffin-embedded (FFPE) blocks were cored with a sterile 16 gauge needle and tumor areas showing at least 50% neoplastic cellularity were selected microscopically. For eight of the cases, benign adjacent urothelium was macrodissected from FFPE blocks. The cores were placed in a 1.5 mL sterile tube for subsequent DNA purification using an AllPrep DNA/RNA Mini Kit (Qiagen, cat. no. 80204). DNA was purified from peripheral blood buffy coats of 15 patients using the same Qiagen kit.

For the second sample set, we prospectively collected urine samples from 15 separate patients undergoing follow-up cystoscopy for previously diagnosed non-muscle-invasive urothelial carcinoma. We purposely biased this cohort to include patients that recurred within the follow-up period. Immediately prior to follow-up cystoscopy, 25 mL of raw urine was collected and subsequently pelleted by centrifugation at 3,000 *g* for 10 minutes. The pellets were stored at -80° C in 1.5 mL tubes for subsequent DNA extraction. For 14 of these patients, matched FFPE from the original diagnostic TURB was retrieved. These included 13 high-grade urothelial carcinomas (pTa HG and pT1 HG in six and seven cases, respectively), and one low-grade papillary urothelial carcinoma (pTa LG). Twenty 8 μm-thick sections were cut from one representative tissue block in each case and areas containing at least 70% neoplastic cells were microdissected and used for DNA purification using a QIAamp DNA FFPE Tissue Kit (Qiagen, cat no. 56404).

Mutation analysis. Due to their tremendous throughput, massively parallel sequencing instruments are highly cost-effective for DNA mutation analysis. However, sample preparation and sequencing steps introduce artifactual mutations into analyses at a low, but significant frequency. To better discriminate genuine *TERT* promoter mutations from artifactual sequencing variants introduced during the sequencing process, we used Safe-SeqS, a sequencing error-reduction technology described previously [11,117]. As depicted in Fig. 1, Safe-SeqS amplification primers were designed to amplify a 126-bp segment containing the region of the *TERT* promoter previously shown to harbor mutations in melanomas and other tumors [122-124]. The forward and reverse amplification primers contained the *TERT*-specific sequences at their 3' ends and a universal priming site (UPS) at their 5' end. The reverse primer additionally contained a 14-base unique identifier (UID) comprised of 14 degenerate N bases (equal likelihood of being an A, C, T, or G) between the UPS and gene-specific sequences. The sequences of the forward and reverse primers were either 5'-CACACAGGAAACAGCTATGACCATGGGCCGCGGAAAGGAAG and 5'-CGACGTAAAACGACGGCCAGTNNNNNNNNNNNNNNCGTCCTGCCCCTTCACC, or CACACAGGAAACAGCTATGACCATGGCGGAAAGGAAAGGGAG and 5'-CGACGTAAAACGACGGCCAGTNNNNNNNNNNNNNNCCGTCCCGACCCCTC (UPS sequences underlined). These primers were used to amplify DNA in 25 μL PCR reactions in 1X Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific, cat. no. F-548L) containing 0.5 μM forward and reverse primers (described above). After incubation at 98ºC for 120 seconds, 10 cycles of PCR were performed in the following manner: 98ºC for 10 seconds, 63ºC for 120 seconds, and 72ºC for 120 seconds was performed. Reactions were purified with AMPure XP beads (Beckman Coulter) and eluted in 100 μL of Buffer EB (Qiagen, cat. no. 19086). For the second stage of amplification, 5 μL of purified PCR

products were amplified in 25 μL reactions containing 1X Phusion Flash High-Fidelity PCR Master Mix and 0.5 μM amplification primers that each contained the first-stage UPS at their 3' ends and the grafting sequences required to hybridize to the sequencing instrument flow cell at their 5' ends [11,117]. The reverse amplification primer additionally included a 6 bp index sequence, unique to each sample, inserted between the UPS and grafting sequences. After incubation at 98ºC for 120 seconds, 17 cycles of PCR were performed in the following manner: 98ºC for 10 seconds, 63ºC for 120 seconds, and 72ºC for 120 seconds. The PCR products were purified with AMPure and sequenced on a MiSeq instrument.

Data were analyzed as previously described [11,117]. Briefly, the amplified *TERT* promoter region of reads containing UIDs, where each base of the UID region had instrument-derived quality scores ≥15, was matched to a reference sequence using a custom script. *TERT* promoter sequences with five or fewer mismatches were retained for further analysis. Tumor samples were considered positive if the fraction of mutations exceeded 1% of alleles (which was a frequency at least 10x higher than found in control DNA templates from FFPE tissues). Urine samples were considered positive when the frequency of mutation exceeded 0.1% of alleles (a frequency at least 10x higher than found in control DNA templates from urine samples of patients without *TERT* mutations in their primary tumors). All sequencing assays scored as positive were confirmed in at least one additional, independent PCR and sequence assay.

Statistical analysis. The data were analyzed using Stata/SE 12 (StataCorp Inc., College Station, TX). Pearson's chi-squared test was used for analysis of association of categorical variables. A two-tailed probability <0.05 was required for statistical significance.

Results

TERT promoter mutation in papillary and "flat" noninvasive urothelial carcinoma. We used a massively parallel sequencing technology to determine the presence and representation of mutant *TERT* promoter alleles in urothelial cancers. A graphical depiction of the method is shown in Fig. 1 and detailed procedures are provided in the Materials and Methods. In addition to revealing whether mutations are present with a population of DNA templates, this technique provides an accurate determination of the fraction of mutant alleles in the sample. Clinicopathologic characteristics of the 76 noninvasive urothelial carcinomas analyzed in the first phase of this study are summarized in Table 1. They included 59 papillary tumors – 28 low-grade (pTa LG) and 31 high-grade (pTa HG) – plus 17 "flat" urothelial carcinoma *in situ* (CIS). These patients were typical of those with this form of cancer; their average age was 66 years and most (82%) were males (Table 1).

TERT promoter mutations were identified in 56/76 (74%) of these urothelial carcinomas (Table 2). In contrast, none of the eight samples of adjacent normal urothelium harbored *TERT* promoter mutations. Additionally, we did not detect *TERT* promoter mutations in 15 samples of peripheral blood from the same patients. Twelve of the blood samples and five of the normal urothelial samples were from patients whose tumors harbored *TERT* promoter mutations. These data demonstrate that the *TERT* promoter mutations in these patients were unequivocally somatic and limited to the neoplastic urothelium in the bladder. The predominant alterations were g.1295228C>T (minus strand of chromosome 5, hg19 assembly) and g.1295250C>T mutations, which accounted for 75% and 20% of the total alterations, respectively. In addition, we identified one g.1295228C>A mutation and two

g.1295242C>T mutations not previously reported (Table S1). The mutations were found in all types and grades of these early cancers: in 76% of papillary lesions and 65% of flat lesions; in 86% of low-grade and in 68% of high-grade lesions (Table 2). None of these differences among subgroups were statistically significant.

The results described above show that *TERT* promoter mutations occur early in bladder cancers and did not correlate with grade or type. Such early mutations would not be likely associated with recurrence or progression, but to evaluate this possibility, our series of samples included cases both with and without recurrence during follow up. In Tables 3 and 4, the relationship between *TERT* promoter mutation status and tumor recurrence or progression, respectively, are displayed: *TERT* promoter mutation status was not associated with likelihood of recurrence or progression in any subgroup.

***TERT* promoter mutation in urine samples.** We next evaluated whether *TERT* promoter mutations could be identified in cells in the urine. As noted in the Introduction, urine samples are routinely taken at follow-up visits following TURB procedures to help determine whether residual tumor cells are present (via cytology or other methods). We first assessed the tumors obtained from 14 patients undergoing TURB for relatively early (non-muscle invasive) disease. Of these, 11 (79%) harbored *TERT* promoter mutations (Table 5), as expected from the evaluation of the first cohort (Table 2). All of the mutations in the second cohort were at either g.1295228C>T or g.1295250C>T (Table 5).

The 14 patients were monitored for recurrence at subsequent visits. Mutations were assessed in the cell pellets from the urines obtained at the first follow-up visit after TURB in these 14 patients, as described in the Materials and Methods. There was a striking correlation between

the presence of a *TERT* promoter mutation in the urine, the presence of the mutation in the original tumor, and recurrence. In the three of 14 patients without a *TERT* promoter mutation in their tumor, no mutation was evident in their urine sample, as expected (Table 5). Of the 11 patients in whom a *TERT* mutation was present in the tumor, seven patients were observed to have a mutation in the DNA isolated from their urine cell pellets; in each case, the mutation was identical to that observed in the primary tumor removed via TURB (Table 5). The bladder cancers in each of these seven patients recurred, either at the first follow-up or thereafter. The proportion of mutant alleles in the cells pelleted from the urine of these patients was often substantial, ranging from 0.17% to 23% with a median of 4.4% (Table 5). We also identified a *TERT* promoter mutation in a urine sample from which no prior tumor was available; this tumor also recurred (Table 5). In contrast, no *TERT* mutations were evident in the urine samples of four patients whose original tumors contained a *TERT* promoter mutation: the tumors of three of these patients never recurred while the fourth developed a recurrence 3.5 months after the urine sample was collected (Table 5). As shown in Table 6, the presence of detectable *TERT* promoter mutations in the urine was strongly associated with recurrence of urothelial carcinoma ($P < 0.001$; Pearson's correlation coefficient =0.87).

Discussion

TERT promoter mutations are detectable in urine, and their presence in urine is strongly associated with bladder cancer recurrence. Muscle-invasive urothelial carcinoma is responsible for the vast majority of bladder cancer related deaths and many of these deaths could be prevented if precursor lesions were detected and surgically excised prior to their invasion into the muscle [125-128]. New strategies for the early detection of such lesions are

therefore urgently needed [9]. Our results show that *TERT* promoter mutations are the most common genetic alteration in noninvasive bladder cancer identified to date, occurring in the majority (74%) of such precursor lesions. They occur in cancers developing through both the papillary and flat routes to tumor progression [129], and occur in low-grade as well as high-grade tumors. We also show that these mutations can be detected in the urine of patients with bladder cancer. Altogether, these results suggest that *TERT* promoter mutations may provide a useful biomarker for the early detection of bladder cancers in the future, and that prospective studies of patients at high risk for this disease are warranted.

Given the high prevalence of *TERT* promoter mutations in early bladder neoplasia, their presence or absence in tumors is of limited prognostic value. However, superficial bladder cancers are currently the most costly solid tumor (per patient) in the US [130,131]. Noninvasive methods to monitor these patients could reduce the cost of caring for these patients as well as the discomfort associated with invasive procedures. Our results are highly encouraging with respect to this potential application. Among patients with *TERT* mutations in their primary tumors, there was a highly significant correlation between the presence of mutations in subsequent urine collections and recurrence (Table 6).

Our results therefore suggest two potential avenues for application of *TERT* promoter mutations in the clinic: early detection in high-risk patients and monitoring of patients with bladder cancer, both through the analysis of urine specimens. It is important to note that both these applications will require further study prior to implementation. For example, we have not yet shown that bladder cancer patients have detectable mutations in urine prior to tumor diagnosis; all of our urine samples were taken at follow-up visits after surgery.

Additionally, our study involved only a small number of patients, and we have yet to demonstrate that the analysis of urine for *TERT* mutations improves upon conventional cytology or clinical criteria, nor whether it could partially replace cystoscopy in certain circumstances. Still, our study provides a strong proof-of-principle: *TERT* promoter mutations occur early, are specific for neoplasia, and can be identified in the urine with currently available technologies. Future large-scale studies will be required to determine the clinical utility of this approach for screening or monitoring purposes.

Acknowledgments

This chapter first appeared in 10.1158/0008-5472.CAN-13-2498[132].

Table 4-1. Clinicopathologic characteristics of patients analyzed in this study

	pTa LG ($N = 28$)[a]	pTa HG ($N = 31$)[a]	CIS ($N = 17$)	Total
Age, mean (range)	65.5 (46–84)	67.6 (18–86)	65.6 (54–80)	66.4 (18–86)
Male (%)	73%	84%	94%	82%
Tumor recurred (%)	18/28 (64%)	17/29 (59%)	11/17 (65%)	46/68 (68%)
Tumor progressed (%)	6/28 (21%)	5/29 (17%)	4/17 (24%)	15/68 (22%)
Follow-up months, median (range)	56.5 (2–103)	40 (1–136)	18 (2–43)	38 (1–136)

Abbreviations: HG, high-grade papillary noninvasive urothelial carcinoma; LG, low-grade papillary noninvasive urothelial carcinoma; CIS, "flat" urothelial carcinoma *in situ* tumor recurred, tumors recurred within indicated follow-up; tumor progressed, the recurrent tumor had progressed with respect to stage or grade.

[a]Recurrence or progression status was not available in 2 cases.

Table 4-2. *TERT* promoter mutations

TERT promoter mutation	pTa LG (*N* = 28)	pTa HG (*N* = 31)	CIS (*N* = 17)	*P*
Present (%)	24/28 (86%)	21/31 (68%)	11/17 (65%)	0.18

Abbreviations: HG, high-grade papillary noninvasive urothelial carcinoma; LG, low-grade papillary noninvasive urothelial carcinoma; CIS, "flat" urothelial carcinoma *in situ*.

Table 4-3. Correlation between *TERT* promoter mutation status and tumor recurrence.

	Recurrence on follow-up	Number of patients	*TERT* mutation present (%)	*TERT* mutation absent (%)	*P*
pTa LG	Yes	17	16/17 (94%)	1/17 (6.0%)	0.21
($N = 28$)[a]	No	9	7/9 (78%)	2/9 (22%)	
pTa HG	Yes	17	13/17 (77%)	4/17 (24%)	0.14
($N = 31$)[a]	No	12	6/12 (50%)	6/12 (50%)	
CIS	Yes	11	7/11 (64%)	4/11 (36%)	0.90
($N = 17$)	No	6	4/6 (67%)	2/6 (33%)	

Abbreviations: HG, high-grade papillary noninvasive urothelial carcinoma; LG, low-grade papillary noninvasive urothelial carcinoma; CIS, "flat" urothelial carcinoma *in situ*.

[a]Recurrence status was not available in 2 cases.

Table 4-4. Correlation of *TERT* promoter mutation status and tumor progression

	Progression on follow-up	Number of patients	*TERT* mutation present (%)	*TERT* mutation absent (%)	*P*
pTa LG	Yes	6	6/6 (100%)	0/6 (0%)	0.31
($N = 28$)	No	22	17/22 (77%)	3/22 (14%)	
pTa HG	Yes	5	4/5 (80%)	1/5 (20%)	0.45
($N = 31$)[a]	No	24	15/24 (63%)	9/24 (38%)	
CIS	Yes	4	3/4 (75%)	1/4 (25%)	0.60
($N = 17$)	No	13	8/13 (62%)	5/13 (39%)	

Abbreviations: HG, high-grade papillary noninvasive urothelial carcinoma; LG, low-grade papillary noninvasive urothelial carcinoma; CIS, "flat" urothelial carcinoma *in situ*; NA, not available.

[a]Progression status was not available in 2 cases.

Table 4-5. Correlation of *TERT* mutation status in original diagnostic transurethral resection biopsy (TURB) tissue and *TERT* mutation status in urine collected at follow-up.

Patient	Original diagnostic TURB mutation (%)	Follow-up urine mutation (%)	Tumor grade	Recurrence at time of urine collection	Recurrence after urine collection
1	g.1295228C>T (11%)	g.1295228C>T (6.3%)	pTa HG	Yes	NA
2	g.1295250C>T (4.1%)	g.1295250C>T (23%)	pTa HG	Yes	No
3	g.1295228C>T (5.9%)	g.1295228C>T (0.17%)	pT1 HG	Yes	NA
4	Absent	Absent	pT1 HG	No	No
5	Absent	Absent	pT1 HG	No	No
6	g.1295228C>T (6.7%)	g.1295228C>T (0.64%)	pT1 HG	Yes	NA
7	g.1295228C>T (8.7%)	g.1295228C>T (4.6%)	pTa LG	Yes	Yes
8	g.1295228C>T (7.8%)	Absent	pTa HG	No	Yes
9	g.1295228C>T (7.0%)	Absent	pT1 HG	No	No
10	g.1295228C>T (5.1%)	Absent	pTa HG	No	NA
11	g.1295228C>T (4.8%)	Absent	pT1 HG	No	No
12	g.1295228C>T (5.5%)	g.1295228C>T (5.1%)	pT1 HG	Yes	Yes
13	Unknown	g.1295228C>T (6.6%)	Unknown	Yes	Yes
14	Absent	Absent	pTa HG	No	No
15	g.1295250C>T (23%)	g.1295250C>T (0.69%)	pT1 HG	Yes	Yes

NOTE: Genomic coordinates refer to the minus (−) strand of chromosome 5 (hg19 assembly).
Abbreviations: HG, high-grade papillary noninvasive urothelial carcinoma; LG, low-grade papillary noninvasive urothelial carcinoma; NA, not applicable.

Table 4-6. Correlation of *TERT* promoter mutation status in follow-up urine samples with recurrence.

TERT mutation in follow-up urine	Number of patients	Recurred	Did not recur	*P*
Present	8	8/8 (100%)	0/8 (0%)	<0.001
Absent	7	1/7 (11%)	6/7 (89%)	($r = 0.87$)[a]

[a]Pearson coefficient of correlation.

Table 4-S1. *TERT* promoter mutation status in 59 pTa and 17 carcinoma *in situ* (CIS) patients.

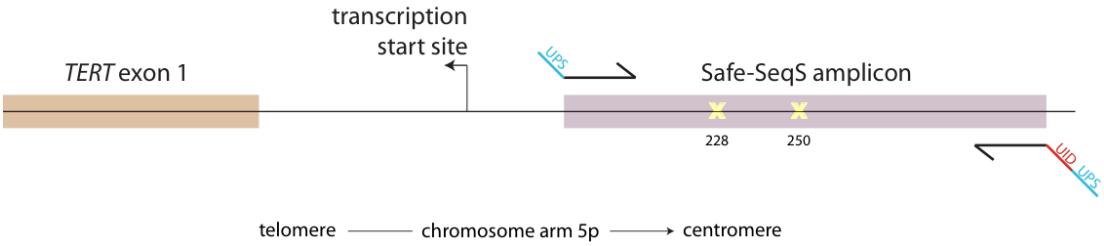
Patient	*TERT* promoter mutation	Fraction of mutant alleles	Tumor grade
1	g.1295228C>T	24%	pTa LG
2	g.1295250C>T	33%	pTa LG
3	g.1295228C>T	24%	pTa LG
4	g.1295228C>T	18%	pTa LG
6	g.1295228C>T	25%	pTa LG
7	g.1295228C>T	29%	pTa LG
8	g.1295228C>T	24%	pTa LG
10	Absent	NA	pTa LG
11	g.1295228C>T	24%	pTa LG
12	g.1295228C>T	21%	pTa LG
13	g.1295228C>A	31%	pTa LG
14	g.1295250C>T	40%	pTa LG
15	g.1295228C>T	23%	pTa LG
16	g.1295228C>T	22%	pTa LG
17	Absent	NA	pTa LG
18	g.1295228C>T	15%	pTa LG
19	Absent	NA	pTa LG
20	g.1295250C>T	31%	pTa LG
21	g.1295228C>T	20%	pTa HG
22	g.1295228C>T	27%	pTa HG
23	g.1295228C>T	30%	pTa HG
24	Absent	NA	pTa HG
25	g.1295228C>T	21%	pTa HG
26	g.1295228C>T	4.9%	pTa HG
27	Absent	NA	pTa HG
28	g.1295228C>T	36%	pTa HG
29	g.1295228C>T	30%	pTa HG
30	g.1295228C>T	32%	pTa HG
31	Absent	NA	pTa HG
32	g.1295228C>T	28%	pTa HG
33	g.1295250C>T	50%	pTa HG
34	g.1295242C>T	49%	pTa HG
35	Absent	NA	pTa HG
36	Absent	NA	pTa HG
37	g.1295228C>T	34%	pTa HG
38	Absent	NA	pTa HG
39	Absent	NA	pTa HG
40	g.1295228C>T	32%	pTa HG
41	g.1295228C>T	7.3%	pTa HG

Table 4-S1. Continued.

42	g.1295228C>T	11%	CIS
43	g.1295250C>T	1.9%	CIS
44	g.1295250C>T	4.2%	CIS
45	g.1295250C>T	2.1%	CIS
46	Absent	NA	CIS
47	g.1295228C>T	7.7%	CIS
48	g.1295250C>T	19%	CIS
49	g.1295228C>T	3.3%	CIS
50	Absent	NA	CIS
51	g.1295228C>T	4.3%	CIS
52	g.1295228C>T	8.4%	CIS
53	g.1295228C>T	6.7%	CIS
54	Absent	NA	CIS
55	Absent	NA	CIS
56	Absent	NA	CIS
57	g.1295250C>T	7.5%	CIS
58	g.1295228C>T	3.7%	CIS
59	Absent	NA	pTa LG
60	g.1295228C>T	6.7%	pTa HG
61	g.1295228C>T	7.0%	pTa LG
62	Absent	NA	pTa HG
63	g.1295228C>T	8.3%	pTa LG
64	g.1295228C>T	7.9%	pTa HG
65	g.1295250C>T	31%	pTa LG
66	g.1295250C>T	23%	pTa HG
67	g.1295228C>T	5.4%	pTa LG
68	Absent	NA	pTa HG
69	g.1295228C>T	7.9%	pTa LG
70	g.1295228C>T	7.1%	pTa HG
71	g.1295242C>T	12%	pTa LG
72	g.1295250C>T	24%	pTa HG
73	g.1295228C>T	20%	pTa LG
74	g.1295228C>T	1.5%	pTa HG
75	g.1295228C>T	8.5%	pTa LG
76	Absent	NA	pTa HG
77	g.1295228C>T	9.6%	pTa LG
78	g.1295228C>T	6.5%	pTa HG

Genomic coordinates refer to the minus (-) strand of chromosome 5 (hg19 assembly). CIS: Carcinoma *in situ*; HG: high-grade noninvasive urothelial carcinoma; LG: low-grade noninvasive urothelial carcinoma.

Figure 4-1. *TERT* promoter locus.



Chapter 5: Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies

Introduction

Cancer will occur in more than 1.6 million individuals this year in the United States alone, but a clinically proven circulating biomarker that can be used to help guide patient management will be available for only a minority of them, even in the setting of widespread metastasis [133-138]. While serum-based protein biomarkers such as carcinoma antigen-125 (CA-125), carcinoembryonic antigen (CEA) and prostate specific antigen (PSA) are commonly used for this purpose, these proteins are also found in the serum of individuals without cancer, albeit in lower concentrations [134-136]. Additionally, these markers are not found to be elevated in a substantial portion of patients with advanced cancers [137,138].

A new generation of biomarkers has become available with the discovery of the genetic alterations that are responsible for the initiation and progression of human cancers [9,139-142]. With the influx of genomic information from recent cancer genome sequencing studies, it is now known that virtually all cancers of every type harbor somatic genetic alterations. These alterations include single base substitutions, insertions, deletions and translocations (the latter including those associated with the creation of gene fusions, gene amplifications or losses of heterozygosity). These somatic mutations occur at negligible frequencies in normal cell populations and therefore provide exquisitely specific biomarkers from a biologic perspective [9].

There are two sources of tumor DNA that can be non-invasively assessed in the circulation: cell-free circulating tumor DNA (ctDNA) and circulating tumor cells (CTCs) [143,144]. Circulating tumor DNA is comprised of small fragments of nucleic acid that are not associated with cells or cell fragments [145]. In contrast, circulating tumor cells represent intact, often viable, cells that can be purified from blood by virtue of physicochemical characteristics or cell surface molecules that distinguish them from normal blood cells [146]. Many studies have shown that both ctDNA and CTCs are present in advanced neoplasia, though only a few studies have compared the amounts of CTCs and ctDNA templates in the same patients [147-150]. The studies comparing the two approaches have reached opposing conclusions, likely due to technical issues that limited interpretation of either the ctDNA or CTC content. Furthermore, the mechanism by which CTCs or ctDNA are released into the circulation are unclear, although it is possible that ctDNA actually comes from CTCs. One of the purposes of the current study was to compare the quantities of ctDNA and CTCs in the circulation of the same patients using an unbiased approach.

Most studies of ctDNA published to date have each evaluated patients with a single tumor type. In light of considerable differences in DNA preparation and analytic techniques in these studies, it has been difficult to directly compare the amounts of ctDNA among tumor types [147,151-157]. Comparisons of studies are also challenging due to differences in the types of data that are reported. For example, it is often impossible to compare real-time PCR results with those reporting the fraction of mutant template molecules assessed, or to compare results based on the analysis of serum with those based on plasma. To directly compare different tumor types and to determine the spectrum of cancers in which ctDNA

measurements could prove clinically useful, we evaluated a large number of tumor types in the current study. We purified plasma and tumor DNA using regimented protocols for all samples and used digital technologies to evaluate ctDNA levels from each tumor so that we could report the number of mutant templates per milliliter of plasma in each case (Fig. 1). This approach also allowed us to directly compare directly the two most commonly used types of tumor-specific genetic alterations found in the circulation – single base substitutions and rearrangements.

One of the most immediate applications of ctDNA has been termed the "liquid biopsy" [151]. In research studies as well as in clinical practice, it is often difficult to obtain tumor samples for genetic analyses. Some tumors are only accessible through fine needle aspirates (lung cancer, for example) with insufficient material available for genotyping, whereas in other cases it can be challenging or time-consuming to acquire samples from different medical centers [158]. Additionally, once a targeted therapy is initiated in a patient with multiple metastases, clinicians frequently search for early evidence of recurrence or mechanisms underlying resistance, a scenario in which liquid biopsies are particularly valuable. For example, they can provide temporal measurements of the total tumor burden as well as identify specific mutations that arise during therapy [147,151,154,159]. Though the liquid biopsy approach has been shown to be promising, its sensitivity and specificity with respect to conventional tumor biopsies has not been evaluated in a large, clinically relevant cohort. In the current study, we evaluated the sensitivity and specificity of this approach in patients with colorectal cancers who were candidates for EGFR blockade. We also used liquid biopsies to identify mutations that were responsible for recurrence in patients who initially responded to EGFR blockade. In aggregate, these studies provide a wealth of information

on the potential utility, as well as the limitations, of ctDNA measurements for the assessment of patients with various cancers.

Results

Patients with metastatic cancers. We began this study with an evaluation of 136 metastatic tumors originating from 14 different tissue types, as well as of 41 patients with primary brain tumors (glioma and medulloblastoma). Primary brain tumors were also included because they are generally lethal but rarely metastasize. We also included 10 additional cases, comprised of stage III ovarian (n=7) and hepatocellular carcinomas (n=3) in this particular evaluation because stage IV cases were rare and Stage III disease is more representative of advanced disease in these two tumor types. The clinical characteristics of these patients are summarized in Table 1. Targeted sequencing, exomic sequencing, or whole genome sequencing was used to identify mutations in the tumors, as described in the Materials and Methods. In these advanced cases, least one genetic alteration - a point mutation (151 cases) or genetic rearrangement (36 cases) - was found in each of the tumors studied (table S1). Except for a subset of mutations at the known hotspots of the *KRAS, NRAS, PIK3CA,* and *BRAF* genes (which are well-known to be somatic), all other genetic alterations were demonstrated to be somatic through evaluation of DNA from non-neoplastic cells of the same patients. Circulating tumor DNA was assessed by one of three digital methods (see Methods). These methods yielded comparable results when applied to the same plasma samples (fig. S1) and all were able to detect one mutant template in the DNA purified from up to 5 mL plasma. The amounts of plasma available from each patient are listed in table S1.

Circulating tumor DNA was detected in the majority of the studied patients with solid tumors outside the brain (112 of 136; 82%). However, the fraction of patients with detectable ctDNA varied with tumor type (Likelihood ratio test, p-value < 0.001). As shown in Fig. 2A and fig. S2, most patients with stage III ovarian and liver cancers and metastatic cancers of the pancreas, bladder, colon, stomach, breast, liver, esophagus, head and neck, as well as patients with neuroblastoma and melanoma, harbored detectable levels of ctDNA, although small sample sizes for some tumor types led to wide confidence intervals. In contrast, less than 50% of patients with medulloblastomas or metastatic cancers of the kidney, prostate, or thyroid, and less than 10% of patients with gliomas, harbored detectable ctDNA. The number of patients with some of the tumor types depicted in Fig. 2A was small, limiting the statistical significance of comparisons among tumor types, but patients with gliomas (low or high grade; table S1) were less likely to harbor ctDNA than patients with metastatic cancers of the pancreas, colon, breast, esophagus/stomach, or ovary (Fig. 2A and Fig. S2).

Though ctDNA was detectable in most patients with metastatic cancers, the concentration of ctDNA varied among patients, even those with the same tumor type (Fig. 2B and table S1). Some of this variability was due to differences in copy number of the genes assayed in different tumors. For example, if the queried gene was amplified 50-fold in the tumor of Patient A, whereas the queried gene in the tumor of Patient B was present at normal copy number, the amount of ctDNA would be expected to be 50-fold higher in Patient A than in Patient B (see section entitled "Comparison of rearrangements with single base substitutions in ctDNA" below). However, great variability was also observed among cancers in which only non-amplified genes (such as *TP53*) were assessed.

Patients with localized disease. We next evaluated ctDNA in patients with localized disease, that is, no clinical or radiographic evidence of distant metastasis at the time of sample collection. Among 223 patients with localized cancers of all types evaluated, detectable levels of ctDNA were found in 55% (122 of 223 patients; table S1). This fraction was lower than observed in patients with metastatic disease from all tumor types in which a sufficient number of samples were available (breast, colon, pancreas, gastroesophageal; Fig. 3A; Cochran-Mantel-Haenszel chi-squared test, $p < 0.001$). Detectable levels of ctDNA were present in 49 to 78% of patients with localized tumors and in 86% to 100% of patients with metastatic tumors of these four types (Fig. 3A).

Differences in the fraction of patients with detectable levels of ctDNA also correlated with stage: 47% of patients with Stage I cancers of any type had detectable ctDNA, whereas the fraction of patients with detectable ctDNA was 55%, 69% and 82% for patients with Stage II, III, and IV, respectively (Fig. 3B; Somers' Dxy rank correlation = 0.337). The concentration of ctDNA in the plasma similarly increased with stage (Fig. 3C).

Comparison of ctDNA with CTCs. For these experiments, DNA was isolated from the cellular compartment of blood obtained after centrifugation; these pellets contained circulating tumor cells as well as WBCs, platelets, and other cellular fragments. In each case, whole genome sequencing of tumor DNA was used to identify somatic rearrangements. PCR-based assays were then used to identify these rearrangements in blood pellets (CTCs) or in the blood supernatants (plasma) of the same patients. This experiment could be performed with tumor-specific rearrangements, but not with tumor-specific point mutations, for the reasons given in the Discussion. We did not identify any cases in which CTCs were

detected but in which ctDNA was absent. However, in many cases in which ctDNA was detected (13 of 16; 81%), no CTCs were detectable with the identical assay (Table 2). Moreover, in the 3 cases wherein both CTC and ctDNA levels were detectable, the average number of mutant fragments in the plasma was >50-fold higher than analogous levels in CTCs (Table 2).

Comparison of rearrangements with single base substitutions in ctDNA. We were also interested in comparing the quantity of two different types of genetically altered DNA fragments in the circulation of the same patients. Though practical issues precluded us from identifying a rearrangement in all patients in this study (see Discussion), tumor-specific rearrangements as well as tumor-specific point mutations were identified in 19 patients (table S2). The rearrangements were identified by whole genome sequencing of tumor DNA and the point mutations identified by targeted sequencing. In each case, the alteration was shown to be somatic via evaluation of normal DNA from the same patients. In 18 of the 19 patients harboring a circulating point mutation, a circulating rearrangement was also detectable (table S2). The one exception was a patient (CRC 37) with a circulating point mutation in *TP53* in which the rearrangement identified in that patient's tumor could not be identified in her plasma (table S2). The absolute number of circulating DNA fragments with point mutations vs. rearrangements was highly correlated (Fig. 4; correlation coefficient = 0.96). However, in four patients, the number of circulating fragments containing rearrangements was > 10-fold that of the queried point mutation (table S2). The reason for this was that the rearrangements we chose for analysis often arose as a result of gene amplification in the tumor, whereas the point mutations were generally present only once per tumor genome.

The sensitivity and specificity of liquid biopsy. The results described above were obtained by first identifying a mutation in a tumor and then determining whether that same mutation was detectable in the plasma. For certain liquid biopsy applications, the mutation in the tumor is not known *a priori* and all mutations of interest are queried at once. To determine the sensitivity of the liquid biopsy approach, we evaluated the plasma and tumors of 206 patients with metastatic colorectal cancer in a blinded fashion (table S3). This cohort of patients was completely distinct from the 410 patients described above and in tables S1 and S2. For each case, we determined whether mutations at codons 12 or 13 of *KRAS* were present in either the primary tumor or in 2 ml plasma drawn prior to treatment. The *KRAS* gene was chosen for this study because of its clinical relevance; the absence of a *KRAS* gene mutation in the primary tumor is a prerequisite for treatment of metastatic CRC patients with antibodies that block EGFR[160]. We identified 69 patients (33% of the 206) who harbored circulating mutant *KRAS* in their plasma. Circulating *KRAS* mutations were not detected in 127 of 128 patients with KRAS wild-type tumors, yielding an uncorrected specificity of 99.2%. Importantly, the mutation identified in the 69 plasma samples was always identical to that identified in the tumors, further emphasizing the specificity of the liquid biopsy. In addition to these 69 tumors, we identified ten cases (of 206) in which mutations were present in the primary tumors but not in the plasma, yielding a sensitivity of 87.2%. Percent concordance between *KRAS* mutation status in the plasma and tumor tissue was 95% and the agreement was highly significant (Kappa statistic 0.88, $p<0.0001$).

We next evaluated 26 clinical and pathologic characteristics to better understand the observed false negative results (tables S3 and S4). The factors associated with a false negative

ctDNA result (mutant KRAS in the tumor but no mutants detectable in the plasma) were a low CEA level, mucinous histology, low ALT, low white blood cell count and younger age (table S4 and S5). CEA levels were also positively correlated with the concentration of mutant *KRAS* fragments in the plasma (table S6 and S7). These observations are consistent with the idea that lower tumor burdens (reflected by normal CEA levels) are associated with lower ctDNA levels.

We next examined the relationship between the concentration of ctDNA and survival. Beginning with a model of known prognostic factors (age, Eastern Cooperative Oncology Group (ECOG) Performance Status (PS), and CEA), and assuming linearity for these adjustment variables, we found that ctDNA concentration provided added value in predicting survival (Likelihood ratio test, $p = 0.00253$, $df=3$). We then estimated the 2-year survival rate for differing concentration of ctDNA, holding the other predictors constant (Fig. 5).We observed a steady decrease in survival rate as ctDNA concentration increased.

Monitoring patients for resistance-conferring mutations. Liquid biopsies can also be used to monitor patients being treated with targeted agents, providing an early warning of recurrence and information about the genetic basis of resistance. For example, *KRAS* codon 12 and 13 mutations were shown to develop in 38% of 24 patients who first responded to EGFR blockade, then progressed [151]. In each case, the *KRAS* gene mutation was not present in the primary tumor but had presumably arisen in a small population of cells within a metastatic lesion and expanded under the influence of the EGFR blockade. In the current study, we wished to determine whether other resistance mutations, besides those at *KRAS* codons 12 and 13, could be identified in liquid biopsies of patients treated with EGFR

blockade. We therefore designed a multiplexed, sequencing-based assay to query known mutated hot-spots of several genes in the EGFR pathway: the regions within and surrounding *KRAS* codons 12, 13, 59, 60, and 61, *NRAS* codons 12, 13, 59, 60, and 61; *BRAF* codons 599 and 600, *EGFR* codons 712 - 721, 738 - 748, 790 - 800, and 847 - 859 and *PIK3CA* codons 538 – 549 and 1039 - 1050. The 24 cases assessed included 17 of those previously assessed for *KRAS* mutations [151] plus seven additional cases of patients who had first responded, then progressed, while being treated with blocking antibodies to EGFR (panitumumab or cetuximab). The primary tumors of 9 of these cases were unavailable, so we used pre-treatment DNA from plasma to assess whether any of the queried mutations were detected prior to administration of EGFR antibodies; none of the mutations listed in Fig. 6 were found prior to antibody treatment.

We identified emergent circulating mutations of at least one MAPK pathway gene in 23 of the 24 patients (96%). The number of different mutations identified in the circulation of individual patients averaged 2.9 (range 0 to 12). The development of different mutations in the same patient is not surprising given that each of these patients had multiple lesions; each lesion that responds to EGFR blockade and then progresses is expected to harbor at least one resistance mutation [151,161].

In total, we observed 70 somatic mutations that were not detected in the tumor or in the plasma prior to EGFR blockade and only appeared after therapy was initiated (table S8; Fig. 6). Half of the mutations (34 of 70) occurred in *KRAS* codon 12. These mutations are known to cause resistance to EGFR blockade when present in the primary tumor, and have been observed to arise after EGFR blockade *in vitro* as well as *in vivo* [151,161]. One mutation in *BRAF* was observed. Several previous studies have shown that BRAF V600E mutations,

when present in primary tumors, are associated with failure to achieve a response to EGFR blockade [162-164]. Two other patients developed mutations in the kinase domain of EGFR (codons 714 and 794; table S8; Fig. 6). Mutations at these residues have been previously observed in primary CRC, albeit infrequently, and resistance to EGFR blockade has been shown to result from genetic alterations in the *EGFR* gene [165,166]. We did not identify treatment-related mutations in the known PIK3CA gene hot spots (exon 9 and 20) [167].

The most surprising observation in the EGFR blockade component of our study was the large number of mutations in codon 61 of either the *KRAS* or *NRAS* gene (table S6; Fig. 6). Fifteen of the 24 patients (62.5%) harbored at least one codon 61 mutation, and the 31 mutations in these 15 patients comprised 45% of the total (69) mutations observed. Forty eight percent of the codon 61 mutations were in *NRAS* and the remainder were in *KRAS* (table S6; Fig. 6).

Discussion

Through the study of 640 patients, we have learned that mutant DNA fragments are found at relatively high concentrations in the circulation of most patients with metastatic cancer and at lower but detectable concentrations in a substantial fraction of patients with localized cancers. These results have several translational implications and suggest important avenues of future research.

Monitoring disease in advanced cancer patients. A genetic alteration could be identified in the tumor of all 410 patients evaluated in this part of study, making ctDNA a widely applicable biomarker for cancer patients. Moreover, >80% of patients with metastatic

disease had detectable levels of ctDNA, higher than that reported for most conventional biomarkers[168]. Unlike proteins such as CEA or CA19-9, which are expressed in normal cells as well as in neoplastic cells, genetic alterations of a clonal nature are only found in neoplasms. Our data indicate that measurements of ctDNA can also provide therapeutic, predictive and prognostic information in patients with metastatic disease. As shown in Fig. 5, metastatic colorectal cancer patients with relatively low levels of ctDNA lived significantly longer than patients with higher levels, and there was a striking correlation between ctDNA concentration and survival. A similar association between survival and ctDNA concentration has recently been reported in patients with advanced breast cancers [147].

Though these advantages of ctDNA render it promising for monitoring patients, there are potential limitations. The specific mutations are defined by evaluation of the primary tumor, adding both time and expense to patient management. This may be less of an obstacle in the future as more cancer patients have their tumors genetically analyzed to guide therapeutic decisions. The genetic alterations used to guide therapies can also be used for ctDNA analysis. A more serious issue relates to the utility of monitoring patients with advanced cancers, either with ctDNA or with other biomarkers [169,170]. On one hand, patients and their physicians are anxious to know, as soon as possible, whether disease has progressed. Imaging studies are often non-informative or slow to reflect progression. Repeated imaging also subjects patients to radiation, while monitoring ctDNA is non-invasive. On the other hand, it has not yet been shown that monitoring patients with advanced disease with *any* biomarker provides clinical as opposed to psychological benefits. Knowing that progression (or response) has occurred prior to changes in clinical symptoms may not prolong survival or improve quality of life.

Methodological comparisons. There are two sources of tumor DNA accessible in the blood (CTCs and ctDNA), and two types of genetic alterations that can be most easily assessed in either source (point mutations and translocations). Previous studies that compared ctDNA with CTCs reached mixed conclusions. For example, one group concluded that ctDNA was present less often than CTCs [148]; this group used state-of-the-art methods to detect CTCs, but did not use a highly sensitive method to detect ctDNA. The second group concluded that ctDNA was present more often than CTCs [147]; this group used a sensitive method for analyzing ctDNA but used a relatively insensitive method for analyzing CTCs. More recently, much higher levels of ctDNA than CTCs were found in 2 of 3 pediatric patients with neuroblastomas [150].

To investigate this issue further, we assessed both ctDNA and CTCs in the same blood sample from patients with typical solid tumors. We simply separated the cellular component from plasma and determined the fraction of cells or cell equivalents, respectively, in which tumor-specific rearrangements could be identified. Because we did not attempt to physically separate tumor cells from normal WBCs, technical issues related to the efficiency of CTC purification were eliminated. The comparison between DNA from CTCs and ctDNA cannot easily be performed with point mutations because the background levels of point mutations in PCR based assays is too high, even with the sensitive methods used in our study. This background precludes the detection of point mutations at levels less than 1 in 100,000 cells [11,171]. Because in cancer patients there exists a mixture of several million normal cells with very few CTC per ml of blood, a technology that is more sensitive is required. The detection of rearrangements is well suited for this task, as it has been shown that one

mutation can be reliably detected among millions of wild-type template molecules; PCR errors do not generate specific rearrangements [172].

Using patient-specific rearrangements as a tool, we were able to show that the level of ctDNA was always higher than the level of CTCs. In 13 of 16 patients, ctDNA levels were relatively high while no CTCs at all could be detected. This does not mean that ctDNA is preferable to CTCs for the detection or monitoring of cancer. Rather, the optimal technology depends on many other factors, including cost and throughput, for which CTC detection has advantages. But this comparison does suggest that the vast majority of ctDNA is not derived directly from CTCs. As the half-life of ctDNA is short (<1.5 hours) [159], in fact shorter than that of CTCs [173], our work suggests that the mutant molecules in the plasma are generally not derived from the circulating tumor cells.

Another comparison of interest concerns translocations and point mutations. Our results (table S2) show that the number of ctDNA fragments per mL of plasma for translocations and point mutations were similar in the majority of cases studied. However, in 1 of 19 cases, a point mutation was detected in a plasma sample in which the studied rearrangement was absent. The likely reason for this was that the point mutation was in a driver gene that occurred relatively early in tumorigenesis while the rearrangement was sub-clonal, perhaps not contributing to the development of the tumor. In 4 other cases, rearrangements were detected at ten-fold higher levels than the point mutations (table S2). In these cases, the rearrangements were found to be components of somatically amplified genes.

From a practical perspective, these data suggest the following conclusions: maximal sensitivity for detecting a genetic alteration can be achieved by using a rearrangement present within an amplicon. Many tumors, particularly advanced ones, contain such amplifications, making them relatively easy to detect with low coverage (10x) genome sequencing. As with the comparison between CTCs and ctDNA, however, this greater sensitivity does not mean that rearrangements are preferred over point mutations for clinical use. The discovery of a rearrangement in a patient's tumor, and the work and time required to develop and test primer pairs that can efficiently detect the rearrangement(s) in the degraded DNA characteristic of plasma, is considerable. In contrast, a panel of assays detecting the most commonly mutated point mutations is currently simpler and less expensive to implement in the clinical setting.

Early detection of localized cancers. Until therapeutic agents with much greater potency and minimal side effects are developed, the current best hope for reducing cancer morbidity and mortality is early detection of neoplastic disease [9]. Prior to metastasis, most solid tumors can be cured by extant surgical methods, and even when occult metastasis has occurred, adjuvant therapy or additional surgery can lead to cure in some patients. One of the encouraging results of our study is that ctDNA was found in the majority of patients with localized disease, when their chances of a favorable outcome are highest (Fig. 3). Even in patients with Stage I disease, who are nearly always curable by surgery alone, 47% of patients were shown to have detectable levels of ctDNA in their plasma. In Stage III disease, which is curable in many patients with certain forms of cancer, more than two-thirds of patients had detectable ctDNA.

Though early detection strategies based on ctDNA are promising, numerous obstacles must be overcome before they can be applied clinically. The fraction of patients with detectable ctDNA represents the maximum obtainable with the amount of plasma collected in this study (Table S1). In a screening setting, with the exception of pancreatic ductal adenocarcinomas (where one gene, *KRAS,* is mutated in almost all cases [174], the mutation of interest would not be known *a priori* and a panel of genes would have to be assessed. Our study on the EGFR blockade cohort shows that it is indeed possible to assess several genes at once for the detection of relatively rare mutations in plasma (table S6).

In addition to these technical challenges, biomedical issues will have to be addressed by any ctDNA-based screening test. False positive findings can be problematic for any screening assay [175]. Experience thus far suggests that benign tumors and non-neoplastic conditions do not generally give rise to ctDNA [176], so the "over-diagnosis" of benign tumors is not likely to pose a major problem. However, other studies suggest that a tumor containing ~50 million malignant (rather than benign) cells releases sufficient DNA for detection in the circulation [151]. A cancer of this size is far below that required for definitive imaging at present. How would a patient who had a positive ctDNA test be managed if follow-up imaging tests were negative? A related issue is the fact that the type of mutation does not provide many clues to the tumor type. For example, a patient with a circulating *TP53* mutation, in the absence of other mutations, could have a cancer in any of several organs. Another question concerns the value of detecting early cancers. In pancreatic ductal adenocarcinomas, for example, it might be argued that most patients with a positive ctDNA test will die from their disease anyway, given the aggressive nature of this form of cancer. Though these obstacles are formidable, we would argue that the presence of detectable amount of a mutant driver gene

is a cause for serious concern given the known causal relationships between such mutations and cancer. Indeed, this point distinguishes mutation-based biomarkers from all other types of biomarkers yet described.

Liquid biopsies. Our studies demonstrate two uses for liquid biopsies. The first - assessing plasma for the presence of specific mutations that can direct patient management - is clinically actionable. We show here that the sensitivity of the liquid biopsy for testing *KRAS* codon 12 is 88.2% in patients with metastatic CRC. Though conventional tumor biopsies are preferable, these often cannot be obtained for logistic or medical reasons. When tumor tissue specimens from metastatic cancer patients are unavailable, liquid biopsies offer an alternative that can be rapidly implemented without the pain, risk, and expense entailed by a biopsy of one of the metastatic lesions. Of note is the fact that ctDNA from neoplasms confined to the CNS (Fig. 2A) and those with mucinous features (table S4) was infrequently detectable. This suggests that physical obstacles such the blood-brain barrier and mucin could prevent ctDNA from entering the circulation.

Tracking Resistance. A second use of liquid biopsies is for identifying resistance mutations that occur when patients first respond to therapy, then progress. The detection of ctDNA requires tumor cells to die, and even tumor cells that are resistant to therapy turn over rapidly; they die almost as frequently as they are born[151]. Thus it is expected, and in fact observed, that the DNA fragments from drug-resistant cancer cells are found in the plasma. Though this approach is mainly of interest for research purposes at present, the obtained information can be clinically informative. A good example of this principle is provided by our discovery of remarkably frequent mutations at codon 61 of *NRAS* and of *KRAS*,

representing 46% of the detected mutations in patients resistant to EGFR blockade. Codon 61 mutations of *KRAS* and *NRAS* have previously been observed to occur in primary colorectal cancers, but very infrequently compared to the prevalence at which we found them in patients progressing after EGFR blockade [164]. *KRAS* codon 61 mutations have been observed to be associated with primary resistance to EGFR blockade when they occur in primary colorectal cancers [163,164,177]. There are no prior studies indicating that *NRAS* codon 61 mutations are associated with acquired resistance, but the results in Fig. 6 leave little doubt as to their role. This finding provides unequivocal evidence that these mutations confer resistance to therapy - the probability that recurrent mutations at these positions occurred by chance alone is essentially nil [151]. It also supports studies showing that *KRAS, BRAF, NRAS* and *EGFR* mutations compromise the efficacy of EGFR blockade in patients with colorectal cancer[177,178].

Collectively, codon 600 mutations of *BRAF*, codon 61 mutations of *KRAS*, and codons 12 or 61 mutations of *NRAS* occur approximately half as often as mutations in *KRAS* 12 or 13 in primary colorectal cancers [179]. These data therefore strongly suggest that patients being considered for treatment with EGFR blockading agents should be tested for these additional mutations. This conclusion was independently supported by a clinical study reported during the review of our manuscript [180]. Patients harboring mutations at these positions are unlikely to benefit from these agents and would be better served by other therapeutic approaches.

Summary. In summary, we demonstrate that ctDNA can be used as a feasible biomarker for a variety of different solid tumor types and clinical indications. The clinical utility of this biomarker, and the risks and benefits accruing from knowledge of ctDNA levels, can only be

addressed through longitudinal studies of ctDNA in appropriate populations of patients, as is currently underway for CTCs [181]. The studies reported here lay the groundwork for such future studies.

Supplementary Materials and Methods

Samples. All samples were collected after Institutional Review Board (IRB) approval at participating institutions, under full compliance with HIPAA guidelines. Tumors and adjacent normal tissues were either frozen at a minimum of -80 C or formalin-fixed and paraffin- embedded (FFPE) according to standard histopathologic procedures. Tumors were macro-dissected under a dissecting microscope to ensure a neoplastic cellularity of >60%. DNA was purified from the macrodissected frozen tumors using AllPrep (Qiagen, cat #80204) and from macrodissected paraffin-embedded tumors with a Qiagen FFPE Kit (Qiagen cat #56494). Translocations, but not point mutations, were previously reported for three of the CRCs [172,182]. Translocation data, but not all clinical correlatives, were previously reported for eight of the nine neuroblastomas recorded in table S1 [112]; these cases were included in the current study for comparative purposes only. For white blood cell DNA extraction, cells were pelleted at 1000 g prior to the preparation of plasma. DNA from these cells was purified using AllPrep (Qiagen, cat #80204). Plasma was used for ctDNA measurements in all experiments except in 17 of the 24 cases described in table S6, in which serum was used. DNA from plasma or serum was purified using QIAamp Circulating Nucleic Acid kit (Qiagen cat# 55114). Total plasma DNA concentration was measured using quantitative PCR as described [111]. The amounts of plasma available from all patients except those used for the liquid biopsy studies are listed in table S1; for the liquid biopsy study in table S3, 2 ml plasma was available and for the liquid biopsy study in table S6, 1 ml

serum or 2 ml plasma was used. One plasma draw was used for each patient except those described in table S6, in which two plasma draws were obtained: one prior to initiating EGFR blockade and one sample when the tumors had recurred after a clinical response.

Tumor Mutational Profiling. A tiered approach was used to identify somatic mutations in tumors. For pancreatic cancers, genomic regions encompassing *KRAS* codons 12,13, 59, 60 and 61 were amplified and the sequence of the PCR products determined via ligation of mutant-specific probes [112] or via SafeSeqS [11], as it is well known that nearly all pancreatic ductal adenocarcinomas harbor mutations in the *KRAS* gene [174], [183]. For colorectal cancers, PCR was used to amplify the *KRAS, BRAF, TP53, SMAD4, PIK3CA,* and *APC* genes and the sequence of the PCR products was determined, generally using SafeSeqS, as described below. For all other cancers, paired-end libraries were generated and regions encompassing 100 genes commonly mutated in cancers were captured as described previously [112]. For tumors that did not contain detectable mutations of these genes, exomic sequencing was performed after capture of the same libraries via SureSelect (Agilent), as previously described [184,185]. In cases in which rearrangements were analyzed using PARE (personalized analysis of rearranged ends), [172,182], genomic libraries were constructed for whole genome sequencing with a physical coverage of ~10x. Whenever possible, we selected rearrangements within amplified segments of the genome. Such rearrangements would be represented more often in tumor DNA than in DNA from normal cells, theoretically increasing the sensitivity of detection of the altered fragment in plasma. Once putative rearrangements were identified on the basis of sequencing data, PCR primers were designed to amplify PCR products of 100 bp that spanned the rearrangement. The rearrangements were confirmed to be somatic by

demonstrating that PCR products were generated from the DNA of the tumor but not from DNA of non-neoplastic cells of the same patient.

Mutation Detection in ctDNA or CTCs. In the early phases of this study, single base substitutions and small insertions or deletions (indels) were assessed either by BEAMing or [112,152] or by a PCR/ligation method [112,152]. For the latter method, 25% of the plasma DNA was aliquotted into wells of a 384-well plate so that an average of 1 ng was contained in each well. After PCR and ligation as described [112,152], all wells were individually evaluated via gel electrophoresis and fluorescence imaging. If all wells contained a mutation, the plasma DNA was re-diluted for more precise quantification. If no wells contained a mutation, then a further 65% of the plasma was aliquotted and the assay repeated; ~10% of the plasma was used to determine DNA concentration and to confirm that the plasma and tumor were derived from the same patient via single nucleotide polymorphism analysis. In the latter stages of this study, mutations were assessed by SafeSeqS, an approach in which template molecules are individually assessed via massively parallel sequencing [11]. For SafeSeqS, 25% of the plasma DNA was aliquotted into wells of a 96-well plate so that an average of 3 ng DNA was contained in each well. The DNA from each well was then amplified using well-specific index primers, and the DNA from all wells was pooled and subjected to massively parallel sequencing and analysis as described [11]. If no mutations were detected, a further 65% of the plasma was aliquotted and the assay repeated; ~10% of the plasma was used to determine DNA concentration and to confirm that the plasma and tumor were derived from the same patient via single nucleotide polymorphism analysis. In each experiment, equivalent amounts of DNA from non-neoplastic cells were included in adjacent wells performed to ensure that the identified mutations were not the result of errors generated during PCR or other steps of

the procedures [11]. SafeSeqS was used to assess all of the 206 metastatic colorectal cancer patients assayed for mutations in the liquid biopsy study (table S3) as well as to assess the 24 patients assayed for resistance mutations after EGFR blockade (table S6).

To assess differences in assay performance among the methods used to assess ctDNA, we quantified the amount of mutations in 20 plasma samples that had been evaluated by all three methods used for detecting point mutations (BEAMing [31], PCR-Ligation [112], or Safe-SeqS [11]). We found that the results were comparable, as evident from the data in fig. S1. All three methods could detect one mutant template in the DNA from 5 ml plasma, as determined by spiking known amounts of mutant KRAS DNA in plasma DNA from normal individuals.

Rearrangements in ctDNA or CTCs were detected and quantified by digital PCR, using PARE (Paired Analysis of Rearranged Ends) as described previously [11,172] with the following modifications. First, 25% of the plasma DNA was aliquotted into wells of a 384-well plate so that an average of 3 ng (plasma) or 300 ng (WBCs containing CTCs) were contained in each well. After amplification, a portion of each well was evaluated by polyacrylamide gel electrophoresis to determine whether a PCR product of the predicted size was present. If all wells contained a mutation, the DNA was re-diluted for more precise quantification. If no wells contained a mutation, then a further 65% of the plasma was aliquotted and the assay repeated. To further verify that the PCR fragments of the expected size contained the intended rearrangement, ligation reactions were performed on each PCR fragment as described [112,152]. The two oligonucleotides used in the ligation reaction spanned the breakpoint so that ligation only occurred if the PCR products assessed contained the

rearrangement (fig. S3). Control experiments with DNA from non-neoplastic cells of the same patients showed that each rearrangement reported in this study was not found in the germ line of that patient.

Statistical Analyses. Proportions of patients with detectable ctDNA, with 95% Wilson confidence intervals, the rank of the proportion and ctDNA concentration by cancer type, with 95% bootstrapped confidence intervals, are listed. Proportions of patients with detectable ctDNA were compared across cancer types with a likelihood ratio chi-square test from a logistic model of detectable ctDNA, across stage of disease with Somers' Dxy rank correlation, and across both cancer type and stage of disease for breast, colon, pancreas, and gastroesophageal cancers using a Cochran-Mantel-Haenszel chi-squared test.

For the liquid biopsy cohort of metastatic colorectal cancer patients, the sensitivity and specificity, along with 95% confidence intervals, for detecting a plasma KRAS mutation compared to the detection of a tissue KRAS mutation were calculated. We also report the percent concordance and kappa statistic for the agreement between liquid biopsy and tissue samples.

Clinical characteristics of the false negative and true negative groups were compared with Pearson's chi-square test for categorical variables and Wilcoxon Mann Whitney tests for continuous variables. Those variables which had <20% missingness of the dependent variable were included in a multivariable logistic regression model of true negative status using lasso penalties.

In cases with detectable levels of mutant KRAS fragments in the plasma, the association of clinical characteristics with the log ctDNA levels was evaluated using univariable linear regression models. Logarithm transformations were made for the dependent variable and some continuous predictor variables to correct for skewness. Those variables which had <20% missingness of the dependent variable were included in a multivariable linear regression using lasso penalties.

Overall survival was calculated from the time of ctDNA measurement to the date of death or last follow-up. The known prognostic factors (age, ECOG PS and CEA), linearity assumed, were included in a multivariable Cox proportional hazards model with ctDNA concentration level transformed with a natural spline function. The 2-year survival probability estimates were plotted against ctDNA concentration levels, fixing the other covariates at the mean (continuous variables) or mode (categorical variables). The other prognostic factors, MSI and BRAF status, had more than 20% missing values and were not adjusted in the multivariable model. The variables were selected based on their clinical relevance, and none were removed by statistical significance testing.

Statistical analyses were performed using the R statistical package (version 2.15.1).

Acknowledgments

This chapter first appeared in 10.1126/scitranslmed.3007094[186].

Table 5-1. Summary of clinical characteristics of 410 patients with various malignancies.

	Parameter value
Age, years	
Mean (SD)	63.0 (13.6)
Median (range)	64 (23–95)
No. unknown (%)	67 (16.3)
Gender, *n* (%)	
Female	163 (39.8)
Male	181 (44.1)
No. unknown (%)	66 (16.1)
Tumor type, *n* (%)	
Bladder	10 (2.4)
Breast	33 (8.0)
Colorectal	64 (15.6)
Endometrial	12 (2.9)
Gastroesophageal	21 (5.1)
Glioma	27 (6.6)
Head and neck	12 (2.9)
Hepatocellular	4 (1.0)
Medulloblastoma	14 (3.4)
Melanoma	20 (4.9)
Neuroblastoma	9 (2.2)
Non–small cell lung cancer	5 (1.2)
Ovary	9 (2.2)
Pancreas	155 (37.8)
Prostate	5 (1.2)
Renal cell carcinoma	5 (1.2)
Small cell lung cancer	1 (0.2)
Thyroid	4 (1.0)
Clinical stage*	
1	49 (13.3)
2	133 (36.0)
3	51 (13.8)
4	136 (36.9)

*Excludes 41 primary brain tumor patients.

Table 5-2. Comparison of CTCs with ctDNA.

Sample ID	Tumor type	Clinical stage	Cellular DNA (mutant fragments per 5 ml)	Plasma DNA (mutant fragments per 5 ml)
BLD 21	Bladder cancer	2	0	226
BLD 24	Bladder cancer	2	0	4
CRC 12	Colorectal cancer	4	0	79
CRC 14	Colorectal cancer	4	0	31
CRC 31	Colorectal cancer	1	0	35
CRC 32	Colorectal cancer	2	0	37
CRC 35	Colorectal cancer	2	0	5
CRC 40	Colorectal cancer	1	0	25
CRC 60	Colorectal cancer	4	680	73,000
CRC BIO 23a*	Colorectal cancer	4	370	21,000
CRC BIO 23b*	Colorectal cancer	4	400	28,000
BR 833	Breast cancer	2	0	2,500
BR 834	Breast cancer	2	0	41
BR 837	Breast cancer	2	0	3
BR 841	Breast cancer	2	0	690
BR 848	Breast cancer	2	0	9,900

*Two independent blood samples from the same patient, drawn 2 months apart, were separately analyzed.

Table 5-S1. Mutations in 410 patients with various malignancies.

Sample ID #	Tumor type (adenocarcinoma unless otherwise indicated)	Plasma volume (µl)	Mutation - nucleotide alteration	Mutation - amino acid alteration	Mutant fragments/5 mL plasma	Method used for mutation detection in plasma	Clinical Stage	Age	Sex	Evaluated as part of the metastatic cohort (fig. 2A and fig. 2B)
BLD 21	Bladder	3000	Intra-chromosomal; Inversion - chr12:68441745-chr12:67966237	Not applicable	226	PARE	2			
BLD 24	Bladder	2000	Inter-chromosomal; chr2:204396710-chr1:154286405	Not applicable	3.9	PARE	2			
BLD 29	Bladder	2000	Inter-chromosomal; chr10:127675281-chr20:17940042	Not applicable	0.0	PARE	2			
BLD 30	Bladder	3000	Inter-chromosomal; chr7:57625307-chr16:10203013	Not applicable	0.0	PARE	2			
BLD 41	Bladder	5000	TP53 c.841G>A	TP53 p.D281N	655	SafeSeqS	4	62	M	X
BLD 44	Bladder	1000	TP53 c.839G>C	TP53 p.R280T	2,450	SafeSeqS	2	77	M	
BLD 46	Bladder	2500	TP53 c.853G>A	TP53 p.E285K	2.0	SafeSeqS	4	82	M	X
BLD 47	Bladder	2000	TP53 c.991C>T	TP53 p.Q331X	308	SafeSeqS	4	77	M	X
BLD 48	Bladder	2000	CTNNB1 c.110C>T	CTNNB1 p.S37F	6.5	SafeSeqS	3	73	M	
BLD 50	Bladder	2000	TP53 c.184G>T	TP53 p.E62X	0.0	SafeSeqS	3	77	M	
CP2	Breast	4000	Intra-chromosomal; Duplication - chr7:86899034-chr7:92758338	Not applicable	10,900	PARE	4			X
CP3	Breast	4000	Intra-chromosomal; Duplication - chr18:52644480-chr18:53770564	Not applicable	2,780	PARE	4			X
CP4	Breast	4000	Intra-chromosomal; Inversion - chr11:33844453-chr11:20019027	Not applicable	970	PARE	4			X
CP5	Breast	4000	Intra-chromosomal; Deletion - chr9:26742593-chr9:14289825	Not applicable	95	PARE	4			X
CP6	Breast	4000	Intra-chromosomal; Deletion - chr8:89365486-chr8:89237886	Not applicable	233	PARE	4			X
CP7	Breast	4000	Intra-chromosomal; Duplication - chr21:43764815-chr21:46124984	Not applicable	600	PARE	4			X
CP8	Breast	4000	Intra-chromosomal; Inversion - chr17:38444351-chr17:38395199	Not applicable	1.3	PARE	4			X
CP9	Breast	4000	Inter-chromosomal; chr10:64401594-chr3:28363857	Not applicable	115	PARE	4			X
BR 801	Breast	2000	NOTCH1 c.7171C>T	NOTCH1 p.Q2391X	0.0	SafeSeqS	3	62	F	
BR 802	Breast	2000	TP53 c.713G>A	TP53 p.C238Y	33	SafeSeqS	3	55	F	
BR 803	Breast	2000	TP53 c.214_215insC	TP53 p.P72fs	21,900	SafeSeqS	3	54	F	
BR 804	Breast	2000	TP53 c.733G>A	TP53 p.G245S	128	SafeSeqS	3	62	F	
BR 805	Breast	2000	NOTCH1 c.4873G>T	NOTCH1 p.E1625X	28	SafeSeqS	3	81	F	
BR 806	Breast	2000	TP53 c.637C>T	TP53 p.R213X	110	SafeSeqS	3	81	F	
BR 807	Breast	2000	TP53 c.329G>C	TP53 p.R110P	0.0	SafeSeqS	2	55	F	
BR 831	Breast	2000	Intra-chromosomal; Inversion - chr5:170838182-chr5:150490356	Not applicable	0.0	PARE	2	77	F	
BR 832	Breast	2000	Intra-chromosomal; Deletion - chr18:32113305-chr18:32008616	Not applicable	0.0	PARE	2	47	F	
BR 833	Breast	2000	Intra-chromosomal; Duplication - chr2:30643115-chr2:30607641	Not applicable	2,480	PARE	2	77	F	
BR 834	Breast	2000	Inter-chromosomal; chr8:111657741-chr5:64600981	Not applicable	41	PARE	2	57	F	
BR 837	Breast	2000	Intra-chromosomal; Inversion - chr1:85446612-chr1:84923571	Not applicable	3.0	PARE	2	43	F	
BR 838	Breast	2000	Intra-chromosomal; Inversion - chr17:35367968-chr17:35389930	Not applicable	0.0	PARE	2	52	F	
BR 839	Breast	2000	Inter-chromosomal; chr5:64292782-chr12:12639702	Not applicable	0.0	PARE	2	82	F	
BR 840	Breast	2000	AKT1 c.49G>A	AKT1 p.E17K	0.0	SafeSeqS	2	76	F	
BR 841	Breast	2000	Inter-chromosomal; chr8:10996593-chr6:106894670	Not applicable	688	PARE	2	42	F	
BR 842	Breast	2000	Intra-chromosomal; Inversion - chr17:35339193-chr17:34860391	Not applicable	0.0	PARE	2	52	F	
BR 843	Breast	2000	TP53 c.659A>G	TP53 p.Y220C	0.0	SafeSeqS	1	67	F	
BR 848	Breast	5000	Intra-chromosomal; Deletion - chr19:33,559,434-cgr19:35,505,204	Not applicable	9,900	PARE	2	28	F	
BREAST10-1	Breast	3000	TP53 c.332T>A	TP53 p.L111Q	1,170	PCR-Ligation	4	56	F	X
BREAST2-1	Breast	3000	FBXL4 c.1187T>A	FBXL4 p.I396N	0.0	SafeSeqS	4	82	F	X
BREAST3-1	Breast	3000	TP53 c.536A>G	TP53 p.H179R	7,500	PCR-Ligation	4	76	F	X
BREAST4-1	Breast	3000	PIK3CA c.3140A>G	PIK3CA p.H1047R	0.0	PCR-Ligation	4	71	F	X
BREAST5-1	Breast	3000	EIF4B c.1756-1G>C	Not applicable	122	SafeSeqS	4	80	F	X
CP10	Breast	4000	TP53 c.637C>T	TP53 p.R213X	2,660	SafeSeqS	4			X
CRC 02	Colorectal	5000	Inter-chromosomal; chr12:73097777-chr1:63099083	Not applicable	22,000	PARE	1	65	M	
CRC 03	Colorectal	5000	Intra-chromosomal; chr11:98914172-chr11:98993460	Not applicable	6,410	PARE	1	66	M	
CRC 06	Colorectal	5000	Intra-chromosomal; chr13:27679603-chr13:25685009	Not applicable	85	PARE	1	71	M	
CRC 07	Colorectal	5000	Intra-chromosomal; chr4:85109657-chr4:85408635	Not applicable	42	PARE	1	57	F	
CRC 11	Colorectal	1000	Intra-chromosomal; Deletion - chr21:35145207-chr21:36324769	Not applicable	103,000	PARE	4	56	F	X
CRC 12	Colorectal	1000	Intra-chromosomal; Deletion - chr6:58080933-chr6:67229823	Not applicable	79	PARE	4	48	F	X
CRC 13	Colorectal	1000	TP53 c.743G>A	TP53 p.R248Q	13	SafeSeqS	4	87	F	X
CRC 14	Colorectal	5000	Intra-chromosomal; Duplication - chr8:141375810-chr8:141405769	Not applicable	31	PARE	4	35	F	X
CRC 21	Colorectal	5000	Intra-chromosomal; chr1:237741691-chr1:244145093	Not applicable	685	PARE	2	60	M	
CRC 27	Colorectal	5000	Intra-chromosomal; chr13:102839339-chr13:105470344	Not applicable	55	PARE	2	49	F	
CRC 30	Colorectal	5000	Intra-chromosomal; chr20:2658058-chr20:2770031	Not applicable	1,470,000	PARE	3	73	F	
CRC 31	Colorectal	2000	Inter-chromosomal; chr1:174709633-chr2:177720185	Not applicable	35	PARE	1	87	F	
CRC 32	Colorectal	2000	Intra-chromosomal; Deletion - chr18:2502911-chr18:5004367	Not applicable	37	PARE	2	82	M	
CRC 33	Colorectal	2000	Intra-chromosomal; Deletion - chr2:149076017-chr2:149290239	Not applicable	0.0	PARE	2	73	F	
CRC 34	Colorectal	2000	Intra-chromosomal; Inversion - chr21:40439891-chr21:40525788	Not applicable	0.0	PARE	2	77	F	
CRC 35	Colorectal	2000	Intra-chromosomal; Deletion - chr1:21881451-chr1:21892730	Not applicable	5.0	PARE	2	72	M	
CRC 36	Colorectal	2000	Intra-chromosomal; Duplication - chr20:52263938-chr20:52260592	Not applicable	0.0	PARE	2	72	F	
CRC 37	Colorectal	2000	Intra-chromosomal; Duplication - chr3:170782510-chr3:170870975	Not applicable	0.0	PARE	1	72	F	
CRC 38	Colorectal	2000	Intra-chromosomal; Deletion - chr3:60080713-chr3:60031900	Not applicable	0.0	PARE	1	66	F	
CRC 39	Colorectal	2000	Intra-chromosomal; Inversion - chr17:61955923-chr17:29976989	Not applicable	0.0	PARE	2	66	M	
CRC 40	Colorectal	2000	Intra-chromosomal; Duplication - chr10:95185791-chr10:99077846	Not applicable	25	PARE	1	77	M	
CRC 41	Colorectal	2000	APC c.3871C>T	APC p.Q1291X	2.8	SafeSeqS	1	52	M	
CRC 42	Colorectal	2000	BRAF c.1799T>A	BRAF p.V600E	8.8	SafeSeqS	1	72	F	
CRC 51	Colorectal	4000	Inter-chromosomal; chr8:30060122-chr1:190065347	Not applicable	3,850	PARE	4	70	M	X
CRC 53	Colorectal	5000	Intra-chromosomal; Deletion - chr16:52977177-chr16:52965488	Not applicable	7,150	PARE	4	50	M	X
CRC 54	Colorectal	5000	Intra-chromosomal; Deletion - chr11:7250750-chr11:1945726	Not applicable	1,240	PARE	4	56	F	X
CRC 55	Colorectal	5000	Intra-chromosomal; Inversion - chr8:38281631-chr8:39225849	Not applicable	295	PARE	4	53	M	X
CRC 58	Colorectal	5000	Intra-chromosomal; Deletion - chr20:14839816-chr20:14863644	Not applicable	1,370	PARE	4	50	F	X
CRC 59	Colorectal	5000	KRAS c.35G>A	KRAS p.G12D	277	SafeSeqS	4	48	F	X
CRC 60	Colorectal	5000	Intra-chromosomal; Inversion - chr17:35277642-chr17:35163714	Not applicable	73,300	PARE	4	53	M	X
CRC 61	Colorectal	5000	TP53 c.818G>A	TP53 p.R273H	427	SafeSeqS	4	35	F	X
CRC 62	Colorectal	5000	TP53 c.455C>T	TP53 p.P152L	361	SafeSeqS	4	63	M	X
CRC 63	Colorectal	5000	TP53 c.844C>T	TP53 p.R282W	1,490	SafeSeqS	4	69	M	X
CRC 64	Colorectal	4000	KRAS c.38G>A	KRAS p.G13D	113	SafeSeqS	4	37	F	X
CRC 65	Colorectal	5000	KRAS c.38G>A	KRAS p.G13D	973	SafeSeqS	4	58	M	X
CRC 66	Colorectal	5000	KRAS c.35G>C	KRAS p.G12A	1,220	SafeSeqS	4	64	F	X
CRC 67	Colorectal	5000	TP53 c.396G>C	TP53 p.K132N	1.9	SafeSeqS	4	43	F	X
CRC Bio 162	Colorectal	1000	APC c.3927_3931delAAAGA	APC p.E1309_I1311fs	909	BEAMing	4	48	M	X
CRC Bio 168	Colorectal	1000	KRAS c.35G>A	KRAS p.G12D	374	BEAMing	4	38	M	X
CRC Bio 180	Colorectal	1000	BRAF c.1799T>A	BRAF p.V600E	22,900	SafeSeqS	4	60	M	X
CRC Bio 203	Colorectal	1000	KRAS c.35G>A	KRAS p.G12D	485	SafeSeqS	4	86	F	X
CRC Bio 204	Colorectal	3000	APC c.4348C>T	APC p.R1450X	1,440	SafeSeqS	4	36	M	X
CRC Bio 23	Colorectal	5000	Intra-chromosomal; Deletion - chr16:6343641-chr16:6727736	Not applicable	28,500	PARE	4	57	F	X
CRC Bio 92	Colorectal	1000	APC c.4216C>T	APC p.Q1406X	377	SafeSeqS	3	54	F	
OLS 13k	Colorectal	3000	TP53 c.406C>T	TP53 p.Q136X	7.5	PCR-Ligation	3			
OLS 14k	Colorectal	3000	TP53 c.613T>A	TP53 p.Y205N	48	PCR-Ligation	3			
OLS 20k	Colorectal	3000	TP53 c.613T>A	TP53 p.Y205N	48	PCR-Ligation	3			
OLS 21k	Colorectal	3000	TP53 c.833C>T	TP53 p.P278L	24	PCR-Ligation	2			
OLS 30k	Colorectal	3000	TP53 c.817C>T	TP53 p.R273C	3.7	PCR-Ligation	1			
OLS 33k	Colorectal	3000	KRAS c.35G>T	KRAS p.G12V	144	PCR-Ligation	3			
OLS 39k	Colorectal	3000	PIK3CA c.1624G>A	PIK3CA p.E542K	34	PCR-Ligation	3			
OLS 47k	Colorectal	3000	BRAF c.1799T>A	BRAF p.V600E	26	PCR-Ligation	2			
OLS 4k	Colorectal	3000	KRAS c.35G>A	KRAS p.G12D	6.5	PCR-Ligation	3			
OLS 52k	Colorectal	3000	TP53 c.404_406dupGCC	Not applicable	6.5	PCR-Ligation	3			
OLS 57k	Colorectal	3000	PIK3CA c.1636C>A	PIK3CA p.Q546K	11	PCR-Ligation	1			
OLS 58k	Colorectal	3000	PIK3CA c.1035T>A	PIK3CA p.N345K	50	PCR-Ligation	2			
OLS 60k	Colorectal	3000	KRAS c.34G>T	KRAS p.G12C	0.0	PCR-Ligation	1			
OLS 61k	Colorectal	3000	APC c.4678G>T	APC p.E1560X	0.0	PCR-Ligation	1			
OLS 62k	Colorectal	3000	KRAS c.35G>T	KRAS p.G12V	11	PCR-Ligation	3			
OLS 66K	Colorectal	5000	APC c.4364delA	APC p.N1455fs	14	PCR-Ligation	3			
OLS 67K	Colorectal	5000	BRAF c.1799T>A	BRAF p.V600E	23	PCR-Ligation	3			
OLS 69k	Colorectal	3000	PIK3CA c.1633G>A	PIK3CA p.E545K	0.0	PCR-Ligation	1			
OLS 72K	Colorectal	4000	KRAS c.34G>A	KRAS p.G12S	4.2	PCR-Ligation	2			
OLS 8k	Colorectal	3000	KRAS c.35G>T	KRAS p.G12V	17	PCR-Ligation	3			
PAP 024	Endometrial	4000	CTNNB1 c.101G>T	CTNNB1 p.G34V	0.0	SafeSeqS	3	47	F	
PAP 025	Endometrial	5000	TP53 c.291_292delCC	TP53 p.V97_P98fs	240	SafeSeqS	1	75	F	
PAP 026	Endometrial	5000	NRAS c.35G>A	NRAS p.G12D	1.1	SafeSeqS	1	87	F	
PAP 030	Endometrial	5000	MSH6 c.2153G>A	MSH6 p.S718N	0.0	SafeSeqS	1	74	F	
PAP 031	Endometrial	5000	CTNNB1 c.110C>A	CTNNB1 p.S37Y	0.0	SafeSeqS	1	61	F	
PAP 032	Endometrial	5000	PTEN c.388C>G	PTEN p.R130G	0.0	SafeSeqS	1	83	F	
PAP 033	Endometrial	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	1	68	F	
PAP 34	Endometrial	5000	KRAS c.100G>A	CTNNB1 p.G34R	1.1	SafeSeqS	1	55	F	
PAP 35	Endometrial	5000	KRAS c.35G>A	KRAS p.G12D	30	SafeSeqS	4	49	F	X
PAP 71	Endometrial	5000	PTEN c.388C>G	PTEN p.R130G	0.0	SafeSeqS	1	56	F	
PAP 80	Endometrial	5000	TP53 c.566delC	TP53 p.A189fs	0.0	SafeSeqS	3	94	F	
PAP 83	Endometrial	5000	PIK3CA c.3140A>G	PIK3CA p.H1047R	0.0	SafeSeqS	1	51	F	
CB ESO 03	Gastroesophageal	2000	TP53 c.818G>A	TP53 p.R273H	65	SafeSeqS	3	52	F	
CB ESO 04	Gastroesophageal	2000	PIK3CA c.263G>A	PIK3CA p.R88Q	7.5	SafeSeqS	3	67	M	
CB ESO 05	Gastroesophageal	2000	TP53 c.437G>A	TP53 p.W146X	0.0	SafeSeqS	3	72	M	
CB ESO 07	Gastroesophageal	2000	APC c.7709C>A	APC p.S2570X	0.0	SafeSeqS	3	62	M	
CB ESO 08	Gastroesophageal	2000	TP53 c.584T>C	TP53 p.I195T	0.0	SafeSeqS	3	87	M	
CB ESO 09	Gastroesophageal	2500	TP53 c.641A>G	TP53 p.H214R	2.3	SafeSeqS	3	67	M	
CB ESO 10	Gastroesophageal	2000	TP53 c.375+2A>G	Not applicable	0.0	SafeSeqS	3	72	F	
CB ESO 11	Gastroesophageal	2000	TP53 c.97-1C>T	Not applicable	48	SafeSeqS	3	82	M	
CBGP1-1	Gastroesophageal	3000	CDH1 c.1569T>A	CDH1 p.Y523X	185	PCR-Ligation	4	71	F	X
CBGP3-1	Gastroesophageal	3000	TP53 c.817C>T	TP53 p.R273C	67	PCR-Ligation	4	57	M	X
CBGP5-1	Gastroesophageal	3000	CDH1 c.563T>A	CDH1 p.V188D	220	PCR-Ligation	4	67	F	X
CBGP9-1	Gastroesophageal	3000	TP53 c.559+1C>T	Not applicable	50	PCR-Ligation	4	95	M	X
ESOPL1-1	Gastroesophageal	5000	TP53 c.818G>A	TP53 p.R273H	31	PCR-Ligation	4	57	M	X
ESOPL2-1	Gastroesophageal	5000	TP53 c.1010G>A	TP53 p.R337H	215	PCR-Ligation	4	66	M	X
G801	Gastroesophageal	2000	XIRP2 c.8144T>C	XIRP2 p.L2715P	1,690	SafeSeqS	3	44	F	
G803	Gastroesophageal	2000	CDKN2A c.334delC	CDKN2A p.R112fs	1,850	SafeSeqS	2	59	M	
G804	Gastroesophageal	2000	RELN c.2005T>C	RELN p.Y669H	239	SafeSeqS	4	39	F	X
G805	Gastroesophageal	2000	SPTB c.2939C>A	SPTB p.T980K	0.0	SafeSeqS	2	79	M	
G806	Gastroesophageal	2000	TP53 c.743G>A	TP53 p.R248Q	99	SafeSeqS	3	61	M	

Table 5-S1. Continued.

G807	Gastroesophageal	2000	PIK3CA c.263G>A	PIK3CA p.R88Q	157	SafeSeqS	2	76	M	
G809	Gastroesophageal	2000	PIK3CA c.278G>A	PIK3CA p.R93Q	0.0	SafeSeqS	3	61	M	
CB GLIOMA 22	Glioma	2000	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II oligodendroglioma	23	M	X
CB GLIOMA 29	Glioma	5000	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II oligodendroglioma	35	M	X
CB GLIOMA 30	Glioma	5000	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II astrocytoma	42	M	X
CB GLIOMA 31	Glioma	5000	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II astrocytoma	24	M	X
CB GLIOMA10	Glioma	2000	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II oligodendroglioma	43	F	X
CB GLIOMA4	Glioma	2000	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II oligodendroglioma	31	M	X
CB GLIOMA5	Glioma	3000	TP53 c.856G>A	TP53 p.E286K	0.0	SafeSeqS	Glioblastoma	44	M	X
CBBRP1	Glioma	4800	Intra-chromosomal; Deletion - chr11:29390774-chr11:13324150	Not applicable	0.0	PARE	Glioblastoma	74	M	X
CBBRP11	Glioma	5000	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Glioblastoma	58	M	X
CBBRP12	Glioma	5000	TP53 c.481G>A	TP53 p.A161T	0.0	PCR-Ligation	Grade II astrocytoma	53	F	X
CBBRP2	Glioma	2400	TP53 c.419C>T	TP53 p.T140I	0.0	PCR-Ligation	Glioblastoma	64	M	X
CBBRP23	Glioma	4500	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II oligodendroglioma	35	F	X
CBBRP24	Glioma	4800	IDH1 c.394C>A	IDH1 p.R132S	0.0	PCR-Ligation	Low grade astrocytoma	32	F	X
CBBRP25	Glioma	4800	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Grade II oligodendroglioma	24	F	X
CBBRP27	Glioma	4800	IDH1 c.394C>A	IDH1 p.R132S	0.0	PCR-Ligation	Grade II astrocytoma	34	F	X
CBBRP28	Glioma	3840	IDH1 c.394C>A	IDH1 p.R132S	0.0	PCR-Ligation	Grade II astrocytoma	36	M	X
CBBRP3	Glioma	5000	Intra-chromosomal; Duplication - chr12:67053391-chr12:56218273	Not applicable	0.0	PARE	Glioblastoma	56	F	X
CBBRP32	Glioma	1920	IDH1 c.395G>A	IDH1 p.R132H	0.0	PCR-Ligation	Gliosarcoma			X
CBBRP9	Glioma	5000	TP53 c.452C>A	TP53 p.P151H	0.0	PCR-Ligation	Grade III astrocytoma	55	M	X
GLI 101	Glioma	5000	TP53 c.731G>A	TP53 p.G244D	0.0	SafeSeqS	Glioblastoma	59	M	X
GLIOMA 102	Glioma	5000	EGFR c.2156G>A	EGFR p.G719D	0.0	SafeSeqS	Glioblastoma	62	M	X
GLIOMA 105	Glioma	5000	PIK3CA c.263G>A	PIK3CA p.R88Q	0.0	SafeSeqS	Glioblastoma	46	M	X
GLIOMA 106	Glioma	5000	TP53 c.569_570CT>TC	TP53 p.P190L	0.0	SafeSeqS	Glioblastoma	62	M	X
GLIOMA 109	Glioma	5000	PTEN c.799A>T	PTEN p.K267X	0.0	SafeSeqS	Glioblastoma	78	F	X
GLIOMA 110	Glioma	3500	TP53 c.569C>T	TP53 p.P190L	0.0	SafeSeqS	Glioblastoma	84	F	X
YN 406	Glioma	5000	Intra-chromosomal; Deletion - chr7:40716157-chr7:40524100	Not applicable	5.7	PARE	Glioblastoma	71	M	X
YN 407	Glioma	5000	Intra-chromosomal; Duplication - chr7:54793913-chr7:56081250	Not applicable	5.0	PARE	Glioblastoma	64	M	X
CB HN 10	Head and Neck Squamous Cell	5000	STK11 c.393C>A	STK11 p.Y131X	5.5	SafeSeqS	2	54	M	
CB HN 9	Head and Neck Squamous Cell	5000	PIK3CA c.1624G>C	PIK3CA p.E542Q	1.6	SafeSeqS	4	79	F	X
CBHNP1-1	Head and Neck Squamous Cell	5000	PIK3CA c.112C>T	PIK3CA p.R38C	44	PCR-Ligation	4	38	M	X
CBHNP2-1	Head and Neck Squamous Cell	5000	TP53 c.1010G>T	TP53 p.R337L	1,500	PCR-Ligation	4	59	F	X
CBHNP3-1	Head and Neck Squamous Cell	3200	BRAF c.1801A>G	BRAF p.K601E	0.0	PCR-Ligation	4	78	M	X
CBHNP4-1	Head and Neck Squamous Cell	4000	TP53 c.536A>G	TP53 p.H179R	50	PCR-Ligation	4	58	M	X
CBHNP6-1	Head and Neck Squamous Cell	5000	TP53 c.579_580TC>AT	TP53 p.H193_L194QF	500	PCR-Ligation	4	52	F	X
CBHNP7-1	Head and Neck Squamous Cell	5000	AKAP9 c.3395G>T	AKAP9 p.R1132L	1,280	SafeSeqS	4	68	M	X
HN 14	Head and Neck Squamous Cell	1500	TP53 c.817C>T	TP53 p.R273C	19	PCR-Ligation	4	44	F	X
HN 305	Head and Neck Squamous Cell	1000	Intra-chromosomal; Deletion - chr11:69177060-chr11:69178251	Not applicable	9.0	PARE	3	50	M	
HN16	Head and Neck Squamous Cell	1500	TP53 c.378C>G	TP53 p.Y126X	0.0	PCR-Ligation	4	68	M	X
HN41	Head and Neck Squamous Cell	1500	PIK3CA c.3140A>T	PIK3CA p.H1047L	0.0	PCR-Ligation	4	42	M	X
HCC 103	Hepatocellular	5000	TP53 c.536A>G	TP53 p.H179R	7.2	SafeSeqS	4	55	F	X
HCC 105	Hepatocellular	2000	PALB2 c.1620C>G	PALB2 p.N540K	15	SafeSeqS	3	87	F	X
HCC 106	Hepatocellular	2000	EGFR c.2014C>G	EGFR p.H672D	7,910	SafeSeqS	3	67	F	X
HCC 107	Hepatocellular	2000	SLC17A9 c.799G>C	SLC17A9 p.E267Q	0.0	SafeSeqS	3	62	M	X
CBMB10-1	Medulloblastoma	1000	CTNNB1 c.98C>G	CTNNB1 p.S33C	38	PCR-Ligation	Medulloblastoma			X
CBMB1-1	Medulloblastoma	2000	CTNNB1 c.98C>A	CTNNB1 p.S33Y	2.5	PCR-Ligation	Medulloblastoma			X
CBMB11-1	Medulloblastoma	1000	PTCH1 c.2778G>C	PTCH1 p.W926C	0.0	PCR-Ligation	Medulloblastoma			X
CBMB12-1	Medulloblastoma	1000	PTCH1 c.707G>A	PTCH1 p.W236X	0.0	PCR-Ligation	Medulloblastoma			X
CBMB13-1	Medulloblastoma	2000	CTNNB1 c.94G>T	CTNNB1 p.D32Y	0.0	PCR-Ligation	Medulloblastoma			X
CBMB2-1	Medulloblastoma	1000	PTCH1 c.3155insCGCGGC	PTCH1 p.T1052fs	5.5	PCR-Ligation	Medulloblastoma			X
CBMB24-1	Medulloblastoma	5000	KDM6A c.4153C>T	KDM6A p.Q1385X	39	SafeSeqS	Medulloblastoma			X
CBMB3-1	Medulloblastoma	1000	MLL2 c.1652C>T	MLL2 p.P551L	10.0	PCR-Ligation	Medulloblastoma			X
CBMB4-1	Medulloblastoma	2000	CTNNB1 c.98C>T	CTNNB1 p.S33F	0.0	PCR-Ligation	Medulloblastoma			X
CBMB5-1	Medulloblastoma	1000	PTCH1 c.981T>A	PTCH1 p.C327X	0.0	PCR-Ligation	Medulloblastoma			X
CBMB6-1	Medulloblastoma	1000	PTEN c.633delCinsGCG	PTEN p.C211fs	0.0	PCR-Ligation	Medulloblastoma			X
CBMB7-1	Medulloblastoma	1000	CTNNB1 c.110C>G	CTNNB1 p.S37C	7.5	PCR-Ligation	Medulloblastoma			X
CBMB8-1	Medulloblastoma	1000	TP53 c.376-2T>C	Not applicable	0.0	PCR-Ligation	Medulloblastoma			X
CBMB9-1	Medulloblastoma	1000	PTCH1 c.3119_3120insT	PTCH1 p.F1040fs	0.0	PCR-Ligation	Medulloblastoma			X
CB MEL 09	Melanoma	2000	NRAS c.182A>G	NRAS p.Q61R	0.0	SafeSeqS	4	57	F	X
CB MEL 10	Melanoma	2000	BRAF c.1799T>A	BRAF p.V600E	165	SafeSeqS	4	52	F	X
CB MEL 11	Melanoma	5000	ALK c.4732C>T	ALK p.P1578S	1.1	SafeSeqS	4	40	F	X
CB MEL2	Melanoma	5000	BRAF c.1799T>A	BRAF p.V600E	1.2	PCR-Ligation	4	32	F	X
MEL 03	Melanoma	5000	KRAS c.34G>T	KRAS p.G12C	0.0	SafeSeqS	3	71	M	
MEL 21	Melanoma	4000	Intra-chromosomal; Deletion - chr4:70925073-chr4:62475744	Not applicable	364	PARE	4	46	F	X
MEL 22	Melanoma	4000	TP53 c.639A>G	TP53 p.R213R	0.0	SafeSeqS	2	51	M	
MEL 23	Melanoma	4000	BRAF c.1799T>A	BRAF p.V600E	1,750	SafeSeqS	4	52	F	X
MEL 24	Melanoma	4000	BRAF c.1799T>A	BRAF p.V600E	6,140	SafeSeqS	4	61	M	X
MEL 25	Melanoma	4000	Inter-chromosomal; chr1:182048490-chr12:25928534	Not applicable	60	PARE	4	39	M	X
MEL 26	Melanoma	4000	Inter-chromosomal; chr9:12557627-chr21:41610646	Not applicable	1,090	PARE	4	41	M	X
MEL 27	Melanoma	4000	Intra-chromosomal; Deletion - chr2:137699753-chr2:137700957	Not applicable	15,600	PARE	4	45	M	X
MEL 28	Melanoma	4000	Inter-chromosomal; chr12:42334536-chr11:63813127	Not applicable	90	PARE	4	55	F	X
MEL 30	Melanoma	5000	TERT promoter, chr5 g.1295250G>A	Not applicable	288	SafeSeqS	4			X
MELP1-1	Melanoma	5000	BRAF c.1799T>A	BRAF p.V600E	345	PCR-Ligation	4	48	F	X
MELP4-1	Melanoma	2000	BRAF c.1799T>A	BRAF p.V600E	0.0	PCR-Ligation	4	35	M	X
MELP5-1	Melanoma	5000	BRAF c.1799T>A	BRAF p.V600E	220	SafeSeqS	4	68	M	X
MELP6-1	Melanoma	5000	BRAF c.1799T>A	BRAF p.V600E	57	PCR-Ligation	4	43	F	X
MELP7-1	Melanoma	5000	BRAF c.1799T>A	BRAF p.V600E	4.7	PCR-Ligation	4	54	M	X
MELP8-1	Melanoma	5000	BRAF c.1405G>C	BRAF p.G469R	0.0	SafeSeqS	4			X
NB 2464	Neuroblastoma	1500	Intra-chromosomal; Inversion - chr12:6549283-chr12:6539393	Not applicable	0.0	PARE	4			X
NB 2870	Neuroblastoma	1000	Intra-chromosomal; Duplication - chr2:16365880-chr2:15815798	Not applicable	4,050	PARE	4			X
NB 2885	Neuroblastoma	1000	Intra-chromosomal; Duplication - chr19:46265851-chr19:46348530	Not applicable	19	PARE	4			X
NB 2885 F	Neuroblastoma	1000	Intra-chromosomal; Duplication - chr19:46265851-chr19:46348530	Not applicable	0.0	PARE	4			X
NB 6321 6	Neuroblastoma	1500	Intra-chromosomal; Duplication - chr8:77658113-chr8:77668843	Not applicable	0.0	PARE	4			X
NB01	Neuroblastoma	1500	Intra-chromosomal; Inversion - chr2:6705809-chr2:19245803	Not applicable	222,000	PARE	4			X
NB02	Neuroblastoma	2200	Intra-chromosomal; Deletion - chr2:30366503-chr17:44326802	Not applicable	680	PARE	4			X
NB03	Neuroblastoma	2000	Intra-chromosomal; Inversion - chr2:15889847-chr2:16118423	Not applicable	925,000	PARE	4			X
NB04	Neuroblastoma	2600	Intra-chromosomal; Inversion - chr2:65488059-chr11:77699660	Not applicable	243,000	PARE	4			X
CB LUNG 17	Non-Small Cell Lung	2000	TP53 c.281C>A	TP53 p.S94X	100	SafeSeqS	2	67	M	
CB LUNG 19	Non-Small Cell Lung	2000	TP53 c.1045G>T	TP53 p.E349X	4.0	SafeSeqS	1	72	M	
CB LUNG 20	Non-Small Cell Lung	2000	TP53 c.569C>T	TP53 p.P190L	2.5	SafeSeqS	2	77	M	
CB LUNG 22	Non-Small Cell Lung	2000	ZBTB4 c.1469G>T	ZBTB4 p.G490V	0.0	SafeSeqS	2	77	F	
CB LUNG 23	Non-Small Cell Lung	1000	SFRP5 c.139T>C	SFRP5 p.Y47H	909	SafeSeqS	2	77	F	
1110	Ovarian	2000	TP53 c.524G>A	TP53 p.R175H	14,000	SafeSeqS	3	47	F	X
CB01-1	Ovarian	2000	TP53 c.614A>G	TP53 p.Y205C	180	PCR-Ligation	3	67	F	X
CB05-1	Ovarian	2000	TP53 c.725G>T	TP53 p.C242F	80,000	PCR-Ligation	3	53	F	X
CB06-1	Ovarian	2000	TP53 c.487_488insGAT	TP53 p.Y163X	405	PCR-Ligation	3	54	F	X
CB09-1	Ovarian	2000	TP53 c.818G>A	TP53 p.R273H	200	PCR-Ligation	3	44	F	X
CB10-1	Ovarian	2000	TP53 c.843C>A	TP53 p.D281E	250	PCR-Ligation	3	74	F	X
CB11-1	Ovarian	2000	TP53 c.742C>T	TP53 p.R248W	10.0	PCR-Ligation	3	64	F	X
OV 101	Ovarian	5000	TP53 c.472C>G	TP53 p.R158G	2.5	SafeSeqS	2			
PAP 36	Ovarian	5000	PIK3CA c.1637A>T	PIK3CA p.Q546L	0.0	SafeSeqS	1			
CBPANC 0	Pancreatic Ductal	1000	KRAS c.35G>A	KRAS p.G12D	625	BEAMing	4	53	M	X
CBPP12-1	Pancreatic Ductal	4500	KRAS c.35G>A	KRAS p.G12D	10.0	PCR-Ligation	4	63	F	X
CBPP15-1	Pancreatic Ductal	4500	KRAS c.35G>T	KRAS p.G12V	2.2	PCR-Ligation	4	62	M	X
CBPP16-1	Pancreatic Ductal	4500	KRAS c.35G>T	KRAS p.G12V	5,350	PCR-Ligation	2	57	M	
CBPP19-1	Pancreatic Ductal	4500	KRAS c.35G>T	KRAS p.G12V	5.0	PCR-Ligation	2	67	M	
CBPP2-1	Pancreatic Ductal	4500	KRAS c.35G>A	KRAS p.G12D	18	PCR-Ligation	2	44	M	
CBPP21-1	Pancreatic Ductal	1350	KRAS c.35G>T	KRAS p.G12V	84	PCR-Ligation	1	62	F	
CBPP24-1	Pancreatic Ductal	4673	KRAS c.35G>A	KRAS p.G12D	2.0	PCR-Ligation	2	74	F	
CBPP25-1	Pancreatic Ductal	2804	KRAS c.35G>A	KRAS p.G12D	24	BEAMing	2	73	M	
CBPP26-1	Pancreatic Ductal	4673	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	2	53	M	
CBPP28-1	Pancreatic Ductal	2804	KRAS c.35G>T	KRAS p.G12V	9.0	PCR-Ligation	2	58	M	
CBPP3-1	Pancreatic Ductal	4500	KRAS c.35G>T	KRAS p.G12V	10.0	PCR-Ligation	2	50	M	
CBPP31-1	Pancreatic Ductal	4800	KRAS c.35G>T	KRAS p.G12V	16	PCR-Ligation	4	78	M	X
CBPP32-1	Pancreatic Ductal	4800	KRAS c.34G>C	KRAS p.G12R	265	PCR-Ligation	4	68	M	X
CBPP34-1	Pancreatic Ductal	4800	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	2	61	F	
CBPP35-1	Pancreatic Ductal	4800	KRAS c.35G>A	KRAS p.G12D	13	PCR-Ligation	2	72	M	
CBPP36-1	Pancreatic Ductal	4800	KRAS c.35G>T	KRAS p.G12V	0.0	PCR-Ligation	2	59	F	
CBPP37-1	Pancreatic Ductal	4800	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	1	63	M	
CBPP38-1	Pancreatic Ductal	4800	KRAS c.35G>A	KRAS p.G12D	30	PCR-Ligation	2	85	M	
CBPP39-1	Pancreatic Ductal	2400	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	2	54	F	
CBPP40-1	Pancreatic Ductal	4320	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	3	59	M	
CBPP41-1	Pancreatic Ductal	4800	KRAS c.34G>C	KRAS p.G12R	1.3	PCR-Ligation	2	71	M	
CBPP42-1	Pancreatic Ductal	4800	KRAS c.35G>T	KRAS p.G12V	2.1	PCR-Ligation	2	54	F	
CBPP44-1	Pancreatic Ductal	4800	KRAS c.35G>T	KRAS p.G12V	0.0	PCR-Ligation	1	73	M	
CBPP46-1	Pancreatic Ductal	1920	KRAS c.35G>A	KRAS p.G12D	19	PCR-Ligation	2	71	M	
CBPP47-1	Pancreatic Ductal	2304	KRAS c.34G>C	KRAS p.G12R	0.0	PCR-Ligation	2	67	F	
CBPP48-1	Pancreatic Ductal	4800	KRAS c.35G>A	KRAS p.G12D	6.0	PCR-Ligation	1	49	M	
CBPP49-1	Pancreatic Ductal	4000	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	2	82	F	
CBPP50-1	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	10.0	PCR-Ligation	4	66	M	X
CBPP52-1	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	2,310	PCR-Ligation	2	66	F	
CBPP53-1	Pancreatic Ductal	4000	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	2	73	M	
CBPP54-1	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	2	72	M	
CBPP55-1	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	19	PCR-Ligation	4	73	F	X
CBPP57	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	3.3	PCR-Ligation	4	46	M	X
CBPP60	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	16	PCR-Ligation	4	66	F	X
CBPP6-1	Pancreatic Ductal	4500	KRAS c.35G>T	KRAS p.G12V	10.0	PCR-Ligation	1	63	M	
CBPP62	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	39	PCR-Ligation	2	53	F	
CBPP64	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	19	PCR-Ligation	4	73	F	X
CBPP65	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	100	PCR-Ligation	2	64	M	
CBPP67	Pancreatic Ductal	2368	KRAS c.35G>T	KRAS p.G12V	46	PCR-Ligation	4	75	F	X
CBPP68	Pancreatic Ductal	1480	KRAS c.35G>T	KRAS p.G12V	0.0	PCR-Ligation	3	67	M	
CBPP7-1	Pancreatic Ductal	4050	KRAS c.35G>T	KRAS p.G12V	3.0	PCR-Ligation	4	77	M	X
CBPP73	Pancreatic Ductal	5000	KRAS c.34G>T	KRAS p.G12C	57	PCR-Ligation	4	66	F	X
CBPP74	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	2.3	PCR-Ligation	2	78	F	
CBPP8-1	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	0.0	PCR-Ligation	2	77	M	
CBPRO8	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	20	SafeSeqS	2	74	M	
PANC 10	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	24	PCR-Ligation	4	74	M	X

Table 5-S1. Continued.

PANC 106	Pancreatic Ductal	2000	KRAS c.35G>T	KRAS p.G12V	68	SafeSeqS	1	68	M	
PANC 112	Pancreatic Ductal	2000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	46	F	
PANC 121	Pancreatic Ductal	2000	KRAS c.38G>A	KRAS p.G13D	5.5	SafeSeqS	1	68	F	
PANC 13	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	9.5	SafeSeqS	4	66	M	X
PANC 135	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	59	F	
PANC 137	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	7.1	SafeSeqS	2	53	F	
PANC 138	Pancreatic Ductal	3000	KRAS c.34G>T	KRAS p.G12C	275	SafeSeqS	2	79	M	
PANC 139	Pancreatic Ductal	4000	KRAS c.34G>T	KRAS p.G12C	6.1	SafeSeqS	2	74	M	
PANC 14	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	4	59	M	X
PANC 141	Pancreatic Ductal	2000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	65	F	
PANC 144	Pancreatic Ductal	2000	KRAS c.35G>A	KRAS p.G12D	135,000	SafeSeqS	2			
PANC 146	Pancreatic Ductal	3000	KRAS c.35G>T	KRAS p.G12V	1.9	SafeSeqS	2	78	M	
PANC 147	Pancreatic Ductal	3000	KRAS c.35G>T	KRAS p.G12V	10	SafeSeqS	4	81	M	X
PANC 148	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	68	M	
PANC 149	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	168	SafeSeqS	2	64	F	
PANC 154	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	1	60	F	
PANC 155	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	517	SafeSeqS	4	53	M	X
PANC 156	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	2,820	SafeSeqS	4	63	F	X
PANC 157	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	18	SafeSeqS	4	79	F	X
PANC 158	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	197	SafeSeqS	4	79	M	X
PANC 160	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	3	79	M	
PANC 161	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	4	65	M	X
PANC 162	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	3,560	SafeSeqS	4	75	M	X
PANC 163	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	1,630	SafeSeqS	4	59	F	X
PANC 164	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	326	SafeSeqS	4	55	M	X
PANC 165	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	3	55	F	
PANC 167	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	1.1	SafeSeqS	2	67	M	
PANC 17	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	74	PCR-Ligation	4	66	M	X
PANC 171	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	1	75	M	
PANC 18	Pancreatic Ductal	4000	KRAS c.35G>T	KRAS p.G12V	0.0	PCR-Ligation	2	68	F	
PANC 203	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	1	80	F	
PANC 233	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	1	80	M	
PANC 29	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	1.9	PCR-Ligation	1	67	M	
PANC 33	Pancreatic Ductal	4000	KRAS c.35G>A	KRAS p.G12D	1.5	PCR-Ligation	2	58	F	
PANC 45	Pancreatic Ductal	4500	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	2	60	F	
PANC 56	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	55	F	
PANC 69	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	63	M	
PANC 70	Pancreatic Ductal	1000	KRAS c.35G>A	KRAS p.G12D	1,600	SafeSeqS	2	73	M	
PANC 71	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	1.2	SafeSeqS	2	78	F	
PANC 72	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	2.1	SafeSeqS	2	47	F	
PANC 76	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	83	F	
PANC 77	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	67	M	
PANC 78	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	46	F	
PANC 79	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	2.8	SafeSeqS	2	76	F	
PANC 80	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	5.2	SafeSeqS	2	83	F	
PANC 81	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	63	M	
PANC 82	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	2.6	SafeSeqS	2	44	M	
PANC 83	Pancreatic Ductal	5000	KRAS c.34G>A	KRAS p.G12S	1.1	SafeSeqS	1	65	F	
PANC 84	Pancreatic Ductal	4000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	46	M	
PANC 85	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	4	60	F	X
PANC 86	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	61	M	
PANC 87	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	1	60	F	
PANC 9	Pancreatic Ductal	3000	KRAS c.35G>T	KRAS p.G12V	875	PCR-Ligation	4	80	M	X
PANC 91	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	122	SafeSeqS	2			
PANC 92	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	1.1	SafeSeqS	2	56	F	
PANC 93	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	3.5	SafeSeqS	4	56	M	X
PANC 94	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	48	M	
PANC 95	Pancreatic Ductal	4000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	84	M	
PANC 96	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	3.3	SafeSeqS	2	55	M	
PANC 97	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	2.1	SafeSeqS	2	62	F	
PANC 98	Pancreatic Ductal	1500	KRAS c.35G>C	KRAS p.G12A	4.7	SafeSeqS	1			
PANC 99	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	1.2	SafeSeqS	2			
PANC100	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	64	F	
PANC101	Pancreatic Ductal	2000	KRAS c.38G>A	KRAS p.G13D	8.3	SafeSeqS	2	68	M	
PANC102	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	54	M	
PANC103	Pancreatic Ductal	3000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	1	74	M	
PANC104	Pancreatic Ductal	3000	KRAS c.35G>T	KRAS p.G12V	45	SafeSeqS	4	75	M	X
PANC105	Pancreatic Ductal	2000	KRAS c.35G>T	KRAS p.G12V	5.5	SafeSeqS	2	61	F	
PANC107	Pancreatic Ductal	2000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	3	66	M	
PANC108	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	81	F	
PANC109	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	75	M	
PANC110	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	71	F	
PANC111	Pancreatic Ductal	3500	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	63	F	
PANC113	Pancreatic Ductal	3000	KRAS c.34G>T	KRAS p.G12C	6.7	SafeSeqS	2	70	M	
PANC114	Pancreatic Ductal	2000	KRAS c.34G>T	KRAS p.G12C	0.0	SafeSeqS	2	88	M	
PANC115	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	1		M	
PANC116	Pancreatic Ductal	2000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	83	M	
PANC117	Pancreatic Ductal	2000	KRAS c.34G>C	KRAS p.G12R	601	SafeSeqS	2	89	M	
PANC118	Pancreatic Ductal	1000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	65	M	
PANC119	Pancreatic Ductal	2000	KRAS c.35G>A	KRAS p.G12D	11	SafeSeqS	2	80	M	
PANC120	Pancreatic Ductal	2000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	67	F	
PANC122	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	3.7	SafeSeqS	2	72	F	
PANC124	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	4	59	M	X
PANC125	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	5.5	SafeSeqS	2	74	M	
PANC126	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	1	60	F	
PANC127	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	54	F	
PANC128	Pancreatic Ductal	2500	KRAS c.35G>T	KRAS p.G12V	4.4	SafeSeqS	2	62	F	
PANC129	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	1.1	SafeSeqS	2	56	F	
PANC130	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	62	F	
PANC131	Pancreatic Ductal	4000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	62	M	
PANC132	Pancreatic Ductal	4000	KRAS c.35G>A	KRAS p.G12D	107	SafeSeqS	4	73	F	X
PANC133	Pancreatic Ductal	3000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	75	M	
PANC134	Pancreatic Ductal	4000	KRAS c.34G>C	KRAS p.G12R	4.0	SafeSeqS	1	65	M	
PANC136	Pancreatic Ductal	4000	KRAS c.34G>C	KRAS p.G12R	0.0	SafeSeqS	2	58	F	
PANC140	Pancreatic Ductal	2000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	1	50	M	
PANC142	Pancreatic Ductal	2500	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	76	M	
PANC143	Pancreatic Ductal	3000	KRAS c.35G>A	KRAS p.G12D	9.7	SafeSeqS	2	67	M	
PANC145	Pancreatic Ductal	3000	KRAS c.34G>C	KRAS p.G12R	7.2	SafeSeqS	2	60	M	
PANC150	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	122	SafeSeqS	4	77	M	X
PANC151	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	6.3	SafeSeqS	2	83	F	
PANC152	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	5.7	SafeSeqS	4	83	M	X
PANC153	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	1.6	SafeSeqS	2	69	F	
PANC159	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	1	49	F	
PANC166	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	76	M	
PANC168	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	2	74	M	
PANC169	Pancreatic Ductal	5000	KRAS c.34G>C	KRAS p.G12R	3.0	SafeSeqS	2	54	M	
PANC170	Pancreatic Ductal	5000	KRAS c.35G>A	KRAS p.G12D	0.0	SafeSeqS	2	71	M	
PANC172	Pancreatic Ductal	5000	KRAS c.35G>T	KRAS p.G12V	0.0	SafeSeqS	1	68	F	
CBPP11-1	Prostate	4500	KRAS c.35G>A	KRAS p.G12D	0.0	PCR-Ligation	4	62	M	X
CBPRP1	Prostate	4800	BRCA2 c.9281C>G	BRCA2 p.S3094X	173,000	PCR-Ligation	4	63	M	X
CBPRP2	Prostate	4800	TP53 c.743G>T	TP53 p.R248L	59	PCR-Ligation	4	63	M	X
CBPRP3	Prostate	4800	SPOP c.304T>G	SPOP p.F102V	0.0	SafeSeqS	4	57	M	X
CBPRP5-1	Prostate	5000	RPP30 c.905G>A	RPP30 p.R302K	0.0	SafeSeqS	4	68	M	X
RCC 10	Renal Cell Carcinoma	5000	VHL c.263G>T	VHL p.W88L	0.0	SafeSeqS	4	70	M	X
RCC 11	Renal Cell Carcinoma	5000	VHL c.263G>T	VHL p.W88L	0.0	SafeSeqS	4	73	M	X
RCC 7	Renal Cell Carcinoma	5000	HOOK2 c.1165G>T	HOOK2 p.E389X	0.0	SafeSeqS	4	77	M	X
RCC PL1	Renal Cell Carcinoma	5000	VHL c.388G>C	VHL p.V130L	24	PCR-Ligation	4	49	M	X
RCC PL2	Renal Cell Carcinoma	5000	MET c.3687_3688GT>TA	MET p.M1229_Y1230IN	665	PCR-Ligation	4	86	M	X
CB LUNG 24	Small Cell Lung Cancer	5000	TP53 c.832C>G	TP53 p.P278A	320	SafeSeqS	4			X
CB THY 4	Thyroid	5000	NRAS c.182A>G	NRAS p.Q61R	0.0	SafeSeqS	4	79	M	X
CB THY 5	Thyroid	5000	NF2 c.20C>G	NF2 p.S7C	0.0	SafeSeqS	4	54	M	X
CB THY 7	Thyroid	5000	TP53 c.743G>A	TP53 p.R248Q	1,350	SafeSeqS	4	56	M	X
THY 1	Thyroid	5000	BRAF c.1799T>A	BRAF p.V600E	0.0	SafeSeqS	4	64	F	X

Table 5-S2. Comparison between circulating tumor DNA fragments containing point mutations vs. rearrangements.

ID#	Tumor Type	Clinical Stage	Rearrangement	Rearrangement (mutant fragments per 5 mL)	Point Mutation (nucleotide)	Point Mutation (codon)	Point mutation (mutant fragments per 5 mL)	Rearrangement fragments/Point mutation fragments (ratio)
BLD 24	Bladder	2	Inter-chromosomal; chr2:204396710-chr1:154286405	3.9	BRAF c.1801A>G	BRAF p.K601E	3.4	1.1
CRC 02	Colorectal	1	Inter-chromosomal; chr12:73097777-chr1:63099083	22,000	TP53 c.730G>A	TP53 p.G244S	16,115	1.4
CRC 03	Colorectal	1	Intra-chromosomal; chr11:98914172-chr11:98993460	6,415	APC c. 4012C>T	APC p.Q1338X	26,155	0.2
CRC 06	Colorectal	1	Intra-chromosomal; chr13:27679603-chr13:25685009	85	KRAS c.182_183AA>CC	KRAS p.Q61P	59	1.4
CRC 07	Colorectal	1	Intra-chromosomal; chr4:85109657-chr4:85408635	42	KRAS c.35G>A	KRAS p.G12D	49	0.9
CRC 12	Colorectal	4	Intra-chromosomal; chr6:58080933-chr6:67229823	79	KRAS c.35G>A	KRAS p.G12D	34	2.3
CRC 14	Colorectal	4	Intra-chromosomal; chr8:141375810-chr8:141405769	31	TP53 c.818G>A	TP53 p.R273H	95	0.3
CRC 21	Colorectal	2	Intra-chromosomal; chr1:237741691-chr1:244145093	685	KRAS c.35G>C	KRAS p.G12A	745	0.9
CRC 23	Colorectal	4	Intra-chromosomal; chr16:6343641-chr16:6727736	28,500	TP53 c.817C>T	TP53 p.R273C	14,500	2.0
CRC 27	Colorectal	2	Intra-chromosomal; chr13:102839339-chr13:105470344	55	KRAS c.35G>A	KRAS p.G12D	51	1.1
CRC 30	Colorectal	3	Intra-chromosomal; chr20:2658058-chr20:2770031	1,466,665	KRAS c.35G>A	KRAS p.G12D	138,445	10.6
CRC 37	Colorectal	1	Intra-chromosomal; Duplication - chr3:170782510-chr3:170870975	0.0	TP53 c.844C>T	TP53 p.R282W	38	0.0
CRC 51	Colorectal	4	Inter-chromosomal; chr8:30060122-chr1:190065347	3,850	APC c.4012C>T	APC p.Q1338X	72	53.5
CRC 53	Colorectal	4	Intra-chromosomal; chr16:52977177-chr16:52965488	7,150	TP53 c.452C>A	TP53 p.P151H	1,085	6.6
CRC 54	Colorectal	4	Intra-chromosomal; chr11:7250750-chr11:19457296	1,240	KRAS c.35G>T	KRAS p.G12V	36	34.4
CRC 55	Colorectal	4	Intra-chromosomal; chr8:38281631-chr8:39225849	295	TP53 c.637C>T	TP53 p.R213X	112	2.6
CRC 58	Colorectal	4	Intra-chromosomal; chr20:14839816-chr20:14863644	1,370	KRAS c.35G>A	KRAS p.G12D	1,380	1.0
CRC 60	Colorectal	4	Intra-chromosomal; chr17:35277642-chr17:35163714	73,335	TP53 c.817C>T	TP53 p.R273C	12,905	5.7
MEL 27	Melanoma	4	Intra-chromosomal; chr2:137699753-chr2:137700957	15,600	BRAF c.1799T>A	BRAF p.V600E	260	60.0

Table 5-S3. Comparison between plasma and tumor tissue KRAS status in 206 patients with metastatic colorectal cancer.

Sample ID	Plasma Mutation	mutant fragments per 5 ml	Tissue Mutation	On Therapy during blood draw?	Age	Gender	Race	ECOG PS	Met Site - liver	Met Site - lungs	Met Site - peritoneum	Met Site - pelvis	Met Site - bone	Met Site - brain	# of Met Sites	Histological subtype	Mucinous	Smoking Status	Family History	MSI Status	BRAF Status	Vital Status (1=deceased)	OS from ctDNA meaurement (days)	OS from diagnosis (days)	Response to 1st line therapy	Prior Surgery	Prior chemoth erapy	Prior Radiation	CEA (+/- 8 weeks)	WBC	ALT	Bilirubin	Albumin	Creatinine
BARD 001	None Detected	Not applicable	WT	Yes	54	M	W	0	1	0	0	0	0	0	1	Poorly	No	No	No	Unknown	WT	0	933	1112.00	Partial	No	Yes	Yes	39	3800	8	0.7	3.7	0.9
BARD 002	None Detected	Not applicable	WT	Yes	58	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	672	706.00	Complete	Yes	Yes	No	1	6500	41	Unknown	5.0	0.9
BARD 003	KRAS p.G13D	9,339	KRAS p.G13D	Yes	48	M	W	0	1	1	0	0	0	0	2	Poorly	No	No	No	Unknown	WT	0	118	308.00	Progression	Yes	Yes	No	3	6500	35	0.7	4.7	0.9
BARD 004	None Detected	Not applicable	WT	Yes	67	F	W	0	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	utation four	0	318	498.00	Partial	Yes	Yes	No	2	4900	14	0.5	3.9	0.7
BARD 005	None Detected	Not applicable	WT	Yes	59	M	W	0	0	0	1	1	0	0	2	Moderately	No	No	No	Unknown	WT	0	298	1600.00	Stable Desease	Yes	Yes	Yes	6	7400	26	0.8	4.4	1.0
BARD 006	None Detected	Not applicable	WT	Yes	79	M	W	0	1	1	1	0	0	0	3	Moderately	No	No	No	Unknown	WT	0	199	2723.00	Partial	Yes	Yes	No	2	8100	15	Unknown	4.2	1.1
BARD 007	None Detected	Not applicable	WT	Yes	69	M	W	0	1	0	1	0	0	0	2	Poorly	Yes	No	No	Unknown	utation four	1	291	956.00	Progression	Yes	Yes	No	Unknown	4300	17	Unknown	4.0	1.0
BARD 008	None Detected	Not applicable	WT	No	72	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	1	95	108.00	Unknown	No	No	No	25	7300	17	Unknown	3.6	2.8
BARD 009	None Detected	Not applicable	WT	Yes	76	M	W	0	1	1	0	0	0	0	2	Moderately	No	Yes	No	Unknown	WT	0	195	418.00	Progression	Yes	Yes	No	28	4400	20	0.9	4.1	0.5
BARD 010	None Detected	Not applicable	WT	No	61	M	W	0	1	0	0	0	0	0	1	Moderately	Yes	No	No	Unknown	WT	0	180	215.00	Stable Desease	No	Yes	No	107	5900	15	Unknown	4.3	0.9
BARD 011	None Detected	Not applicable	WT	No	69	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	733	783.00	Partial	No	No	No	5	5300	22	Unknown	3.7	0.7
BARD 012	KRAS p.G12D	93	KRAS p.G12D	No	83	F	W	0	0	0	1	0	0	0	1	Moderately	Yes	No	No	Unknown	WT	0	350	1498.00	Unknown	Yes	No	No	1	5600	16	1.0	3.7	0.7
BARD 013	None Detected	Not applicable	WT	Yes	85	F	W	0	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	1	290	2097.00	Progression	Yes	Yes	No	52	9600	13	0.6	4.4	0.9
BARD 014	KRAS p.G13D	2,400	KRAS p.G13D	Yes	60	M	W	0	0	1	1	0	0	0	2	Moderately	No	No	No	Unknown	WT	1	208	1809.00	Partial	Yes	Yes	No	34	7200	10	0.7	4.1	1.0
BARD 015	None Detected	Not applicable	KRAS p.G12D	Yes	60	M	W	0	0	1	0	0	0	0	1	Poorly	No	No	No	Unknown	WT	0	428	888.00	Partial	Yes	Yes	No	1	3900	16	0.6	4.7	1.0
BARD 016	None Detected	Not applicable	WT	No	61	M	W	0	1	0	1	0	0	0		Moderately	No	No	No	Unknown	WT	0	56	864.00	Progression	Yes	Yes	No	335	2600	156	1.8	3.4	0.8
BARD 017	KRAS p.G12D	1,050	KRAS p.G12D	Yes	82	M	W	0	0	0	1	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	149	520.00	Progression	Yes	Yes	No	4	4800	9	Unknown	4.1	0.8
BARD 018	None Detected	Not applicable	WT	Yes	49	M	W	0	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	523	1279.00	Stable Desease	Yes	Yes	No	5	8900	19	1.0	4.2	0.9
BARD 019	None Detected	Not applicable	WT	Yes	71	M	W	0	1	1	0	0	0	0	2	Moderately	Yes	No	No	Unknown	WT	0	910	2211.00	Stable Desease	Yes	Yes	No	6	3800	15	1.7	4.3	1.1
BARD 020	None Detected	Not applicable	WT	Yes	71	M	W	0	0	1	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	1	717	799.00	Stable Desease	Yes	No	No	2	7600	16	Unknown	4.6	0.7
BARD 021	None Detected	Not applicable	WT	Yes	69	F	W	0	1	0	0	0	0	0	1	Moderately	Yes	No	No	Unknown	utation four	0	383	1170.00	Unknown	Yes	Yes	Yes	1	10800	16	0.6	4.2	0.5
BARD 022	None Detected	Not applicable	WT	Yes	75	M	W	0	1	1	1	0	0	0	3	Unknown	Unknown	No	No	Unknown	WT	0	489	914.00	Stable Desease	No	Yes	No	5	10500	20	0.5	4.5	1.0
BARD 023	None Detected	Not applicable	WT	Yes	71	M	W	0	1	1	1	0	0	0	3	Moderately	No	No	No	Unknown	WT	0	616	1992.00	Progression	Yes	Yes	No	145	7400	9	Unknown	3.2	2.7
BARD 024	KRAS p.G13D	2,330	KRAS p.G13D	Yes	80	M	W	0	0	1	1	1	0	0	4	Poorly	Yes	No	No	Unknown	WT	1	149	828.00	Progression	Yes	Yes	Yes	83	8200	Unknown	0.5	4.2	0.8
BARD 026	None Detected	Not applicable	WT	No	61	M	W	0	1	0	0	0	0	0	1	Moderately	Unknown	No	No	Unknown	WT	0	73	158.00	Unknown	Yes	No	No	2	7300	11	Unknown	4.2	0.8
BARD 027	None Detected	Not applicable	WT	No	25	F	W	0	0	0	1	0	0	0	1	Moderately	No	No	No	Unknown	WT	1	330	831.00	Stable Desease	Yes	Yes	No	25	4200	45	Unknown	4.9	0.7
BARD 028	KRAS p.G12S	17,300	KRAS p.G12S	Yes	61	F	W	0	1	1	1	0	0	0	3	Moderately	Unknown	No	No	Unknown	WT	1	293	329.00	Progression	Yes	Yes	No	1084	3700	27	Unknown	4.1	0.5
BARD 029	KRAS p.G12D	13,700	KRAS p.G12D	Yes	51	F	W	0	1	1	1	0	1	0	4	Moderately	No	No	No	Unknown	WT	0	159	145.00	Partial	Yes	No	No	529	10900	7	Unknown	3.8	0.9
BARD 030	None Detected	Not applicable	KRAS p.G12V	No	69	M	W	0	0	0	0	0	1	0	1	Moderately	No	No	No	Unknown	WT	0	380	418.00	Unknown	Yes	No	No	2	6600	9	Unknown	4.1	1.1
BARD 031	None Detected	Not applicable	WT	No	49	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	297	330.00	Partial	Yes	No	No	743	9700	19	Unknown	3.8	0.9
BARD 032	None Detected	Not applicable	WT	No	82	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	14	29.00	Unknown	No	Yes	No	27	7900	14	Unknown	3.8	1.0
BARD 033	KRAS p.G12D	5,260	KRAS p.G12D	No	38	F	W	0	1	1	1	0	0	0	3	Moderately	Yes	No	No	Unknown	WT	1	368	666.00	Progression	Yes	Yes	No	7	10800	17	Unknown	4.5	0.6
BARD 034	KRAS p.G12A	158,000	WT	Yes	74	M	W	0	1	0	1	0	1	1	4	Moderately	No	No	No	Unknown	WT	1	22	40.00	Unknown	No	Yes	Yes	1000	7100	126	0.9	2.7	0.5
BARD 035	KRAS p.G13D	381	KRAS p.G13D	No	61	M	W	0	1	0	1	1	0	0	3	ll to moderat	No	No	No	Unknown	WT	0	586	1210.00	Partial	Yes	Yes	Yes	Unknown	5700	40	Unknown	4.5	1.0
BARD 036	None Detected	Not applicable	WT	Yes	60	F	W	0	1	1	1	0	0	0	3	Moderately	No	No	No	Unknown	WT	1	41	1329.00	Partial	Yes	Yes	No	65	3700	19	0.7	4.3	0.7
BARD 037	KRAS p.G12C	60,800	KRAS p.G12V	Yes	62	F	W	0	1	0	1	0	0	0	2	Poorly	Unknown	No	No	Unknown	WT	1	113	137.00	Unknown	No	Yes	Yes	2	12900	22	0.7	3.7	0.5
BARD 038	None Detected	Not applicable	WT	No	60	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	616	1328.00	Partial	No	Yes	No	100	4300	16	Unknown	4.7	0.9
BARD 039	None Detected	Not applicable	WT	Yes	38	F	W	0	0	0	1	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	255	1007.00	Partial	Yes	Yes	No	72	4800	39	Unknown	3.5	0.6
BARD 040	None Detected	Not applicable	WT	Yes	67	M	W	0	1	0	1	0	0	0	2	ll to moderat	No	No	No	Unknown	WT	0	202	773.00	Partial	Yes	Yes	No	4	8600	15	Unknown	4.1	1.4
BARD 041	None Detected	Not applicable	WT	Yes	76	M	W	0	1	1	1	0	0	0	3	Moderately	No	No	No	Unknown	WT	0	851	1606.00	Partial	Yes	Yes	No	25	8500	Unknown	1.2	4.0	1.1
BARD 042	None Detected	Not applicable	WT	Yes	66	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	988	1624.00	Stable Desease	Yes	Yes	No	3	8200	56	Unknown	4.8	1.0
BARD 043	None Detected	Not applicable	WT	Yes	63	F	W	0	1	1	0	0	0	0	2	Poorly	No	No	No	Unknown	WT	0	105	333.00	Partial	Yes	Yes	No	8	7400	8	Unknown	Unknown	0.6
BARD 044	None Detected	Not applicable	WT																															
BARD 045	KRAS p.G12D	224	KRAS p.G12D	No	71	M	W	1	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	1	199	1265.00	Partial	Yes	Yes	Yes	21	5700	25	0.6	3.6	0.6
BARD 046	None Detected	Not applicable	WT	Yes	73	M	W	0	0	1	0	0	0	0	1	Poorly	No	No	No	Unknown	WT	1	585	2106.00	Stable Desease	Yes	Yes	Yes	2	7800	26	0.8	4.4	1.0
BARD 047	None Detected	Not applicable	WT	Yes	61	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	937	1424.00	Partial	Yes	Yes	No	7	6300	25	0.7	4.4	0.8
BARD 048	None Detected	Not applicable	WT	Yes	72	M	W	1	0	1	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	289	1980.00	Stable Desease	Yes	Yes	No	4	7000	17	Unknown	4.9	0.8
BARD 049	None Detected	Not applicable	WT	Yes	58	M	W	0	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	1025	2458.00	Stable Desease	Yes	Yes	No	3	4200	Unknown	0.7	4.0	0.9
BARD 050	KRAS p.G12D	485	KRAS p.G12D	Yes	51	F	W	1	0	1	0	0	1	0	2	Unknown	No	No	No	Unknown	WT	0	507	931.00	Partial	No	Yes	Yes	Unknown	8100	17	Unknown	Unknown	Unknown
BARD 051	None Detected	Not applicable	WT	Yes	38	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	0	288	433.00	Partial	Yes	Yes	No	1	9500	196	1.9	4.5	0.5
BARD 053	None Detected	Not applicable	WT	No	74	M	W	1	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	0	703	745.00	Partial	Yes	No	No	617	4800	50	1.0	4.0	1.1
BARD 054	KRAS p.G12S	64	KRAS p.G12S	No	78	M	W	1	1	0	1	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	740	1290.00	Partial	Yes	Yes	No	3	3600	12	0.5	4.7	1.1
BARD 055	KRAS p.G12D	1,760	KRAS p.G12D	No	76	F	W	2	0	1	1	1	1	0	4	Unknown	Unknown	No	No	Unknown	WT	1	229	3264.00	Progression	Yes	Yes	Yes	541	9200	8	0.7	4.3	0.6
BARD 056	KRAS p.G12A	1,250	KRAS p.G12A	Yes	75	M	W	0	1	0	0	0	0	0	1	Moderately	Yes	No	No	Unknown	WT	0	289	420.00	Stable Desease	Yes	Yes	No	13	8900	18	Unknown	3.9	0.8
BARD 057	None Detected	Not applicable	WT	No	56	F	W	0	0	1	1	0	0	0	2	Moderately	Unknown	No	No	Unknown	WT	0	393	1227.00	Partial	Yes	Yes	No	3	3400	7	Unknown	3.9	0.6
BARD 058	None Detected	Not applicable	WT	No	51	F	M	0	1	0	1	0	0	0	2	Unknown	Unknown	Yes	Unknown	Unknown	WT	0	22	197.00	Progression	No	Yes	No	836	9000	25	1.0	3.7	0.6
BARD 059	KRAS p.G12D	11,800	KRAS p.G12D	Yes	78	M	W	3	0	1	1	0	0	0	2	Moderately	No	No	No	Unknown	WT	1	59	264.00	Progression	Yes	Yes	No	29	5400	8	0.7	4.3	0.8
BARD 060	KRAS p.G12D	14,500	KRAS p.G12D	No	76	F	W	1	1	1	0	0	0	0	3	Moderately	Yes	No	No	Unknown	Unknown	0	121	1192.00	Progression	Yes	Yes	No	66	5200	24	0.7	4.3	0.7
BARD 061	None Detected	Not applicable	WT	Yes	75	M	W	1	0	1	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	463	831.00	Progression	Yes	Yes	No	3	8100	Unknown	Unknown	3.8	0.9
BARD 062	None Detected	Not applicable	WT	No	69	M	W	1	1	0	1	1	0	0	3	ll to moderat	No	No	No	Unknown	WT	1	201	890.00	Progression	Yes	Yes	No	238	9200	17	0.7	Unknown	1.0
BARD 063	None Detected	Not applicable	WT	Yes	75	M	W	1	0	1	1	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	98	1965.00	Partial	Yes	Yes	Yes	27	4000	Unknown	0.6	4.2	1.0
BARD 064	None Detected	Not applicable	WT	Yes	70	M	W	2	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	578	1249.00	Partial	Yes	Yes	No	33	6700	Unknown	Unknown	4.4	0.7
BARD 065	None Detected	Not applicable	WT	No	57	M	W	2	0	0	1	1	0	0	2	Poorly	No	No	No	Unknown	Unknown	0	191	742.00	Unknown	Yes	Yes	Yes	49	9800	22	Unknown	3.4	0.7

Table 5-S3. Continued.

BARD 066	None Detected	Not applicable	WT	No	37	M	W	2	1	1	1	0	0	0	3	Poorly	No	No	No	Unknown	WT	0	65	408.00	Stable Desease	Yes	Yes	No	2057	10800	32	Unknown	Unknown	0.9
BARD 067	None Detected	Not applicable	WT	Yes	29	F	W	1	0	1	1	1	0	0	3	Moderately	No	No	No	Unknown	WT	0	657	2733.00	Partial	Yes	Yes	Yes	Unknown	5600	37	0.5	4.2	0.7
BARD 068	KRAS p.G13D	223,000	KRAS p.G13D	No	35	M	W	3	1	0	1	0	0	0	2	Moderately	Unknown	No	Yes	Unknown	WT	0	14	799.00	Partial	Yes	Yes	No	20543	19900	63	3.4	2.8	0.6
BARD 069	KRAS p.G12V	49,700	KRAS p.G12V	No	67	F	W	1	1	1	0	0	0	0	2	Moderately	Unknown	No	No	Unknown	WT	0	64	1651.00	Partial	Yes	Yes	No	874	10600	23	Unknown	Unknown	1.2
BARD 070	KRAS p.G12D	1,270	KRAS p.G12D	Yes	57	M	W	2	1	1	1	0	1	0	4	Moderately	Unknown	No	No	Unknown	Unknown	0	248	693.00	Unknown	Yes	Yes	Yes	3	7400	47	0.6	4.2	0.9
BARD 071	None Detected	Not applicable	WT	Yes	70	F	W	2	1	1	0	0	1	0	3	Moderately	Unknown	No	No	Unknown	Unknown	0	45	926.00	Partial	Yes	Yes	Yes	2930	7800	44	0.7	4.0	0.6
BARD 072	None Detected	Not applicable	WT	No	79	M	W	0	1	0	0	0	0	0	1	Moderately	Unknown	No	No	Unknown	WT	0	220	290.00	Unknown	Yes	Yes	No	73	20100	17	Unknown	3.9	1.0
BARD 073	KRAS p.G13D	178,000	KRAS p.G13D	Yes	42	F	W	2	1	0	0	0	0	0	1	Moderately	Unknown	No	No	MSS	Unknown	0	0	680.00	Partial	Yes	Yes	No	991	17400	53	2.9	3.4	0.6
BARD 074	KRAS p.G12V	3,370	KRAS p.G12V	Yes	39	M	W	1	1	0	0	0	0	0	1	Poorly	No	No	No	Unknown	WT	0	822	860.00	Partial	Yes	Yes	No	16	5300	52	0.6	Unknown	0.7
BARD 075	None Detected	Not applicable	WT	Yes	68	F	W	3	1	0	0	0	0	0	1	Unknown	Unknown	No	No	Unknown	WT	0	47	986.00	Progression	Yes	Yes	Yes	2581	8500	26	Unknown	3.5	0.5
BARD 077	KRAS p.G12V	799,000	KRAS p.G12V	No	58	M	W	1	1	0	0	0	0	0	1	Moderately	Unknown	No	No	Unknown	WT	0	0	289.00	Partial	No	Yes	No	28	6400	89	0.9	3.5	0.8
BARD 078	KRAS p.G12V	134	KRAS p.G12V	No	41	F	W	0	0	1	1	0	0	0	2	Moderately	Yes	No	Yes	Unknown	WT	0	176	527.00	Partial	Yes	Yes	No	19	4500	30	Unknown	4.3	0.6
CRC 085	KRAS p.G12S	1,380	KRAS p.G12S	No	58	F	B	0	1	0	0	0	0	0	1	Moderately	No	Yes	Yes	Unknown	Unknown	1	1177	1199.00	No Therapy	No	No	No	793	5650	12	0.5	4.2	0.7
CRC 086	None Detected	Not applicable	WT	Yes	41	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	0	994	1220.00	Progression	Yes	Yes	No	4	20350	39	0.9	4.4	0.6
CRC 086	None Detected	Not applicable	WT																																
CRC 087	None Detected	Not applicable	WT	No	65	M	W	0	0	0	0	0	0	0	0	ll to moderat	No	No	Yes	Unknown	WT	1	1408	2220.00	No Therapy	Yes	No	No	119	8730	16	0.3	4.3	1.2
CRC 088	None Detected	Not applicable	WT	No	64	M	B	2	1	1	0	0	0	0	2	Unknown	No	Distant	Yes	Unknown	Unknown	1	341	393.00	Progressed	No	No	No	26	12580	23	0.3	2.8	0.8
CRC 089	None Detected	Not applicable	WT	No	55	M	B	1	1	0	1	0	0	0	2	Moderately	Unknown	Distant	Yes	Unknown	Unknown	1	?	321.00	Progressed	No	Yes	No	73	9520	17	Unknown	Unknown	Unknown
CRC 090	None Detected	Not applicable	WT	No	47	M	W	Unknown	1	0	0	0	0	0	1	Unknown	Unknown	No	Yes	Unknown	Unknown	1	59	340.00	Stable	No	Yes	No	3	12070	37	1.4	3.6	0.7
CRC 091	None Detected	Not applicable	WT	No	52	F	W	Unknown	1	0	0	0	0	0	1	Unknown	Unknown	No	No	Unknown	Unknown	1	170	208.00	No Therapy	No	No	No	191	13770	32	0.3	4.0	0.6
CRC 092	KRAS p.G12V	813	KRAS p.G12V	No	56	M	W	0	1	1	1	0	0	0	3	Unknown	Unknown	No	No	Unknown	Unknown	1	95	1307.00	Stable	Yes	Yes	Yes	21	5900	61	0.6	3.9	1.1
CRC 093	None Detected	Not applicable	WT	No	56	F	B	0	1	0	1	0	0	0	2	Moderately	Unknown	No	Yes	Unknown	Unknown	1	350	1597.00	No Therapy	Yes	No	No	89	7670	30	0.7	4.5	0.8
CRC 094	KRAS p.G12V	40,000	KRAS p.G12V	No	43	F	W	0	1	0	0	0	0	0	1	erately to po	No	Distant	Yes	Unknown	WT	1	302	385.00	No Therapy	No	No	No	12	11300	18	0.5	4.0	0.8
CRC 095	KRAS p.G12D	35,800	KRAS p.G12D	No	44	F	W	Unknown	1	0	0	0	0	0	1	Poorly	No	No	No	Unknown	Unknown	1	96	124.00	No Therapy	No	No	No	27	7240	31	0.3	2.9	0.6
CRC 096	None Detected	Not applicable	WT	No	30	M	W	0	1	0	0	0	0	0	1	Moderately	Unknown	No	Yes	Unknown	Unknown	0	1364	1404.00	No Therapy	No	No	No	5	18350	129	1.0	3.2	0.9
CRC 097	None Detected	Not applicable	WT	Yes	60	F	B	0	1	0	0	0	0	0	1	Moderately	No	Distant	Yes	Unknown	Unknown	0	205	500.00	Partial	Yes	Yes	Yes	Unknown	13300	341	0.7	2.2	0.6
CRC 098	None Detected	Not applicable	WT	No	64	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	0	55	1021.00	Stable	Yes	Yes	No	2	11890	42	0.8	4.2	1.2
CRC 099	KRAS p.G12S	323	KRAS p.G12S	No	41	F	W	0	1	0	0	0	0	0	1	Moderately	No	Unknown	Unknown	Unknown	Unknown	1	580	1063.00	Progression	Yes	Yes	No	Unknown	16860	1169	1.0	2.9	0.5
CRC 100	None Detected	Not applicable	WT	Yes	59	F	W	0	1	0	0	0	0	0	1	Moderately	No	Distant	No	Unknown	Unknown	0	180	676.00	Partial	Yes	Yes	No	6	11220	571	1.2	2.6	0.5
CRC 101	None Detected	Not applicable	WT	No	59	M	W	0	0	1	0	0	0	0	1	Poorly	No	Yes	No	Unknown	Unknown	1	553	561.00	No Therapy	Yes	No	No	54	14100	20	0.5	3.4	0.8
CRC 102	None Detected	Not applicable	WT	No	82	M	W	0	0	1	0	0	0	0	1	ll to moderat	No	Distant	Yes	Unknown	Unknown	1	120	136.00	No Therapy	No	No	No	6	5670	20	0.3	3.2	1.0
CRC 104	KRAS p.G12C	1,200	KRAS p.G12C	No	57	F	W	0	1	0	0	0	0	0	1	Unknown	Unknown	No	Yes	Unknown	WT	1	170	648.00	Partial	Yes	Yes	No	5	7770	854	0.9	3.3	0.6
CRC 105	KRAS p.G13D	218	KRAS p.G13D	No	68	M	W	0	1	0	0	0	0	0	1	Moderately	Yes	Unknown	Unknown	Unknown	Unknown	0	934	1174.00	Partial	Yes	Yes	No	Unknown	6660	724	1.6	2.6	1.7
CRC 106	KRAS p.G12D	37,800	KRAS p.G12D	No	71	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	1	217	438.00	Progression	Yes	Yes	No	7	6360	45	0.6	4.1	0.8
CRC 107	None Detected	Not applicable	KRAS p.G12D	Yes	64	F	W	0	1	0	0	0	0	0	1	ll to moderat	Yes	No	No	Unknown	Unknown	0	1064	1381.00	No Therapy	Yes	Yes	No	3	6120	31	0.7	4.6	0.7
CRC 108	None Detected	Not applicable	WT	No	74	F	W	1	0	0	0	1	0	1	1	Moderately	No	No	No	Unknown	utation four	0	1001	1091.00	No Therapy	Yes	No	Yes	7	6290	24	0.3	4.4	0.6
CRC 110	None Detected	Not applicable	WT	No	62	F	A	0	1	0	0	0	0	0	1	ll to moderat	No	No	No	Unknown	Unknown	1	774	3504.00	Stable	Yes	Yes	No	141	7180	14	0.4	4.7	0.5
CRC 111	None Detected	Not applicable	WT	No	38	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	WT	0	845	859.00	No Therapy	No	No	No	272	6520	345	1.0	2.7	0.6
CRC 112	None Detected	Not applicable	WT	No	52	F	W	0	1	0	0	0	0	0	1	Unknown	Unknown	No	No	Unknown	Unknown	0	335	863.00	Stable	Yes	Yes	No	6	5010.0	38.0	Unknown	Unknown	Unknown
CRC 113	KRAS p.G12V	638	KRAS p.G12V	No	35	M	W	0	1	0	0	0	0	0	1	Unknown	Unknown	No	Yes	Unknown	WT	1	494	539.00	No Therapy	No	No	No	2	5830	26	0.3	4.3	0.8
CRC 114	None Detected	Not applicable	WT	No	66	M	W	1	1	1	0	0	0	0	2	Moderately	No	Distant	No	Unknown	WT	0	532	562.00	No Therapy	No	No	No	6	8160	9	0.4	3.7	1.0
CRC 115	None Detected	Not applicable	WT	No	60	M	B	1	1	0	0	0	0	0	1	Moderately	No	Distant	Yes	Unknown	WT	1	346	383.00	No Therapy	No	No	No	335	6410	17	0.4	3.5	0.9
CRC 116	None Detected	Not applicable	WT	No	49	M	W	0	0	0	0	0	0	0	0	Poorly	Yes	Distant	Yes	Unknown	WT	0	645	668.00	No Therapy	No	No	No	Unknown	9840	18	0.3	4.4	0.8
CRC 117	None Detected	Not applicable	WT	No	39	F	W	0	0	0	1	0	0	0	1	Poorly	Yes	No	Yes	MSS	WT	0	428	439.00	No Therapy	No	No	No	55	4410	11	0.1	3.8	0.5
CRC 118	None Detected	Not applicable	WT	No	66	F	B	2	1	0	0	0	0	0	1	Unknown	Unknown	No	No	Unknown	WT	0	333	337.00	No Therapy	No	No	No	13864	15260.0	31.0	0.7	4.1	1.1
CRC 119	None Detected	Not applicable	WT	No	62	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	984	39365.00	No Therapy	Yes	No	No	3	3270	36	1.4	4.3	1.0
CRC 120	None Detected	Not applicable	WT	No	42	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	MSS	WT	0	949	984.00	No Therapy	No	No	No	3	8520	17	0.3	4.0	0.7
CRC 121	None Detected	Not applicable	WT	No	68	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	1	39	233.00	Partial	No	Yes	No	Unknown	7450	17	0.4	4.5	0.9
CRC 122	None Detected	Not applicable	WT	No	70	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	WT	0	217	723.00	Partial	Yes	Yes	No	Unknown	4870	31	0.6	4.2	0.6
CRC 123	None Detected	Not applicable	WT	No	47	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	MSS	Unknown	0	342	809.00	Stable	Yes	Yes	No	Unknown	7490	54	0.7	4.5	0.6
CRC 124	None Detected	Not applicable	WT	No	50	F	B	0	1	0	0	0	0	0	1	ll to moderat	Yes	No	Yes	MSS	WT	0	819	534.00	Complete	No	No	No	Unknown	3910	62	0.6	3.9	0.8
CRC 125	None Detected	Not applicable	KRAS p.G12C	No	56	M	W	0	1	1	0	0	0	0	2	Moderately	No	No	Yes	Unknown	WT	0	823	856.00	No Therapy	No	No	No	3	4170	23	0.7	4.7	0.8
CRC 126	None Detected	Not applicable	WT	No	54	M	W	0	1	0	0	0	0	0	1	ll to moderat	No	Distant	No	Unknown	WT	0	777	1171.00	No Therapy	No	No	No	5	10820	178	1.0	3.0	0.9
CRC 127	KRAS p.G12V	28,400	KRAS p.G12V	No	39	F	W	0	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	818	869.00	No Therapy	No	No	No	Unknown	7130	33	0.5	4.2	0.7
CRC 128	None Detected	Not applicable	WT	Yes	18	M	W	0	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	1	616	841.00	Partial	No	Yes	No	Unknown	4900	22	0.5	5.0	0.7
CRC 129	KRAS p.G12V	211	KRAS p.G12V	No	73	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	1	468	1012.00	Mixed	Yes	Yes	No	Unknown	5620	20	1.2	4.1	0.8
CRC 130	KRAS p.G12V	2,220	KRAS p.G12V	No	45	M	B	0	1	0	0	0	0	0	1	ll to moderat	No	No	Yes	Unknown	WT	0	738	790.00	No Therapy	No	No	No	3	3180	23	0.7	4.4	0.9
CRC 131	None Detected	Not applicable	KRAS p.G13D	No	44	M	W	0	1	0	0	0	0	0	1	Moderately	Yes	Distant	No	MSS	WT	0	427	537.00	Partial	No	Yes	No	Unknown	4600	33	0.4	4.9	1.0
CRC 132	None Detected	Not applicable	WT	No	54	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	WT	0	477	575.00	No Therapy	No	No	No	Unknown	8840	23	0.5	3.6	0.6
CRC 133	None Detected	Not applicable	WT	No	69	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	1	32	612.00	Partial	Yes	Yes	No	96	8870	17	0.4	3.1	1.5
CRC 134	KRAS p.G12V	1,280	KRAS p.G12V	No	56	F	W	0	1	0	1	0	0	0	2	Moderately	No	No	Yes	Unknown	Unknown	1	42	1060.00	Unknown	Yes	Yes	No	7	8680	43	0.5	4.6	1.0
CRC 135	None Detected	Not applicable	WT	Yes	56	M	W	0	1	0	0	0	0	0	1	Poorly	Yes	Yes	No	Unknown	Unknown	1	74	270.00	Partial	Yes	Yes	No	10	5480	31	0.5	3.4	0.8
CRC 136	None Detected	Not applicable	WT	No	44	F	W	0	1	0	0	0	0	0	1	Moderately	No	Unknown	Yes	Unknown	Unknown	0	548	853.00	Partial	Yes	Yes	No	146	10270	610	0.6	3.3	0.5
CRC 137	None Detected	Not applicable	WT	No	50	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	1	173	703.00	Unknown	Yes	Yes	Yes	6	5850	22	0.4	4.3	0.6
CRC 139	KRAS p.G12D	267	KRAS p.G12D	No	69	M	W	1	1	0	0	0	0	0	1	Moderately	Yes	No	Yes	Unknown	Unknown	1	45	419.00	Progression	Yes	Yes	No	27	4580	33	0.4	4.3	0.9
CRC 140	None Detected	Not applicable	WT	No	51	F	B	0	1	0	0	0	0	0	1	Unknown	Unknown	No	Yes	Unknown	WT	0	1606	1877.00	Partial	Yes	Yes	No	55	1692	861	0.9	3.1	0.6
CRC 141	KRAS p.G12A	9,040	KRAS p.G12A	No	64	F	W	0	1	0	0	0	0	0	1	Moderately	No	Yes	Yes	Unknown	Unknown	1	152	704.00	Stable	Yes	Yes	No	Unknown	7090	19	0.4	4.6	0.9
CRC 142	None Detected	Not applicable	WT	Yes	53	M	W	0	1	0	0	0	0	0	1	Moderately	No	Yes	Yes	Unknown	Unknown	1	140	822.00	Partial	Yes	Yes	No	118	6910	41	1.0	4.2	1.0
CRC 143	None Detected	Not applicable	WT	No	63	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	1	2130	2543.00	Stable	Yes	Yes	No	5	5820	31	0.6	4.3	1.1
CRC 144	None Detected	Not applicable	WT	Yes	43	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	Unknown	1	30	184.00	Partial	Yes	Yes	No	30	4510	31	0.6	4.9	0.7

Table 5-S3. Continued.

CRC 145	None Detected	Not applicable	WT	No	58	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	0	1636	38810.00	No Therapy	Yes	No	No	Unknown	6550	51	0.2	4.1	0.7
CRC 146	None Detected	Not applicable	WT	Yes	50	M	W	0	1	0	0	0	0	0	1	Moderately	Unknown	No	Yes	Unknown	Unknown	1	489	711.00	Partial	Yes	Yes	No	4	5800	39	0.6	4.6	0.9
CRC 147	None Detected	Not applicable	WT	No	63	F	W	0	1	0	0	0	0	0	1	lerately to po	Unknown	No	No	MSS	Unknown	0	2083	2720.00	Partial	Yes	Yes	No	3	10880	922	0.9	2.6	1.0
CRC 148	None Detected	Not applicable	WT	Yes	36	F	W	0	1	0	0	0	0	0	1	Moderately	Yes	No	Yes	Unknown	Unknown	1	15	400.00	Partial	No	Yes	No	5	19290	504	2.3	2.2	0.8
CRC 150	KRAS p.G13D	539	KRAS p.G13D	No	41	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	0	1756	2242.00	Partial	Yes	Yes	Yes	3	11090	617	2.0	3.3	0.7
CRC 151	None Detected	Not applicable	WT	No	59	M	W	Unknown	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	1	207	988.00	Complete	Yes	Yes	No	12	13420	428	2.1	3.5	0.9
CRC 152	KRAS p.G12S	32,900	KRAS p.G12S	No	59	F	W	1	1	1	0	0	0	0	2	Well	Unknown	No	Yes	Unknown	Unknown	1	0	1453.00	Progressed	Yes	Yes	No	Unknown	11810	8	0.3	2.6	0.4
CRC 153	None Detected	Not applicable	WT	No	63	M	W	3	0	0	1	0	0	0	1	Poorly	Yes	No	Yes	Unknown	Unknown	1	482	539.00	No Therapy	Yes	No	Yes	3	3980	11	0.3	3.9	1.9
CRC 154	None Detected	Not applicable	WT	Yes	64	F	W	1	1	1	1	0	0	0	3	Unknown	Unknown	Distant	No	Unknown	Unknown	1	67	765.00	Progressed	Yes	Yes	Yes	7	5700	25	0.1	4.3	0.9
CRC 155	None Detected	Not applicable	WT	No	48	M	W	Unknown	0	1	1	0	0	0	2	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	1	217	1236.00	Stable	Yes	Yes	Yes	546	6250	14	0.2	3.7	0.9
CRC 156	KRAS p.G12D	2,560	KRAS p.G12D	No	38	M	W	Unknown	1	1	0	0	0	0	2	Moderately	Unknown	No	Yes	Unknown	Unknown	1	987	1085.00	No Therapy	Yes	No	No	54	5530	48	0.5	4.3	0.8
CRC 157	None Detected	Not applicable	WT	No	46	F	B	0	1	1	0	0	0	0	2	Unknown	Unknown	No	Yes	Unknown	Unknown	0	113	2069.00	Stable	Yes	Yes	No	1069	9300	24	0.5	4.1	0.8
CRC 159	KRAS p.G12D	858	KRAS p.G12D	No	86	F	W	Unknown	0	1	1	0	1	0	3	Unknown	Unknown	Distant	Yes	Unknown	Unknown	1	64	108.00	No Therapy	No	No	No	8	6920	12	0.5	4.1	0.9
CRC 160	KRAS p.G12D	17,900	KRAS p.G12D	No	59	F	W	1	1	0	0	0	0	0	1	lerately to po	Yes	Distant	Yes	Unknown	Unknown	1	252	318.00	No Therapy	Yes	No	No	19	5950	15	0.4	4.2	1.3
CRC 161	None Detected	Not applicable	KRAS p.G13D	No	58	F	W	0	0	0	1	0	0	0	1	Well	Yes	Distant	Yes	Unknown	Unknown	0	0	104.00	No Therapy	Yes	No	No	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
CRC 162	None Detected	Not applicable	KRAS p.G13D	No	32	F	A	0	0	0	1	0	0	0	1	Moderately	Yes	No	No	Unknown	Unknown	0	759	998.00	No Therapy	Yes	No	No	1	4170	34	0.7	4.5	0.7
CRC 164	None Detected	Not applicable	WT	No	64	F	W	0	0	0	1	0	0	0	1	Poorly	Yes	Distant	Yes	Unknown	WT	0	0	39.00	No Therapy	Yes	No	No	2	5480	38	0.5	4.1	0.6
CRC 165	None Detected	Not applicable	KRAS p.G12V	No	63	M	W	0	0	1	0	0	0	1	2	Moderately	Yes	Distant	No	Unknown	Unknown	1	377	3508.00	No Therapy	Yes	Yes	Yes	Unknown	11070	17	0.7	4.6	0.8
CRC 166	None Detected	Not applicable	WT	No	76	M	W	4	1	0	1	0	0	0	2	Unknown	Unknown	Yes	No	Unknown	Unknown	1	14	Unknown	No Therapy	No	No	No	Unknown	19270	42	3.5	1.8	2.9
CRC 167	None Detected	Not applicable	WT	No	66	M	W	0	1	0	0	0	0	0	1	Moderately	Yes	Distant	No	Unknown	Unknown	1	897	1607.00	Mixed	Yes	No	Yes	Unknown	7980	15	0.4	4.7	0.8
CRC 168	None Detected	Not applicable	WT	No	54	M	W	0	1	0	0	0	0	0	1	Moderately	No	Unknown	Unknown	Unknown	Unknown	0	424	Unknown	No Therapy	Yes	No	No	1	4790	25	0.8	4.6	1.1
CRC 169	None Detected	Not applicable	WT	No	48	M	B	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	1	288	2302.00	Stable	Yes	Yes	Unknown	Unknown	10540	664	1.1	3.3	0.7
CRC 171	KRAS p.G12D	14,600	KRAS p.G12D	Yes	63	M	W	0	1	0	0	0	0	0	1	Moderately	No	Unknown	Unknown	Unknown	Unknown	1	207	1347.00	Unknown	Yes	Yes	No	92	5720	41	0.9	4.2	0.7
CRC 172	KRAS p.G12V	1,230	KRAS p.G12V	No	63	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	0	828	1132.00	Partial	Yes	Yes	No	Unknown	9580	22	0.4	4.4	0.7
CRC 173	None Detected	Not applicable	KRAS p.G12D	Yes	59	M	W	0	1	0	0	0	0	0	1	Moderately	Yes	No	Yes	Unknown	Unknown	0	1106	1238.00	Stable	Yes	Yes	No	3	3480	18	0.5	4.6	0.8
CRC 175	None Detected	Not applicable	WT	Yes	66	M	W	0	1	0	0	0	0	0	1	Poorly	Yes	Yes	Yes	Unknown	Unknown	0	1022	1155.00	Partial	Yes	Yes	No	Unknown	3470	13	Unknown	Unknown	Unknown
CRC 176	None Detected	Not applicable	WT	Yes	56	M	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	1	55	1404.00	Partial	Yes	Yes	No	Unknown	5820	27	0.7	4.7	0.9
CRC 177	None Detected	Not applicable	WT	Yes	37	F	W	0	1	0	0	0	0	0	1	Moderately	No	No	Yes	Unknown	Unknown	1	318	487.00	Progression	Yes	Yes	No	Unknown	8270	44	0.5	4.8	0.7
CRC 178	None Detected	Not applicable	WT	No	51	F	W	0	1	0	0	0	0	0	1	Unknown	Unknown	No	No	Unknown	Unknown	1	10	209.00	Partial	Yes	Yes	No	Unknown	4740	39	0.7	4.2	0.7
CRC 179	None Detected	Not applicable	WT	No	45	F	W	0	1	0	0	0	0	0	1	Poorly	No	No	Yes	Unknown	Unknown	0	544	664.00	No Therapy	Yes	No	No	5	5320	23	0.4	4.5	1.0
CRC 180	None Detected	Not applicable	WT	No	53	F	W	0	1	0	0	0	0	0	1	ll to moderat	No	No	No	Unknown	Unknown	0	503	577.00	No Therapy	No	No	No	Unknown	14140	514	1.6	2.7	0.7
CRC 181	KRAS p.G12S	7,170	KRAS p.G12S	No	56	M	B	0	1	0	0	0	0	0	1	Moderately	Unknown	No	No	Unknown	Unknown	0	781	815.00	No Therapy	No	No	No	40	12440	12	0.5	4.5	1.2
CRC 182	None Detected	Not applicable	WT	No	81	F	W	0	0	0	0	0	0	0	0	Unknown	Unknown	Distant	Unknown	Unknown	Unknown	1	28	619.00	Progression	Yes	Yes	Yes	Unknown	1700.0	10.0	Unknown	Unknown	Unknown
CRC 183	None Detected	Not applicable	WT	No	55	M	W	1	1	1	0	0	0	0	2	ll to moderat	No	Distant	No	Unknown	WT	0	319	2704.00	Complete	Yes	Yes	Yes	21	4770	33	0.4	4.9	0.9
CRC 184	KRAS p.G12D	1,350	KRAS p.G12D	No	44	F	O	0	1	0	0	0	0	0	1	Moderately	No	No	No	MSS	WT	0	32	Unknown	No Therapy	No	No	No	380	7180	45	0.2	4.8	0.8
CRC 185	None Detected	Not applicable	WT	No	32	F	A	0	1	0	0	0	0	0	1	ll to moderat	No	No	No	MSS	WT	0	242	Unknown	No Therapy	No	Yes	No	1	2250	51	0.8	4.5	0.7
CRC 186	None Detected	Not applicable	WT	No	56	M	W	0	1	0	0	0	0	0	1	ll to moderat	No	Distant	No	Unknown	WT	0	233	Unknown	No Therapy	No	No	No	Unknown	2830	27	0.3	4.5	0.8
CRC 187	None Detected	Not applicable	WT	No	59	F	A	0	1	0	0	0	0	0	1	ll to moderat	Yes	No	Yes	MSS	utation four	0	200	273.00	No Therapy	No	No	No	Unknown	5770	16	0.2	4.1	0.7
CRC 192	KRAS p.G13D	9,970	KRAS p.G13D																															
LCR 001	KRAS p.G12S	730	KRAS p.G12S	No	74	M	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	672	791.00	Not Applicable	Yes	No	No	1	7600	21	0.8	3.4	0.8
LCR 002	None Detected	Not applicable	WT	No	66	F	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	602	4011.00	SD	Yes	Yes	No	8	6200	15	1.2	3.7	0.5
LCR 004	None Detected	Not applicable	WT	No	41	F	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	MSS	WT	0	533	556.00	SD	No	No	Yes	22	4300	13	0.6	3.8	0.4
LCR 006	KRAS p.G13D	500	KRAS p.G13D	No	76	M	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	507	707.00	Not Applicable	Yes	No	No	5	7400	34	0.8	3.8	0.8
LCR 010	KRAS p.G12V	500	KRAS p.G12V	No	76	M	Unknown	Unknown	1	0	0	0	0	0	1	Moderately	No	Unknown	Unknown	Unknown	WT	0	294	1086.00	Not Applicable	Yes	Unknown	No	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
LCR 011	None Detected	Not applicable	WT	No	65	F	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	355	553.00	SD	Yes	No	No	9	10100	28	0.5	3.0	0.9
LCR 013	KRAS p.G12V	500	KRAS p.G12V	No	68	M	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	301	1312.00	Not Applicable	Yes	Yes	Yes	1	6000	32	1.0	4.0	0.8
LCR 014	None Detected	Not applicable	KRAS p.G12D	No	47	M	Unknown	0	1	0	0	0	0	0	1	Moderately	Yes	Yes	Yes	MSI	WT	0	385	1135.00	Not Applicable	Yes	Yes	No	2	6700	14	1.0	4.2	1.0
LCR 015	KRAS p.G12C	1,210	KRAS p.G12C	No	66	M	Unknown	0	1	0	0	0	0	0	1	Moderately	No	Distant	No	Unknown	WT	0	380	584.00	Not Applicable	Yes	No	No	15	9500	41	0.8	3.8	0.8
LCR 016	KRAS p.G12D	63	KRAS p.G12D	No	70	F	Unknown	0	1	0	0	0	0	0	1	Well	No	Unknown	Unknown	Unknown	WT	0	129	1268.00	SD	Yes	Unknown	No	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
LCR 017	KRAS p.G12A	4,330	KRAS p.G12A	No	45	F	Unknown	0	1	0	0	0	0	0	1	Moderately	Unknown	No	No	Unknown	WT	0	371	934.00	Not Applicable	Yes	Yes	No	2	5400	20	0.8	4.2	0.6
LCR 019	KRAS p.G12V	4,060	KRAS p.G12V	No	78	M	Unknown	0	1	0	0	0	0	0	1	Moderately	No	Distant	No	MSS	WT	0	336	1629.00	Not Applicable	Yes	No	No	3	7100	12	0.6	4.0	0.8
LCR 071	None Detected	Not applicable	WT	no	73	M		0	0	1	0	0	0	0	1	Moderately	No	No	No	Stable	WT	0			Progression	Yes	Yes	No	98	7.2	11.0	8.0	39.0	65.0
LCR 072	KRAS p.G12D	11,600	KRAS p.G12D	no	39	F		0	1	1	0	0	0	0	2	Moderately	No	No	No	Unknown	WT	1	344		Stable	No	No	No	200	12.5	23.0	6.0	33.0	62.0
LCR 073	KRAS p.G13D	500	KRAS p.G13D	No	68	M	Unknown	1	0	1	0	0	0	0	1	Poorly	Yes	Distant	No	MSI	WT	1	370	542.00	SD	Yes	No	No	2	5800	28	1.1	3.6	0.7
LCR 076	KRAS p.G12D	1,260	KRAS p.G12D	no	71	M		0	1	0	1	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	Uknown		Progression	No	No	No	580	8.3	8.0	7.0	37.0	42.0
LCR 077	None Detected	Not applicable	WT	No	63	F	Unknown	0	0	1	1	0	0	0	2	Poorly	Yes	Distant	Yes	Unknown	WT	0	140	1183.00	SD	Yes	Yes	No	64	8600	20	0.6	3.7	0.7
LCR 078	KRAS p.G13D	500	KRAS p.G13D	No	49	M	Unknown	0	0	1	0	0	0	0	1	Moderately	Yes	No	No	MSS	WT	0	378	2166.00	SD	Yes	Yes	No	1	8200	33	0.9	3.9	0.8
LCR 079	KRAS p.G12D	11,680	KRAS p.G12D	No	57	M	Unknown	0	1	0	0	0	0	0	1	Moderately	No	Distant	No	MSS	WT	1	211	255.00	SD	No	No	No	2750	7700	18	1.4	3.1	1.0
LCR 080	None Detected	Not applicable	WT	No	64	M	Unknown	0	1	0	1	0	0	0	2	Poorly	No	Distant	No	MSS	utation four	1	56	139.00	ed prior to restagi	Yes	No	No	235	13400	143	1.4	2.7	0.7
LCR 081	None Detected	Not applicable	WT	No	55	F	Unknown	0	0	1	1	0	0	0	2	Moderately	No	No	No	Unknown	utation four	0	335	342.00	SD	No	No	No	18	6800	9	1.0	2.8	0.4
LCR 082	None Detected	Not applicable	WT	No	60	M	Unknown	0	1	0	1	0	0	0	2	Moderately	No	No	No	Unknown	utation four	0	262	800.00	SD	Yes	Yes	No	7	4300	40	0.6	3.8	0.7
LCR 083	KRAS p.G12S	500	KRAS p.G12S	No	55	M	Unknown	0	0	1	0	0	0	0	1	Moderately	No	Distant	No	Unknown	WT	0	321	698.00	SD	Yes	Yes	No	31	8800	21	1.0	2.9	0.7
LCR 084	KRAS p.G12D	139,000	KRAS p.G12D	Yes	58	M	Unknown	1	1	1	1	0	0	0	3	Moderately	No	No	Yes	MSS	WT	1	136	185.00	SD	No	Yes	No	26	Unknown	Unknown	Unknown	Unknown	Unknown
LCR 108	None Detected	Not applicable	WT	No	64	F	Unknown	0	0	1	1	0	1	0	3	Moderately	No	Distant	No	MSS	WT	0	234	425.00	Partial	Yes	No	No	55	4500	15	1.3	3.3	0.6
LCR 109	None Detected	Not applicable	WT	No	71	M	Unknown	0	1	1	0	0	0	0	2	Moderately	No	Distant	No	MSS	WT	1	46	77.00	ed prior to restagi	Yes	No	No	17	8900	16	1.7	3.2	1.2
LCR 110	KRAS p.G12D	628	KRAS p.G12D	No	54	F	Unknown	0	1	0	0	0	0	0	1	Moderately	No	Yes	No	MSS	WT	0	208	273.00	Partial	Yes	No	No	90	5200	27	0.4	4.2	0.6
LCR 113	None Detected	Not applicable	WT	No	57	M	Unknown	1	1	0	1	0	0	0	2	Poorly	No	Yes	No	MSI	utation four	0	194	254.00	Partial	Yes	No	No	2	11200	15	0.5	3.7	0.6
LCR 116	KRAS p.G12C	768	KRAS p.G12C	No	75	M	Unknown	0	1	0	1	0	0	0	2	Moderately	No	No	No	Unknown	WT	0	189	202.00	Partial	No	No	No	10	6900	13	1.2	3.8	0.9

Table 5-S3. Continued.

LCR 117	None Detected	Not applicable	WT	No	71	M	Unknown	1	1	1	1	0	0	0	3	Moderately	No	Unknown	Unknown	Unknown	WT	0	189	257.00	SD	Yes	No	No	963	Unknown	Unknown	Unknown	Unknown	Unknown
LCR 118	None Detected	Not applicable	WT	No	68	M	Unknown	0	1	1	1	0	0	0	3	Moderately	No	No	No	Unknown	WT	0	155	832.00	Partial	Yes	Yes	No	2	Unknown	Unknown	Unknown	Unknown	Unknown
LCR 119	None Detected	Not applicable	WT	No	64	M	Unknown	1	1	0	1	0	0	0	2	Moderately	No	Unknown	Unknown	Unknown	WT	0	189	306.00	SD	No	No	No	9000	Unknown	Unknown	Unknown	Unknown	Unknown
LCR 121	KRAS p.G13D	500	KRAS p.G13D	No	31	F	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	MSS	WT	0	305	356.00	Partial	Yes	No	No	7	11800	18	0.5	3.6	0.5
LCR 122	KRAS p.G12D	500	KRAS p.G12D	No	82	M	Unknown	0	1	0	0	0	0	0	1	Moderately	No	No	No	Unknown	WT	0	271	777.00	SD	Yes	No	No	28	6800	14	0.8	3.8	0.5

Table 5-S4. Clinical characteristics of patients with discordant tissue and plasma KRAS mutation data.

Univariable Analysis

	False Negative N = 10	True Negative N = 127	P-Value
Age			
Mean (SD)	55.2 (11.08)	59.02 (12.83)	
Median (Range)	58.5 (32, 69)	60 (18, 85)	0.293
No. Missing (%)	0 (0)	2 (2)	
Gender, N (%)			
Female	3 (30)	51 (41)	0.737
Male	7 (70)	74 (59)	
Missing	0	2	
Race, N (%)			
A	1 (11)	3 (3)	0.447
B	0 (0)	10 (9)	
M	0 (0)	1 (1)	
O	0 (0)	0 (0)	
W	8 (89)	97 (87)	
Missing	1	16	
ECOG, N (%)			
0	10 (100)	96 (80)	0.653
1	0 (0)	15 (12)	
2	0 (0)	6 (5)	
3	0 (0)	2 (2)	
4	0 (0)	1 (1)	
Missing	0	7	
Met Site - Liver			
No	5 (50)	25 (20)	0.072
Yes	5 (50)	100 (80)	
Missing	0	2	
Met Site - Lungs			
No	7 (70)	88 (70)	>0.99
Yes	3 (30)	37 (30)	
Missing	0	2	
Met Site - Peritoneum			
No	8 (80)	89 (71)	0.818
Yes	2 (20)	36 (29)	
Missing	0	2	
Met Site - Pelvis			
No	10 (100)	121 (97)	>0.99
Yes	0 (0)	4 (3)	
Missing	0	2	
Met Site - Bone			
No	9 (90)	122 (98)	0.693
Yes	1 (10)	3 (2)	
Missing	0	2	
Met Site - Brain			
No	9 (90)	125 (100)	0.103
Yes	1 (10)	0 (0)	
Missing	0	2	
# of Met Sites			
0	0 (0)	3 (2)	0.604
1	8 (80)	77 (62)	
2	2 (20)	31 (25)	
3	0 (0)	13 (10)	
4	0 (0)	0 (0)	
Missing	0	3	
Histological Subtype			
Moderately	7 (70)	80 (73)	0.023
Moderately to Poorly	0 (0)	1 (1)	
Poorly	1 (10)	17 (15)	
Well	1 (10)	0 (0)	
Well to Moderately	1 (10)	12 (11)	
Missing	0	17	

Table 5-S4. Continued.

Mucinous			
No	3 (30)	87 (85)	< 0.001*
Yes	7 (70)	15 (15)	
Missing	0	25	
Smoking Status			
Distant	3 (30)	19 (16)	0.446
No	6 (60)	93 (78)	
Yes	1 (10)	8 (7)	
Missing	0	7	
Family History			
No	6 (60)	83 (70)	0.776
Yes	4 (40)	36 (30)	
Missing	0	8	
MSI Status			
MSI	1 (50)	1 (8)	0.254
MSS	1 (50)	11 (85)	
Stable	0 (0)	1 (8)	
Missing	8	114	
BRAF Status			
Mutation Found	0 (0)	9 (12)	0.95
WT	5 (100)	69 (88)	
Missing	5	49	
	1	6	
Prior Surgery			
No	2 (20)	35 (28)	0.859
Yes	8 (80)	90 (72)	
Missing	0	2	
Prior Chemotherapy			
No	4 (40)	44 (35)	>0.99
Yes	6 (60)	81 (65)	
Missing	0	2	
Prior Radiation			
No	9 (90)	105 (85)	>0.99
Yes	1 (10)	19 (15)	
Missing	0	3	
CEA (+/- 8 weeks)			
Mean (SD)	2.14 (0.9)	380.82 (1663.89)	
Median (Range)	2 (1, 3)	17 (1, 13864)	< 0.001*
No. Missing (%)	3 (30)	24 (19)	
WBC			
Mean (SD)	5645.56 (2367.93)	7710.57 (3831.43)	
Median (Range)	4600 (3480, 11070)	7300 (7.2, 20350)	0.056
No. Missing (%)	1 (10)	5 (4)	
ALT			
Mean (SD)	21.67 (9.06)	77.56 (165.37)	
Median (Range)	18 (9, 34)	24 (7, 922)	0.299
No. Missing (%)	1 (10)	10 (8)	
Bilirubin			
Mean (SD)	0.66 (0.18)	0.85 (0.91)	
Median (Range)	0.7 (0.4, 1)	0.6 (0.1, 8)	>0.99
No. Missing (%)	2 (20)	34 (27)	
Albumin			
Mean (SD)	4.54 (0.25)	4.22 (3.34)	
Median (Range)	4.6 (4.1, 4.9)	4 (1.8, 39)	0.002
No. Missing (%)	1 (10)	12 (9)	
Creatinine			
Mean (SD)	0.88 (0.15)	1.41 (5.92)	
Median (Range)	0.8 (0.7, 1.1)	0.8 (0.4, 65)	0.308
No. Missing (%)	1 (10)	9 (7)	

*When a Bonferroni correction was applied, mucinous pathology and CEA remained significant (P<0.0017)

Table 5-S5. Clinical characteristics of patients with false negatives in plasma *KRAS* mutation compared with tissue samples.

	Multivariable Analysis
	Odds Ratio Estimate
Age	0.989
CEA (within 8 weeks of blood DNA measurement)	0.897
WBC	0.999
ALT	0.959

Table 5-S6. Association between clinical characteristics and ctDNA concentration (log scale) in metastatic colorectal cancer patients.

	Univariable Analysis	
	Estimate (95% CI)	**P-Value**
Age	-0.035 (-0.07, 0)	0.051
Gender		
Female	ref	
Male	-0.625 (-1.672, 0.421)	0.237
Race		
White	ref	
Non-White	-0.456 (-2.813, 1.9)	0.699
ECOG PS		
0	ref	
1, 2	1.025 (-0.208, 2.259)	0.102
≥3	3.31 (0.306, 6.314)	0.031
Met Site - Liver	1.433 (0.113, 2.752)	0.034
Met Site - Lungs	0.399 (-0.686, 1.484)	0.465
Met Site - Peritoneum	0.252 (-0.846, 1.351)	0.648
Met Site - Pelvis	-0.857 (-3.383, 1.669)	0.5
Met Site - Bone	0.267 (-1.445, 1.978)	0.757
Met Site - Brain	4.155 (-0.046, 8.357)	0.053
# of Met Sites		
0-1	ref	
≥2	0.872 (-0.158, 1.903)	0.096
Histological Subtype		
Poor	ref	
Moderate	-1.309 (-2.974, 0.356)	0.121
Well	-1.868 (-5.334, 1.597)	0.285
Mucinous		

Table 5-S6. Continued.

No	ref	
Yes	-0.761 (-2.077, 0.554)	0.251
Smoking Status		
Distant	ref	
No	-0.018 (-1.639, 1.603)	0.983
Yes	-0.449 (-3.339, 2.44)	0.757
Family History		
No	ref	
Yes	0.166 (-1.035, 1.368)	0.783
MSI Status		
MSI	ref	
MSS	2.246 (-3.837, 8.329)	0.412
Prior Surgery		
No	ref	
Yes	-1.58 (-2.694, -0.466)	0.006
Prior Chemotherapy		
No	ref	
Yes	0.378 (-0.693, 1.449)	0.484
Prior Radiation		
No	ref	
Yes	-0.492 (-1.9, 0.917)	0.488
Log CEA (within 8 weeks of blood DNA measurement)	0.439 (0.22, 0.658)	<.001
WBC	0 (0, 0)	0.01
Log ALT	-0.07 (-0.566, 0.427)	0.78
Log Bilirubin	0.267 (-0.581, 1.115)	0.531
Log Albumin	-0.211 (-1.479, 1.058)	0.741
Log Creatinine	-0.06 (-0.731, 0.61)	0.857

Table 5-S7. Association between clinical characteristics and ctDNA concentration (log scale) in metastatic colorectal cancer patients.

Multivariable Analysis

	Estimate
Age	-0.013
Prior Surgery	
No	ref
Yes	-0.0799
Log CEA (within 8 weeks of blood DNA measurement)	0.3045
WBC	0.0002

Table 5-S8. Patient characteristics and plasma mutations detected post-EGFR blockade.

Sample	Age	Gender	Tumor	EGFR Blockade	Duration on anti-EGFR therapy (weeks)	Plasma Mutation (nucleotide)	Plasma Mutation (codon)	Mutant fragments per 5 mL
Patient #5	60	M	Colorectal Cancer	Panitumumab	7	KRAS c.35G>A	KRAS p.G12D	66
Patient #16	72	F	Colorectal Cancer	Panitumumab	23	EGFR c.2380C>T	EGFR p.P794S	168
						KRAS c.183C>T	KRAS p.Q61H	90
Patient #17	57	F	Colorectal Cancer	Panitumumab	31	None Detected	None Detected	NA
Patient #18	47	F	Colorectal Cancer	Panitumumab	28	KRAS c.35G>T	KRAS p.G12V	120
						NRAS c.34G>A	NRAS p.G12S	129
Patient #19	42	M	Colorectal Cancer	Panitumumab	15	NRAS c.183A>T	NRAS p.Q61H	40
Patient #21	57	M	Colorectal Cancer	Panitumumab	15	KRAS c.35G>T	KRAS p.G12V	80
						NRAS c.183A>T	NRAS p.Q61H	30
						NRAS c.182A>G	NRAS p.Q61R	82
Patient #22	59	F	Colorectal Cancer	Panitumumab	15	EGFR c.2142G>C	EGFR p.K714N	948
						KRAS c.35G>C	KRAS p.G12A	40
						KRAS c.34G>T	KRAS p.G12C	30
						KRAS c.34G>A	KRAS p.G12S	120
						KRAS c.35G>A	KRAS p.G12D	104
						KRAS c.35G>T	KRAS p.G12V	100
Patient #24	57	M	Colorectal Cancer	Panitumumab	52	KRAS c.35G>C	KRAS p.G12A	15
						NRAS c.181C>A	NRAS p.Q61K	28
Patient #26	59	F	Colorectal Cancer	Panitumumab	23	KRAS c.35G>T	KRAS p.G12V	40
	73	M	Colorectal Cancer	Panitumumab	23	KRAS c.35G>T	KRAS p.G12V	114
Patient #1	64	M	Colorectal Cancer	Panitumumab	32	KRAS c.35G>C	KRAS p.G12A	1590
						KRAS c.34G>T	KRAS p.G12C	2160
						KRAS c.34G>C	KRAS p.G12R	660
						KRAS c.35G>A	KRAS p.G12D	3900
						KRAS c.35G>T	KRAS p.G12V	4940
						KRAS c.183A>T	KRAS p.Q61H	710
						KRAS c.182A>T	KRAS p.Q61L	640
						KRAS c.182A>G	KRAS p.Q61R	688
						NRAS c.181C>A	NRAS p.Q61H	1340
						NRAS c.183A>C	NRAS p.Q61K	4100
						NRAS c.182A>T	NRAS p.Q61L	6760
						NRAS c.182A>G	NRAS p.Q61R	625
Patient #2	53	M	Colorectal Cancer	Panitumumab	22	KRAS c.35G>T	KRAS p.G12V	93
Patient #4	67	F	Colorectal Cancer	Panitumumab	11	KRAS c.34G>C	KRAS p.G12R	30
						KRAS c.183A>T	KRAS p.Q61H	220
Patient #7	55	M	Colorectal Cancer	Panitumumab	23	KRAS c.35G>C	KRAS p.G12A	135
						KRAS c.35G>T	KRAS p.G12V	133
						NRAS c.183A>T	NRAS p.Q61H	848
						NRAS c.181C>A	NRAS p.Q61K	98
						NRAS c.182A>T	NRAS p.Q61L	374
Patient #9	50	M	Colorectal Cancer	Panitumumab	31	KRAS c.35G>A	KRAS p.G12D	61
						NRAS c.181C>A	NRAS p.Q61K	25
Patient #10	67	M	Colorectal Cancer	Panitumumab	23	KRAS c.35G>T	KRAS p.G12V	244
						KRAS c.183A>T	KRAS p.Q61H	83
						KRAS c.183A>C	KRAS p.Q61H	57
						KRAS c.35G>C	KRAS p.G12A	400

Table 5-S8. Continued.

Patient #12	49	M	Colorectal	Cancer	Panitumumab	23	KRAS c.34G>T	KRAS p.G12C	100
							KRAS c.183A>C	KRAS p.Q61H	429
							KRAS c.34G>C	KRAS p.G12R	13
BARD 101 PLS	64	M	Colorectal	Cancer	Cetuximab	72	KRAS c.183A>T	KRAS p.Q61H	394
							NRAS c.182A>G	NRAS p.Q61R	4
							KRAS c.34G>C	KRAS p.G12R	208
BARD 102 PLS	45	F	Colorectal	Cancer	Panitumumab	52	KRAS c.34G>C	KRAS p.G12R	308
							KRAS c.181C>G	KRAS p.Q61E	139
							KRAS c.182_183AA>CC	KRAS p.Q61P	265
BARD 103 PLS	67	F	Colorectal	Cancer	Panitumumab	24	KRAS c.35G>A	KRAS p.G12D	13
							KRAS c.35G>T	KRAS p.G12V	130
CRC 188 PLS	41	F	Colorectal	Cancer	Cetuximab	9	KRAS c.34G>T	KRAS p.G12C	3
CRC 189 PLS	35	F	Colorectal	Cancer	Cetuximab	14	KRAS c.183A>T	KRAS p.Q61H	28
							KRAS c.183A>C	KRAS p.Q61H	13
							BRAF c.1799T>A	BRAF p.V600E	45
							KRAS c.35G>C	KRAS p.G12A	131
CRC 190 PLS	54	M	Colorectal	Cancer	Cetuximab	14	KRAS c.183A>C	KRAS p.Q61H	10
							KRAS c.182A>G	KRAS p.Q61R	11
							NRAS c.182A>T	NRAS p.Q61L	2
							NRAS c.182A>G	NRAS p.Q61R	12
							KRAS c.35G>A	KRAS p.G12D	173
CRC 191 PLS	56	F	Colorectal	Cancer	Panitumumab	7	KRAS c.34G>C	KRAS p.G12R	31
							KRAS c.35G>T	KRAS p.G12V	58
							KRAS c.183A>C	KRAS p.Q61H	250

Figure 5-1. Depiction of circulating tumor DNA.

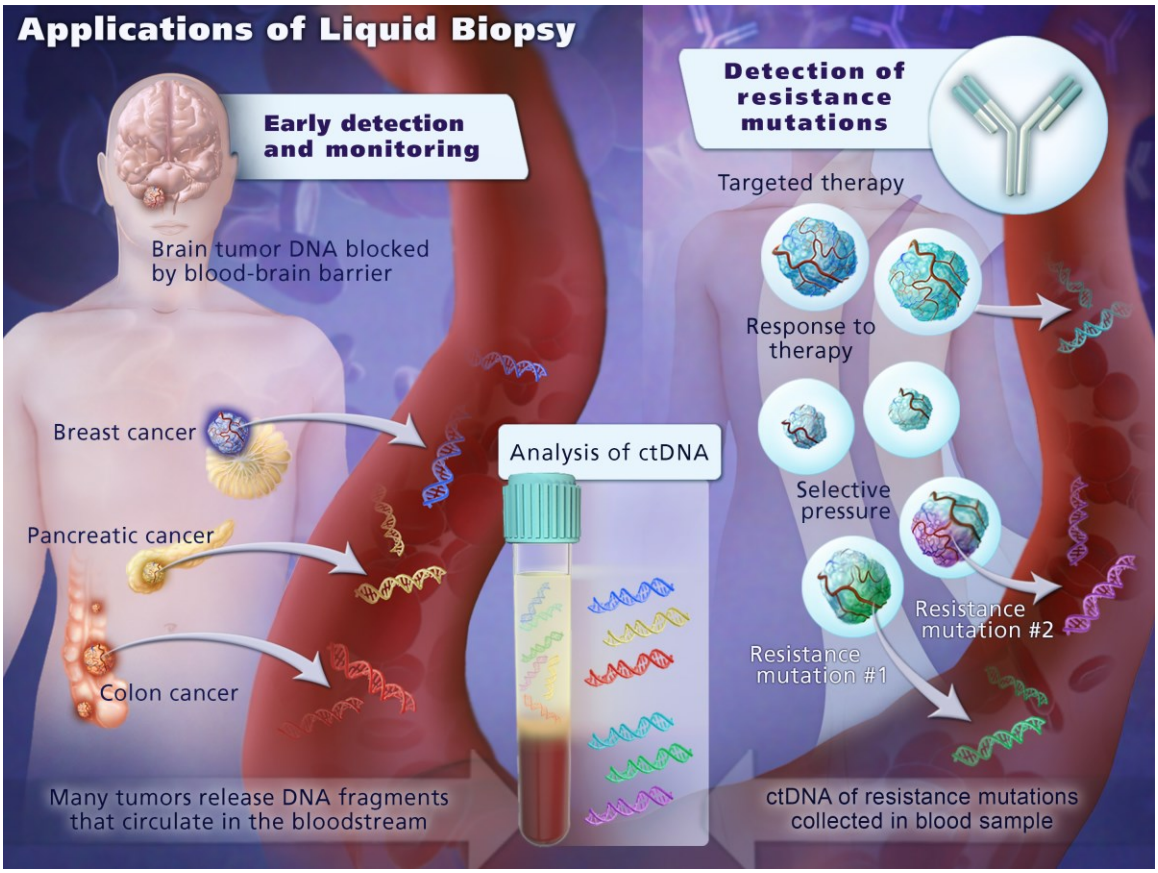
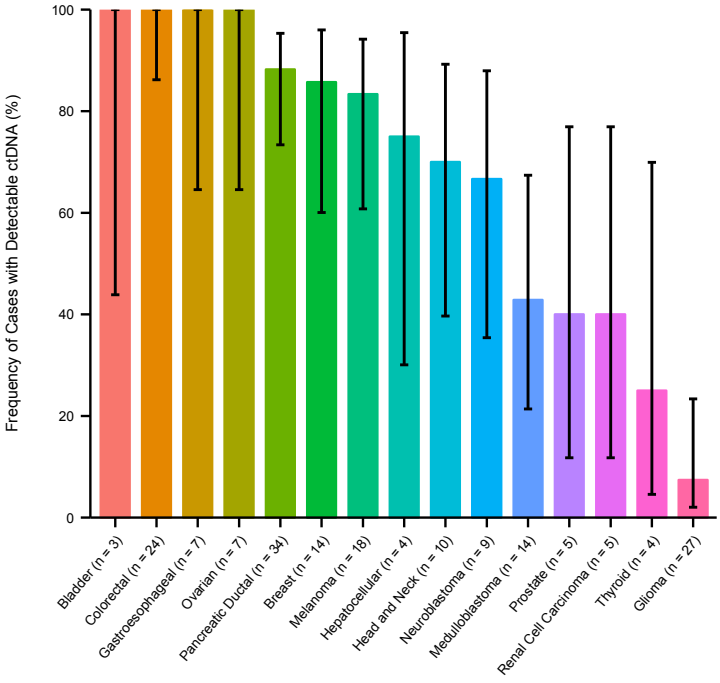


Figure 5-2. Circulating tumor DNA in advanced malignancies.

(A) Fraction of patients with detectable ctDNA and (B) quantification of mutant fragments. Error bars represent the 95% bootstrapped confidence interval of the mean.

Figure 5-2. Continued.

A



B

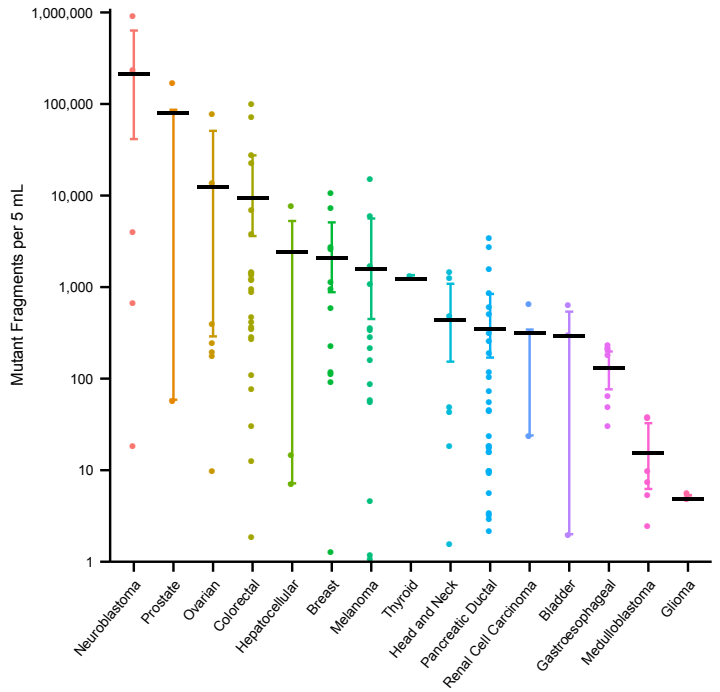


Figure 5-3. Circulating tumor DNA in localized and non-localized malignancies. (A) Fraction of patients with detectable ctDNA in localized (stages I-III) and metastatic (stage IV) colorectal, gastroesophageal, pancreatic and breast cancers; (B) fraction of patients with detectable ctDNA and (C) quantification of mutant fragments in cancer cases categorized by stage. Error bars represent standard error of the mean (SEM).

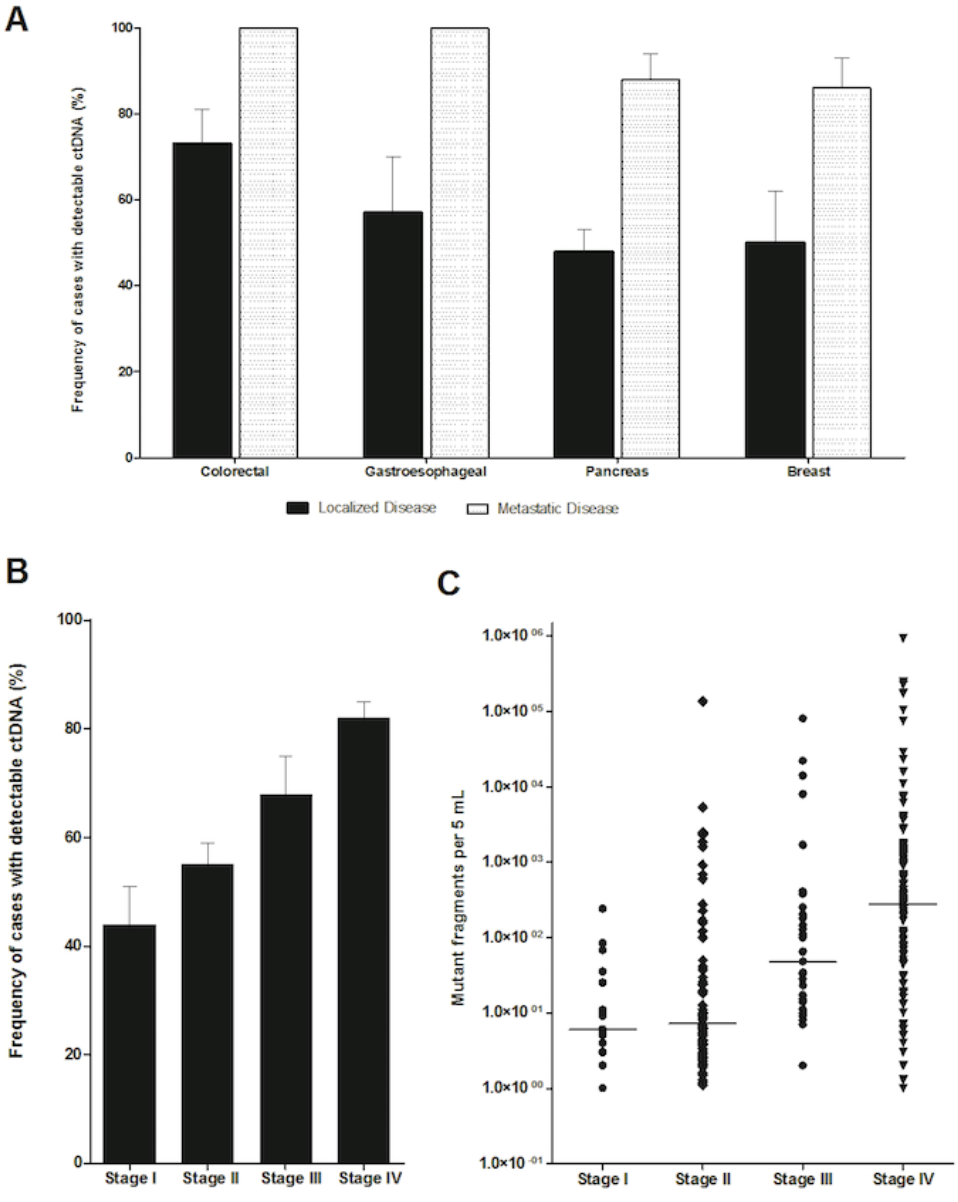


Figure 5-4. Scatter plot correlating point mutations with rearrangements in the same plasma specimens.

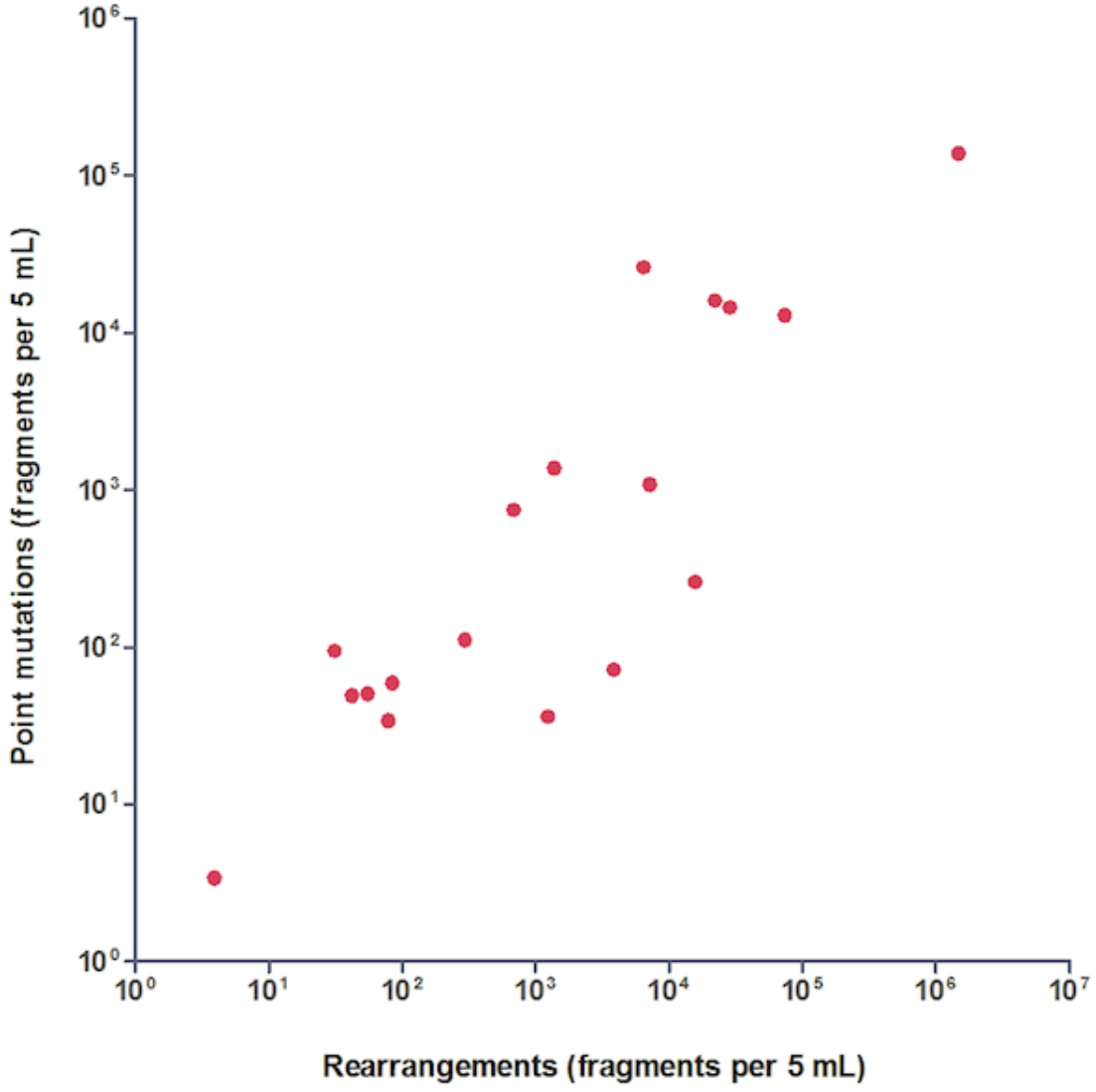


Figure 5-5. The relationship between ctDNA concentration (mutant fragments per mL) and 2-year survival.

The association between survival and ctDNA concentration was assessed holding known prognostic factors (age, ECOG PS, and CEA) constant. The 2-year survival was estimated based on a multivariable Cox regression model, in which ctDNA concentration level was transformed with a natural spline function.

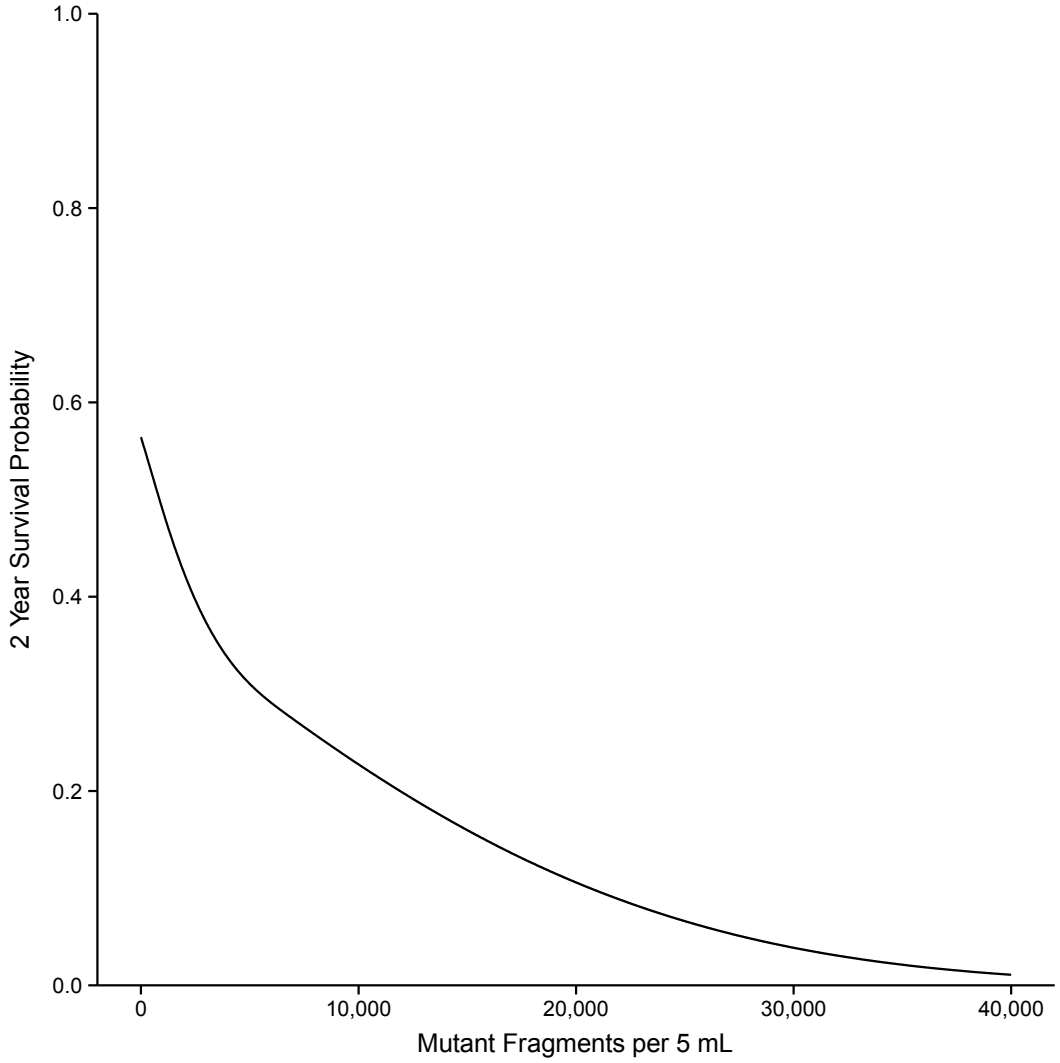


Figure 5-6. Heat map of acquired resistance mutations to EGFR blockade in ctDNA from patients with metastatic colorectal cancer.

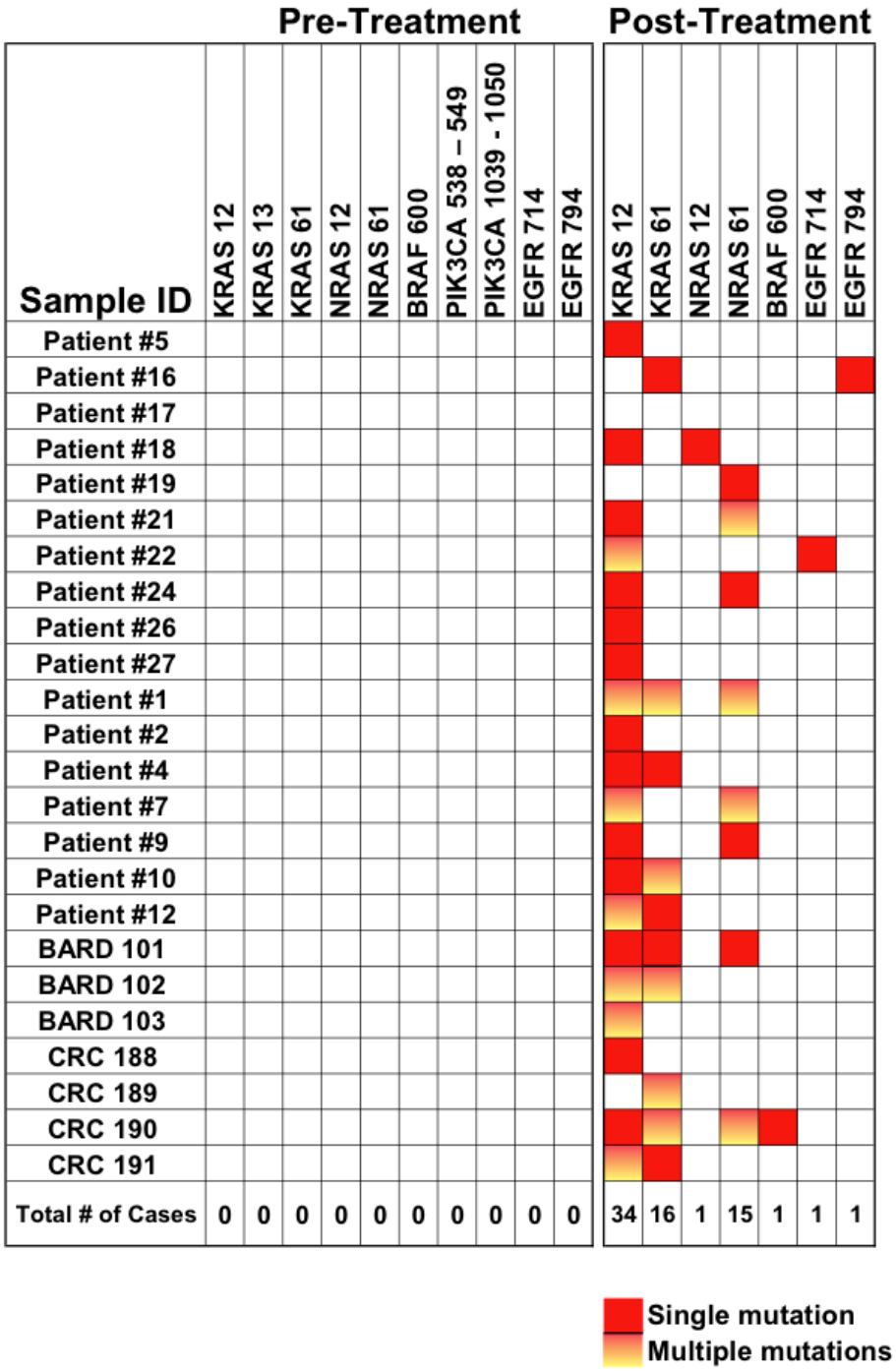


Figure 5-S1. Comparison of methods for analysis of point mutations in plasma DNA. Duplicate aliquots of plasma or serum from 20 different patients were collected. A mutation present in the corresponding tumors was quantified in the first aliquot by Quantification Method #1 (PCR-ligation) and in the second aliquot by Quantification Method #2 (BEAMing in 11 patients and SafeSeqS in 9 patients). Mutant templates per 5 ml plasma or serum are plotted on both the x- and y-axes. In addition to the 20 samples displayed in this graph, each of which yielded at least one mutation with both quantification methods, we tested 10 other duplicate samples containing low amounts of mutant DNA. In five of these cases, neither of the aliquots tested by the two methods revealed any mutations. In five other cases, one method revealed a single mutant template molecule while the other method revealed zero. These results are consistent with expectations based on a Poisson distribution of mutant templates in the circulation.

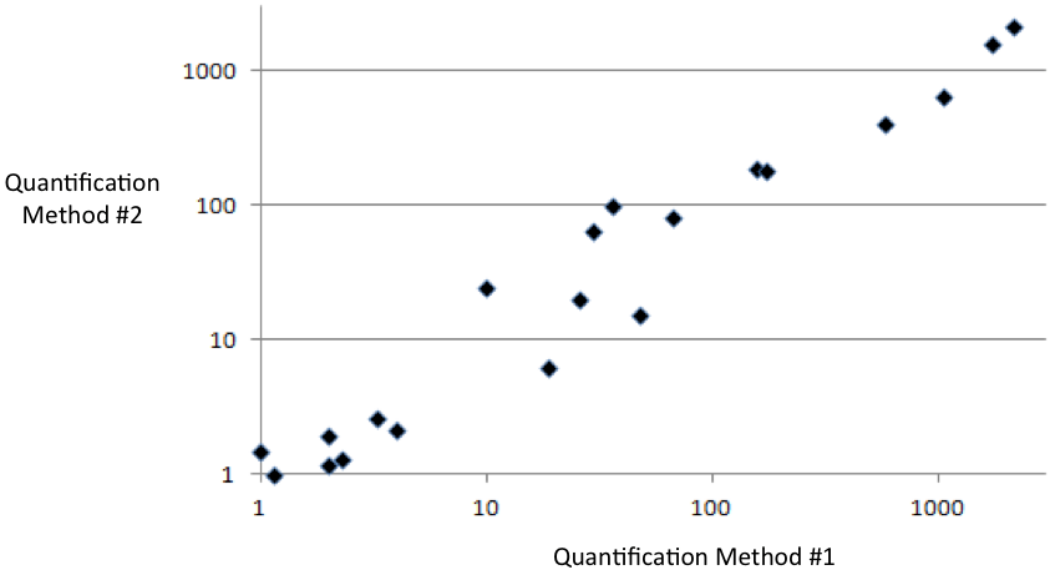


Figure 5-S2. Circulating Tumor DNA in advanced malignancies, ranking of the fraction of patients with detectable ctDNA.

For each tumor type, the rank of the proportion of detectable ctDNA is reported, where a rank of 1 means having the highest proportion and a rank of 15 means having the lowest proportion among the 15 tumor types. Ties were handled by taking the average ranking. The error bars represent the 95% bootstrap confidence interval of the rank of each tumor type.

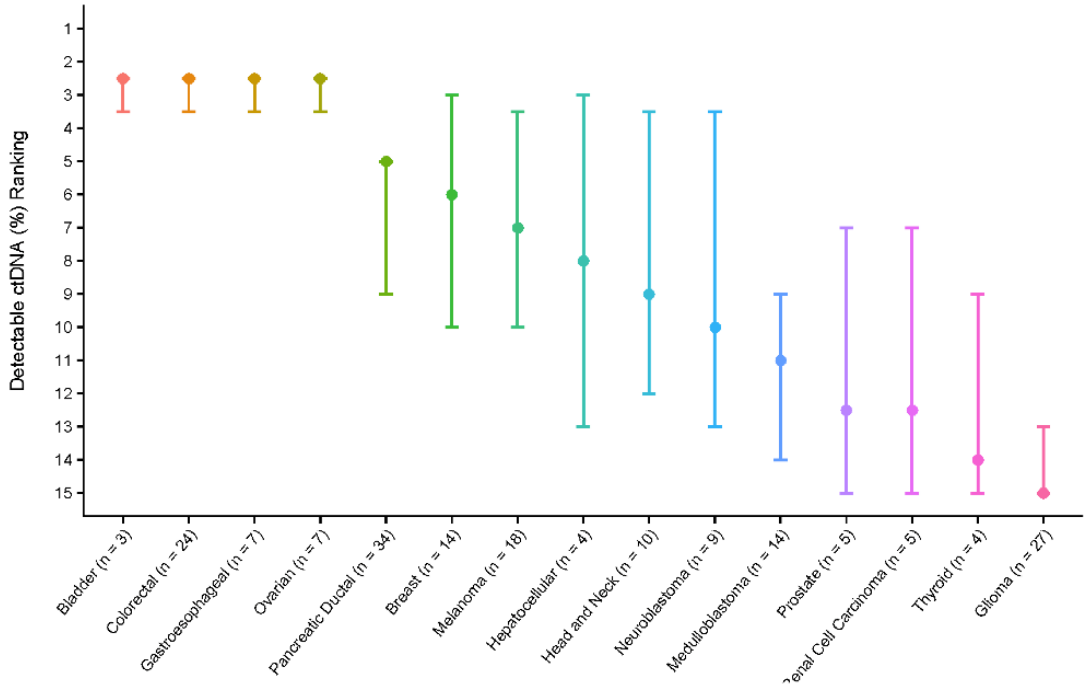
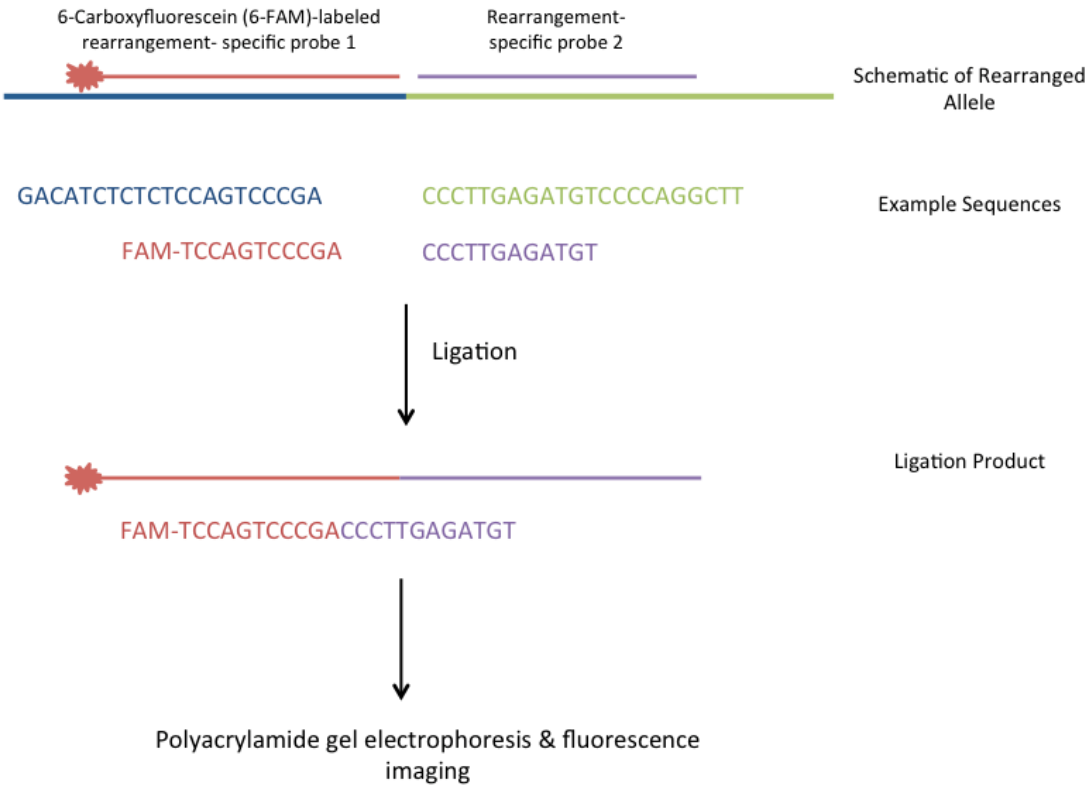


Figure 5-S3. Diagram of the assay used to confirm rearrangements in plasma DNA.



References

1 Luria, S. E. & Delbruck, M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**, 491-511, (1943).

2 Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L. B., Hood, L. & Galas, D. J. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636-639, (2010).

3 Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D., Gibbs, R. A., Hurles, M. E. & McVean, G. A. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073, (2010).

4 Shibata, D. Mutation and epigenetic molecular clocks in cancer. *Carcinogenesis* **32**, 123-128, (2011).

5 McMahon, M. A., Jilek, B. L., Brennan, T. P., Shen, L., Zhou, Y., Wind-Rotolo, M., Xing, S., Bhat, S., Hale, B., Hegarty, R., Chong, C. R., Liu, J. O., Siliciano, R. F. & Thio, C. L. The HBV drug entecavir - effects on HIV-1 replication and resistance. *N Engl J Med* **356**, 2614-2621, (2007).

6 Eastman, P. S., Shapiro, D. E., Coombs, R. W., Frenkel, L. M., McSherry, G. D., Britto, P., Herman, S. A. & Sperling, R. S. Maternal viral genotypic zidovudine resistance and infrequent failure of zidovudine therapy to prevent perinatal transmission of human immunodeficiency virus type 1 in pediatric AIDS Clinical Trials Group Protocol 076. *J Infect Dis* **177**, 557-564, (1998).

7 Sjoblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell,

C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W. & Velculescu, V. E. The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268-274, (2006).

8 Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E. & Vogelstein, B. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-1113, (2007).

9 Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr. & Kinzler, K. W. Cancer genome landscapes. *Science* **339**, 1546-1558, (2013).

10 Diehl, F., Schmidt, K., Durkee, K. H., Moore, K. J., Goodman, S. N., Shuber, A. P., Kinzler, K. W. & Vogelstein, B. Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. *Gastroenterology* **135**, 489-498, (2008).

11 Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 9530-9535, (2011).

12 Hoque, M. O., Lee, J., Begum, S., Yamashita, K., Engles, J. M., Schoenberg, M., Westra, W. H. & Sidransky, D. High-throughput molecular analysis of urine

sediment for the detection of bladder cancer by high-density single-nucleotide polymorphism array. *Cancer Res* **63**, 5723-5726, (2003).

13 Thunnissen, F. B. Sputum examination for early detection of lung cancer. *J Clin Pathol* **56**, 805-810, (2003).

14 Barnes, W. M. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112**, 29-35, (1992).

15 Araten, D. J., Golde, D. W., Zhang, R. H., Thaler, H. T., Gargiulo, L., Notaro, R. & Luzzatto, L. A quantitative measurement of the human somatic mutation rate. *Cancer Res* **65**, 8111-8117, (2005).

16 Campbell, F., Appleton, M. A., Shields, C. J. & Williams, G. T. No difference in stem cell somatic mutation between the background mucosa of right- and left-sided sporadic colorectal carcinomas. *J Pathol* **186**, 31-35, (1998).

17 Tindall, K. R. & Kunkel, T. A. Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. *Biochemistry* **27**, 6008-6013, (1988).

18 Kunkel, T. A. The mutational specificity of DNA polymerase-beta during in vitro DNA synthesis. Production of frameshift, base substitution, and deletion mutations. *J Biol Chem* **260**, 5787-5796, (1985).

19 van Dongen, J. J. & Wolvers-Tettero, I. L. Analysis of immunoglobulin and T cell receptor genes. Part II: Possibilities and limitations in the diagnosis and management of lymphoproliferative diseases and related disorders. *Clin Chim Acta* **198**, 93-174, (1991).

20 Grist, S. A., McCarron, M., Kutlaca, A., Turner, D. R. & Morley, A. A. In vivo human somatic mutation: frequency and spectrum with age. *Mutat Res* **266**, 189-196, (1992).

21 Liu, Q. & Sommer, S. S. Detection of extremely rare alleles by bidirectional pyrophosphorolysis-activated polymerization allele-specific amplification (Bi-PAP-A): measurement of mutation load in mammalian tissues. *Biotechniques* **36**, 156-166, (2004).

22 Monnat, R. J., Jr. & Loeb, L. A. Nucleotide sequence preservation of human mitochondrial DNA. *Proc Natl Acad Sci U S A* **82**, 2895-2899, (1985).

23 Shi, C., Eshleman, S. H., Jones, D., Fukushima, N., Hua, L., Parker, A. R., Yeo, C. J., Hruban, R. H., Goggins, M. G. & Eshleman, J. R. LigAmp for sensitive detection of single-nucleotide differences. *Nat Methods* **1**, 141-147, (2004).

24 Keohavong, P. & Thilly, W. G. Fidelity of DNA polymerases in DNA amplification. *Proc Natl Acad Sci U S A* **86**, 9253-9257, (1989).

25 Sidransky, D., Von Eschenbach, A., Tsai, Y. C., Jones, P., Summerhayes, I., Marshall, F., Paul, M., Green, P., Hamilton, S. R., Frost, P. & et al. Identification of p53 gene mutations in bladder cancers and urine samples. *Science* **252**, 706-709, (1991).

26 Bielas, J. H. & Loeb, L. A. Quantification of random genomic mutations. *Nat Methods* **2**, 285-290, (2005).

27 Vogelstein, B. & Kinzler, K. W. Digital PCR. *Proc Natl Acad Sci U S A* **96**, 9236-9241, (1999).

28 Mitra, R. D., Butty, V. L., Shendure, J., Williams, B. R., Housman, D. E. & Church, G. M. Digital genotyping and haplotyping with polymerase colonies. *Proc Natl Acad Sci U S A* **100**, 5926-5931, (2003).

29 Chetverina, H. V., Samatov, T. R., Ugarov, V. I. & Chetverin, A. B. Molecular colony diagnostics: detection and quantitation of viral nucleic acids by in-gel PCR. *Biotechniques* **33**, 150-152, 154, 156, (2002).

30 Zimmermann, B. G., Grill, S., Holzgreve, W., Zhong, X. Y., Jackson, L. G. & Hahn, S. Digital PCR: a powerful new tool for noninvasive prenatal diagnosis? *Prenat Diagn* **28**, 1087-1093, (2008).

31 Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* **100**, 8817-8822, (2003).

32 Ottesen, E. A., Hong, J. W., Quake, S. R. & Leadbetter, J. R. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **314**, 1464-1467, (2006).

33 Quail, M. A., Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Swerdlow, H. & Turner, D. J. A large genome center's improvements to the Illumina sequencing system. *Nat Methods* **5**, 1005-1010, (2008).

34 Nazarian, R., Shi, H., Wang, Q., Kong, X., Koya, R. C., Lee, H., Chen, Z., Lee, M. K., Attar, N., Sazegar, H., Chodon, T., Nelson, S. F., McArthur, G., Sosman, J. A., Ribas, A. & Lo, R. S. Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973-977, (2010).

35 He, Y., Wu, J., Dressman, D. C., Iacobuzio-Donahue, C., Markowitz, S. D., Velculescu, V. E., Diaz, L. A., Jr., Kinzler, K. W., Vogelstein, B. & Papadopoulos, N. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* **464**, 610-614, (2010).

36 Gore, A., Li, Z., Fung, H. L., Young, J. E., Agarwal, S., Antosiewicz-Bourget, J., Canto, I., Giorgetti, A., Israel, M. A., Kiskinis, E., Lee, J. H., Loh, Y. H., Manos, P. D., Montserrat, N., Panopoulos, A. D., Ruiz, S., Wilbert, M. L., Yu, J., Kirkness, E. F., Izpisua Belmonte, J. C., Rossi, D. J., Thomson, J. A., Eggan, K., Daley, G. Q.,

Goldstein, L. S. & Zhang, K. Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63-67, (2011).

37 Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**, e105, (2008).

38 Erlich, Y., Mitra, P. P., delaBastide, M., McCombie, W. R. & Hannon, G. J. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat Methods* **5**, 679-682, (2008).

39 Rougemont, J., Amzallag, A., Iseli, C., Farinelli, L., Xenarios, I. & Naef, F. Probabilistic base calling of Solexa sequencing data. *BMC Bioinformatics* **9**, 431, (2008).

40 Druley, T. E., Vallania, F. L., Wegner, D. J., Varley, K. E., Knowles, O. L., Bonds, J. A., Robison, S. W., Doniger, S. W., Hamvas, A., Cole, F. S., Fay, J. C. & Mitra, R. D. Quantification of rare allelic variants from pooled genomic DNA. *Nat Methods* **6**, 263-265, (2009).

41 Vallania, F. L., Druley, T. E., Ramos, E., Wang, J., Borecki, I., Province, M. & Mitra, R. D. High-throughput discovery of rare insertions and deletions in large cohorts. *Genome Res* **20**, 1711-1718, (2010).

42 McCloskey, M. L., Stoger, R., Hansen, R. S. & Laird, C. D. Encoding PCR products with batch-stamps and barcodes. *Biochem Genet* **45**, 761-767, (2007).

43 Parameswaran, P., Jalili, R., Tao, L., Shokralla, S., Gharizadeh, B., Ronaghi, M. & Fire, A. Z. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Res* **35**, e130, (2007).

44 Craig, D. W., Pearson, J. V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J. J., Pawlowski, T. L., Laub, T., Nunn, G., Stephan, D. A., Homer, N. & Huentelman, M.

J. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* **5**, 887-893, (2008).

45 Miner, B. E., Stoger, R. J., Burden, A. F., Laird, C. D. & Hansen, R. S. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. *Nucleic Acids Res* **32**, e135, (2004).

46 Herman, D. S., Hovingh, G. K., Iartchouk, O., Rehm, H. L., Kucherlapati, R., Seidman, J. G. & Seidman, C. E. Filter-based hybridization capture of subgenomes enables resequencing and copy-number detection. *Nat Methods* **6**, 507-510, (2009).

47 Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683-692, (2007).

48 de Boer, J. G. & Ripley, L. S. An in vitro assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence. *Genetics* **118**, 181-191, (1988).

49 Eckert, K. A. & Kunkel, T. A. High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Res* **18**, 3739-3744, (1990).

50 Kosuri, S., Eroshenko, N., Leproust, E. M., Super, M., Way, J., Li, J. B. & Church, G. M. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotechnol* **28**, 1295-1299, (2010).

51 Matzas, M., Stahler, P. F., Kefer, N., Siebelt, N., Boisguerin, V., Leonard, J. T., Keller, A., Stahler, C. F., Haberle, P., Gharizadeh, B., Babrzadeh, F. & Church, G. M. High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. *Nat Biotechnol* **28**, 1291-1294, (2010).

52 Li, J., Wang, L., Mamon, H., Kulke, M. H., Berbeco, R. & Makrigiorgos, G. M. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nat Med* **14**, 579-584, (2008).

53 Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. & Turner, S. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138, (2009).

54 Chory, J. & Pollard, J. D., Jr. Separation of small DNA fragments by conventional gel electrophoresis. *Curr Protoc Mol Biol* **Chapter 2**, Unit2 7, (2001).

55 Collins, A. R. Oxidative DNA damage, antioxidants, and cancer. *Bioessays* **21**, 238-246, (1999).

56 Morley, A. A., Cox, S. & Holliday, R. Human lymphocytes resistant to 6-thioguanine increase with age. *Mech Ageing Dev* **19**, 21-26, (1982).

57 Trainor, K. J., Wigmore, D. J., Chrysostomou, A., Dempsey, J. L., Seshadri, R. & Morley, A. A. Mutation frequency in human lymphocytes increases with age. *Mech Ageing Dev* **27**, 83-86, (1984).

58 Williams, G. T., Geraghty, J. M., Campbell, F., Appleton, M. A. & Williams, E. D. Normal colonic mucosa in hereditary non-polyposis colorectal cancer shows no generalised increase in somatic mutation. *Br J Cancer* **71**, 1077-1080, (1995).

59 Araten, D. J., Nafa, K., Pakdeesuwan, K. & Luzzatto, L. Clonal populations of hematopoietic cells with paroxysmal nocturnal hemoglobinuria genotype and

phenotype are present in normal individuals. *Proc Natl Acad Sci U S A* **96**, 5209-5214, (1999).

60 Bodenteich, A., Mitchell, L. G. & Merril, C. R. A lifetime of retinal light exposure does not appear to increase mitochondrial mutations. *Gene* **108**, 305-309, (1991).

61 Howell, N., Kubacka, I. & Mackey, D. A. How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* **59**, 501-509, (1996).

62 Khrapko, K., Coller, H. A., Andre, P. C., Li, X. C., Hanekamp, J. S. & Thilly, W. G. Mitochondrial mutational spectra in human cells and tissues. *Proc Natl Acad Sci U S A* **94**, 13798-13803, (1997).

63 Heyer, E., Zietkiewicz, E., Rochowski, A., Yotova, V., Puymirat, J. & Labuda, D. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet* **69**, 1113-1126, (2001).

64 Howell, N., Smejkal, C. B., Mackey, D. A., Chinnery, P. F., Turnbull, D. M. & Herrnstadt, C. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet* **72**, 659-670, (2003).

65 Taylor, R. W., Barron, M. J., Borthwick, G. M., Gospel, A., Chinnery, P. F., Samuels, D. C., Taylor, G. A., Plusa, S. M., Needham, S. J., Greaves, L. C., Kirkwood, T. B. & Turnbull, D. M. Mitochondrial DNA mutations in human colonic crypt stem cells. *J Clin Invest* **112**, 1351-1360, (2003).

66 Arbyn, M., Anttila, A., Jordan, J., Ronco, G., Schenck, U., Segnan, N., Wiener, H., Herbert, A. & von Karsa, L. European Guidelines for Quality Assurance in Cervical

Cancer Screening. Second edition--summary document. *Ann Oncol* **21**, 448-458, (2010).

67 DeMay, R. M. *Practical principles of cytopathology*. (ASCP Press, 2010).

68 Howlader, N., Noone, A. M., Krapcho, M., Neyman N., Aminou, R., Altekruse, S. F., Kosary, C. L., Ruhl, J., Tatalovich, Z., Cho, H., Mariotto, A., Eisner, M. P., Lewis, D. R., Chen, H. S., Feuer, E. J. & Cronin, K. A. *SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations)*, <http://seer.cancer.gov/csr/1975_2009_pops09/> (2012).

69 Bray, F., Ren, J. S., Masuyer, E. & Ferlay, J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int J Cancer* **132**, 1133-1145, (2013).

70 Ferlay, J., Shin, H. R., Bray, F., Forman, D., Mathers, C. & Parkin, D. M. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* **127**, 2893-2917, (2010).

71 Sams, S. B., Currens, H. S. & Raab, S. S. Liquid-based Papanicolaou tests in endometrial carcinoma diagnosis. Performance, error root cause analysis, and quality improvement. *Am J Clin Pathol* **137**, 248-254, (2012).

72 Smith, P., Bakos, O., Heimer, G. & Ulmsten, U. Transvaginal ultrasound for identifying endometrial abnormality. *Acta Obstet Gynecol Scand* **70**, 591-594, (1991).

73 Mitchell, H., Giles, G. & Medley, G. Accuracy and survival benefit of cytological prediction of endometrial carcinoma on routine cervical smears. *Int J Gynecol Pathol* **12**, 34-40, (1993).

74 Carlson, K. J., Skates, S. J. & Singer, D. E. Screening for ovarian cancer. *Ann Intern Med* **121**, 124-132, (1994).

75 Cramer, D. W., Bast, R. C., Jr., Berg, C. D., Diamandis, E. P., Godwin, A. K., Hartge, P., Lokshin, A. E., Lu, K. H., McIntosh, M. W., Mor, G., Patriotis, C., Pinsky, P. F., Thornquist, M. D., Scholler, N., Skates, S. J., Sluss, P. M., Srivastava, S., Ward, D. C., Zhang, Z., Zhu, C. S. & Urban, N. Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer prevention research* **4**, 365-374, (2011).

76 Meden, H. & Fattahi-Meibodi, A. CA 125 in benign gynecological conditions. *Int J Biol Markers* **13**, 231-237, (1998).

77 Buys, S. S., Partridge, E., Black, A., Johnson, C. C., Lamerato, L., Isaacs, C., Reding, D. J., Greenlee, R. T., Yokochi, L. A., Kessel, B., Crawford, E. D., Church, T. R., Andriole, G. L., Weissfeld, J. L., Fouad, M. N., Chia, D., O'Brien, B., Ragard, L. R., Clapp, J. D., Rathmell, J. M., Riley, T. L., Hartge, P., Pinsky, P. F., Zhu, C. S., Izmirlian, G., Kramer, B. S., Miller, A. B., Xu, J. L., Prorok, P. C., Gohagan, J. K. & Berg, C. D. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. *JAMA* **305**, 2295-2303, (2011).

78 ACOG Practice Bulletin. Clinical Management Guidelines for Obstetrician-Gynecologists. Number 60, March 2005. Pregestational diabetes mellitus. *Obstet Gynecol* **105**, 675-685, (2005).

79 Partridge, E., Kreimer, A. R., Greenlee, R. T., Williams, C., Xu, J. L., Church, T. R., Kessel, B., Johnson, C. C., Weissfeld, J. L., Isaacs, C., Andriole, G. L., Ogden, S., Ragard, L. R. & Buys, S. S. Results from four rounds of ovarian cancer screening in a randomized trial. *Obstet Gynecol* **113**, 775-782, (2009).

80 *Detailed guide: ovarian cancer — can ovarian cancer be found early?*, <http://www.cancer.org/cancer/ovariancancer/detailedguide/ovarian-cancer-detection> (2009).

81 Force, U. S. P. S. T. Screening for ovarian cancer: recommendation statement. *Am Fam Physician* **71**, 759-762, (2005).

82 ACOG Committee Opinion: number 280, December 2002. The role of the generalist obstetrician-gynecologist in the early detection of ovarian cancer. *Obstet Gynecol* **100**, 1413-1416, (2002).

83 *Practice Guidelines in Oncology: ovarian cancer and genetic screening*, <http://www.nccn.org/professionals/physician_gls/PDF/genetics_screening.pdf> (2012).

84 Lindor, N. M., Petersen, G. M., Hadley, D. W., Kinney, A. Y., Miesfeldt, S., Lu, K. H., Lynch, P., Burke, W. & Press, N. Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. *JAMA* **296**, 1507-1517, (2006).

85 Marques, J. P., Costa, L. B., Pinto, A. P., Lima, A. F., Duarte, M. E., Barbosa, A. P. & Medeiros, P. L. Atypical glandular cells and cervical cancer: systematic review. *Rev Assoc Med Bras* **57**, 234-238, (2011).

86 Insinga, R. P., Glass, A. G. & Rush, B. B. Diagnoses and outcomes in cervical cancer screening: a population-based study. *Am J Obstet Gynecol* **191**, 105-113, (2004).

87 Sharpless, K. E., Schnatz, P. F., Mandavilli, S., Greene, J. F. & Sorosky, J. I. Dysplasia associated with atypical glandular cells on cervical cytology. *Obstet Gynecol* **105**, 494-500, (2005).

88 DeSimone, C. P., Day, M. E., Tovar, M. M., Dietrich, C. S., 3rd, Eastham, M. L. & Modesitt, S. C. Rate of pathology from atypical glandular cell Pap tests classified by the Bethesda 2001 nomenclature. *Obstet Gynecol* **107**, 1285-1291, (2006).

89 Geier, C. S., Wilson, M. & Creasman, W. Clinical evaluation of atypical glandular cells of undetermined significance. *Am J Obstet Gynecol* **184**, 64-69, (2001).

90 Malpica, A., Deavers, M. T., Lu, K., Bodurka, D. C., Atkinson, E. N., Gershenson, D. M. & Silva, E. G. Grading ovarian serous carcinoma using a two-tier system. *Am J Surg Pathol* **28**, 496-504, (2004).

91 Ries, L. A. G., Young, J. L., Keel, G. E., Eisner, M. P., Lin, Y. D. & Horner, M.-J. *SEER Survival Monograph: Cancer Survival Among Adults: US SEER Program, 1988-2001, Patient and Tumor Characteristics*, 2007).

92 Ichigo, S., Takagi, H., Matsunami, K., Murase, T., Ikeda, T. & Imai, A. Transitional cell carcinoma of the ovary (Review). *Oncology letters* **3**, 3-6, (2012).

93 Kommoss, F., Kommoss, S., Schmidt, D., Trunk, M. J., Pfisterer, J., du Bois, A. & Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe, O. Survival benefit for patients with advanced-stage transitional cell carcinomas vs. other subtypes of ovarian carcinoma after chemotherapy with platinum and paclitaxel. *Gynecol Oncol* **97**, 195-199, (2005).

94 Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615, (2011).

95 Jones, S., Wang, T. L., Shih Ie, M., Mao, T. L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L. A., Jr., Vogelstein, B., Kinzler, K. W., Velculescu, V. E. & Papadopoulos, N. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-231, (2010).

96 Jones, S., Wang, T. L., Kurman, R. J., Nakayama, K., Velculescu, V. E., Vogelstein, B., Kinzler, K. W., Papadopoulos, N. & Shih Ie, M. Low-grade serous carcinomas of the ovary contain very few point mutations. *J Pathol* **226**, 413-420, (2012).

97 Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., Teague, J. W., Campbell, P. J., Stratton, M. R. & Futreal, P. A. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-950, (2011).

98 Barrow, E., Robinson, L., Alduaij, W., Shenton, A., Clancy, T., Lalloo, F., Hill, J. & Evans, D. G. Cumulative lifetime incidence of extracolonic cancers in Lynch syndrome: a report of 121 families with proven mutations. *Clin Genet* **75**, 141-149, (2009).

99 Oda, K., Stokoe, D., Taketani, Y. & McCormick, F. High frequency of coexistent mutations of PIK3CA and PTEN genes in endometrial carcinoma. *Cancer Res* **65**, 10669-10673, (2005).

100 Kuhn, E., Wu, R. C., Guan, B., Wu, G., Zhang, J., Wang, Y., Song, L., Yuan, X., Wei, L., Roden, R. B., Kuo, K. T., Nakayama, K., Clarke, B., Shaw, P., Olvera, N., Kurman, R. J., Levine, D. A., Wang, T. L. & Shih, I. M. Identification of Molecular Pathway Aberrations in Uterine Serous Carcinoma by Genome-wide Analyses. *J Natl Cancer Inst*, (2012).

101 Hamilton, C. A., Cheung, M. K., Osann, K., Chen, L., Teng, N. N., Longacre, T. A., Powell, M. A., Hendrickson, M. R., Kapp, D. S. & Chan, J. K. Uterine papillary serous and clear cell carcinomas predict for poorer survival compared to grade 3 endometrioid corpus cancers. *British journal of cancer* **94**, 642-646, (2006).

102 Traut, H. F. & Papanicolaou, G. N. Cancer of the Uterus: The Vaginal Smear in Its Diagnosis. *Cal West Med* **59**, 121-122, (1943).

103 Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat Med* **10**, 789-799, (2004).

104 Cooper, J. M. & Erickson, M. L. Endometrial sampling techniques in the diagnosis of abnormal uterine bleeding. *Obstet Gynecol Clin North Am* **27**, 235-244, (2000).

105 Gunderson, C. C., Fader, A. N., Carson, K. A. & Bristow, R. E. Oncologic and reproductive outcomes with progestin therapy in women with endometrial hyperplasia and grade 1 adenocarcinoma: a systematic review. *Gynecol Oncol* **125**, 477-482, (2012).

106 Bristow, R. E., Tomacruz, R. S., Armstrong, D. K., Trimble, E. L. & Montz, F. J. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol* **20**, 1248-1259, (2002).

107 Mayrand, M. H., Duarte-Franco, E., Rodrigues, I., Walter, S. D., Hanley, J., Ferenczy, A., Ratnam, S., Coutlee, F. & Franco, E. L. Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *N Engl J Med* **357**, 1579-1588, (2007).

108 Naucler, P., Ryd, W., Tornberg, S., Strand, A., Wadell, G., Elfgren, K., Radberg, T., Strander, B., Johansson, B., Forslund, O., Hansson, B. G., Rylander, E. & Dillner, J. Human papillomavirus and Papanicolaou tests to screen for cervical cancer. *N Engl J Med* **357**, 1589-1597, (2007).

109 Willis, B. H., Barton, P., Pearmain, P., Bryan, S. & Hyde, C. Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic

analysis and a simulation modelling exercise applied to the UK. *Health Technol Assess* **9**, 1-207, iii, (2005).

110 Pecorelli, S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet* **105**, 103-104, (2009).

111 Rago, C., Huso, D. L., Diehl, F., Karim, B., Liu, G., Papadopoulos, N., Samuels, Y., Velculescu, V. E., Vogelstein, B., Kinzler, K. W. & Diaz, L. A., Jr. Serial Assessment of Human Tumor Burdens in Mice by the Analysis of Circulating DNA. *Cancer Res* **67**, 9364-9370, (2007).

112 Wu, J., Matthaei, H., Maitra, A., Dal Molin, M., Wood, L. D., Eshleman, J. R., Goggins, M., Canto, M. I., Schulick, R. D., Edil, B. H., Wolfgang, C. L., Klein, A. P., Diaz, L. A., Jr., Allen, P. J., Schmidt, C. M., Kinzler, K. W., Papadopoulos, N., Hruban, R. H. & Vogelstein, B. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. *Sci Transl Med* **3**, 92ra66, (2011).

113 He, J., Wu, J., Jiao, Y., Wagner-Johnston, N., Ambinder, R. F., Diaz, L. A., Jr., Kinzler, K. W., Vogelstein, B. & Papadopoulos, N. IgH gene rearrangements as plasma biomarkers in Non- Hodgkin's lymphoma patients. *Oncotarget* **2**, 178-185, (2011).

114 Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. & Sirotkin, K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311, (2001).

115 Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology* **132**, 365-386., (2000).

116 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194, (1998).

117 Kinde, I., Bettegowda, C., Wang, Y., Wu, J., Agrawal, N., Shih Ie, M., Kurman, R., Dao, F., Levine, D. A., Giuntoli, R., Roden, R., Eshleman, J. R., Carvalho, J. P., Marie, S. K., Papadopoulos, N., Kinzler, K. W., Vogelstein, B. & Diaz, L. A., Jr. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci Transl Med* **5**, 167ra164, (2013).

118 Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA: a cancer journal for clinicians* **63**, 11-30, (2013).

119 Netto, G. J. Clinical applications of recent molecular advances in urologic malignancies: no longer chasing a "mirage"? *Advances in anatomic pathology* **20**, 175-203, (2013).

120 Lotan, Y. & Roehrborn, C. G. Sensitivity and specificity of commonly available bladder tumor markers versus cytology: results of a comprehensive literature review and meta-analyses. *Urology* **61**, 109-118, (2003).

121 Friedrich, M. G., Toma, M. I., Hellstern, A., Pantel, K., Weisenberger, D. J., Noldus, J. & Huland, H. Comparison of multitarget fluorescence in situ hybridization in urine with other noninvasive tests for detecting bladder cancer. *BJU international* **92**, 911-914, (2003).

122 Horn, S., Figl, A., Rachakonda, P. S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., Schadendorf, D. & Kumar, R. TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-961, (2013).

123 Huang, F. W., Hodis, E., Xu, M. J., Kryukov, G. V., Chin, L. & Garraway, L. A. Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-959, (2013).

124 Killela, P. J., Reitman, Z. J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz, L. A., Jr., Friedman, A. H., Friedman, H., Gallia, G. L., Giovanella, B. C., Grollman, A. P., He, T. C., He, Y., Hruban, R. H., Jallo, G. I., Mandahl, N., Meeker, A. K., Mertens, F., Netto, G. J., Rasheed, B. A., Riggins, G. J., Rosenquist, T. A., Schiffman, M., Shih Ie, M., Theodorescu, D., Torbenson, M. S., Velculescu, V. E., Wang, T. L., Wentzensen, N., Wood, L. D., Zhang, M., McLendon, R. E., Bigner, D. D., Kinzler, K. W., Vogelstein, B., Papadopoulos, N. & Yan, H. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci U S A* **110**, 6021-6026, (2013).

125 Oosterhuis, J. W., Schapers, R. F., Janssen-Heijnen, M. L., Pauwels, R. P., Newling, D. W. & ten Kate, F. Histological grading of papillary urothelial carcinoma of the bladder: prognostic value of the 1998 WHO/ISUP classification system and comparison with conventional grading systems. *Journal of clinical pathology* **55**, 900-905, (2002).

126 Herr, H. W., Donat, S. M. & Reuter, V. E. Management of low grade papillary bladder tumors. *The Journal of urology* **178**, 1201-1205, (2007).

127 Miyamoto, H., Brimo, F., Schultz, L., Ye, H., Miller, J. S., Fajardo, D. A., Lee, T. K., Epstein, J. I. & Netto, G. J. Low-grade papillary urothelial carcinoma of the urinary bladder: a clinicopathologic analysis of a post-World Health Organization/International Society of Urological Pathology classification cohort from a single academic center. *Archives of pathology & laboratory medicine* **134**, 1160-1163, (2010).

128 Chaux, A., Karram, S., Miller, J. S., Fajardo, D. A., Lee, T. K., Miyamoto, H. & Netto, G. J. High-grade papillary urothelial carcinoma of the urinary tract: a

clinicopathologic analysis of a post-World Health Organization/International Society of Urological Pathology classification cohort from a single academic center. *Human pathology* **43**, 115-120, (2012).

129 Wu, X. R. Urothelial tumorigenesis: a tale of divergent pathways. *Nature reviews. Cancer* **5**, 713-725, (2005).

130 Botteman, M. F., Pashos, C. L., Redaelli, A., Laskin, B. & Hauser, R. The health economics of bladder cancer: a comprehensive review of the published literature. *PharmacoEconomics* **21**, 1315-1330, (2003).

131 Eble, J. N., Sauter, G. & Epstein, J. I. *Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs*. (International Agency for Research on Cancer Press, 2004).

132 Kinde, I., Munari, E., Faraj, S. F., Hruban, R. H., Schoenberg, M., Bivalacqua, T., Allaf, M., Springer, S., Wang, Y., Diaz, L. A., Jr., Kinzler, K. W., Vogelstein, B., Papadopoulos, N. & Netto, G. J. TERT promoter mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine. *Cancer Res* **73**, 7162-7167, (2013).

133 Cancer Facts & Figures 2012. *American Cancer Society*, (2012).

134 Mazzucchelli, R., Colanzi, P., Pomante, R., Muzzonigro, G. & Montironi, R. Prostate tissue and serum markers. *Adv Clin Path* **4**, 111-120, (2000).

135 Ruibal Morell, A. CEA serum levels in non-neoplastic disease. *Int J Biol Markers* **7**, 160-166, (1992).

136 Sikaris, K. A. CA125--a test with a change of heart. *Heart Lung Circ* **20**, 634-640, (2011).

137 Ballehaninna, U. K. & Chamberlain, R. S. Serum CA 19-9 as a Biomarker for Pancreatic Cancer-A Comprehensive Review. *Indian J Surg Oncol* **2**, 88-100, (2011).

138 Wanebo, H. J., Rao, B., Pinsky, C. M., Hoffman, R. G., Stearns, M., Schwartz, M. K. & Oettgen, H. F. Preoperative carcinoembryonic antigen level as a prognostic indicator in colorectal cancer. *N Engl J Med* **299**, 448-451, (1978).

139 Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nat Rev Genet* **13**, 795-806, (2012).

140 Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**, 685-696, (2010).

141 Stratton, M. R. Exploring the genomes of cancer cells: progress and promise. *Science* **331**, 1553-1558, (2011).

142 Mardis, E. R. Genome sequencing and cancer. *Curr Opin Genet Dev* **22**, 245-250, (2012).

143 Fleischhacker, M. & Schmidt, B. Circulating nucleic acids (CNAs) and cancer—A survey. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1775**, 181-232, (2007).

144 Alix-Panabieres, C., Schwarzenbach, H. & Pantel, K. Circulating tumor cells and circulating tumor DNA. *Annu Rev Med* **63**, 199-215, (2012).

145 Stroun, M., Lyautey, J., Lederrey, C., Olson-Sand, A. & Anker, P. About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release. *Clinica Chimica Acta* **313**, 139-142, (2001).

146 Racila, E., Euhus, D., Weiss, A. J., Rao, C., McConnell, J., Terstappen, L. W. & Uhr, J. W. Detection and characterization of carcinoma cells in the blood. *Proc Natl Acad Sci U S A* **95**, 4589-4594, (1998).

147 Dawson, S. J., Tsui, D. W., Murtaza, M., Biggs, H., Rueda, O. M., Chin, S. F., Dunning, M. J., Gale, D., Forshew, T., Mahler-Araujo, B., Rajan, S., Humphray, S., Becq, J., Halsall, D., Wallis, M., Bentley, D., Caldas, C. & Rosenfeld, N. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**, 1199-1209, (2013).

148 Maheswaran, S., Sequist, L. V., Nagrath, S., Ulkus, L., Brannigan, B., Collura, C. V., Inserra, E., Diederichs, S., Iafrate, A. J., Bell, D. W., Digumarthy, S., Muzikansky, A., Irimia, D., Settleman, J., Tompkins, R. G., Lynch, T. J., Toner, M. & Haber, D. A. Detection of mutations in EGFR in circulating lung-cancer cells. *N Engl J Med* **359**, 366-377, (2008).

149 Punnoose, E. A., Atwal, S., Liu, W., Raja, R., Fine, B. M., Hughes, B. G. M., Hicks, R. J., Hampton, G. M., Amler, L. C., Pirzkall, A. & Lackner, M. R. Evaluation of Circulating Tumor Cells and Circulating Tumor DNA in Non–Small Cell Lung Cancer: Association with Clinical Endpoints in a Phase II Clinical Trial of Pertuzumab and Erlotinib. *Clinical Cancer Research* **18**, 2391-2401, (2012).

150 Sausen, M., Leary, R. J., Jones, S., Wu, J., Reynolds, C. P., Liu, X., Blackford, A., Parmigiani, G., Diaz, L. A., Jr., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., Velculescu, V. E. & Hogarty, M. D. Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat Genet* **45**, 12-17, (2013).

151 Diaz, L. A., Jr., Williams, R. T., Wu, J., Kinde, I., Hecht, J. R., Berlin, J., Allen, B., Bozic, I., Reiter, J. G., Nowak, M. A., Kinzler, K. W., Oliner, K. S. & Vogelstein, B. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540, (2012).

152 Diehl, F., Schmidt, K., Choti, M. A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S. A., Kinzler, K. W., Vogelstein, B. & Diaz, L. A., Jr. Circulating mutant DNA to assess tumor dynamics. *Nature Medicine* **14**, 985-990, (2008).

153 Forshew, T., Murtaza, M., Parkinson, C., Gale, D., Tsui, D. W., Kaper, F., Dawson, S. J., Piskorz, A. M., Jimenez-Linan, M., Bentley, D., Hadfield, J., May, A. P., Caldas, C., Brenton, J. D. & Rosenfeld, N. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* **4**, 136ra168, (2012).

154 Murtaza, M., Dawson, S. J., Tsui, D. W., Gale, D., Forshew, T., Piskorz, A. M., Parkinson, C., Chin, S. F., Kingsbury, Z., Wong, A. S., Marass, F., Humphray, S., Hadfield, J., Bentley, D., Chin, T. M., Brenton, J. D., Caldas, C. & Rosenfeld, N. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108-112, (2013).

155 Kuang, Y., Rogers, A., Yeap, B. Y., Wang, L., Makrigiorgos, M., Vetrand, K., Thiede, S., Distel, R. J. & Jänne, P. A. Noninvasive Detection of EGFR T790M in Gefitinib or Erlotinib Resistant Non–Small Cell Lung Cancer. *Clinical Cancer Research* **15**, 2630-2636, (2009).

156 Taniguchi, K., Uchida, J., Nishino, K., Kumagai, T., Okuyama, T., Okami, J., Higashiyama, M., Kodama, K., Imamura, F. & Kato, K. Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. *Clin Cancer Res* **17**, 7808-7815, (2011).

157 Chang, H. W., Lee, S. M., Goodman, S. N., Singer, G., Cho, S. K., Sokoll, L. J., Montz, F. J., Roden, R., Zhang, Z., Chan, D. W., Kurman, R. J. & Shih Ie, M.

Assessment of plasma DNA levels, allelic imbalance, and CA 125 as diagnostic tests for cancer. *J Natl Cancer Inst* **94**, 1697-1703, (2002).

158 Holdhoff, M., Schmidt, K., Donehower, R. & Diaz, L. A., Jr. Analysis of circulating tumor DNA to confirm somatic KRAS mutations. *J Natl Cancer Inst* **101**, 1284-1285, (2009).

159 Diehl, F., Schmidt, K., Choti, M. A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S. A., Kinzler, K. W., Vogelstein, B. & Diaz, L. A., Jr. Circulating mutant DNA to assess tumor dynamics. *Nat Med* **14**, 985-990, (2008).

160 Karapetis, C. S., Khambata-Ford, S., Jonker, D. J., O'Callaghan, C. J., Tu, D., Tebbutt, N. C., Simes, R. J., Chalchal, H., Shapiro, J. D., Robitaille, S., Price, T. J., Shepherd, L., Au, H.-J., Langer, C., Moore, M. J. & Zalcberg, J. R. K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer. *New England Journal of Medicine* **359**, 1757-1765, (2008).

161 Misale, S., Yaeger, R., Hobor, S., Scala, E., Janakiraman, M., Liska, D., Valtorta, E., Schiavo, R., Buscarino, M., Siravegna, G., Bencardino, K., Cercek, A., Chen, C. T., Veronese, S., Zanon, C., Sartore-Bianchi, A., Gambacorta, M., Gallicchio, M., Vakiani, E., Boscaro, V., Medico, E., Weiser, M., Siena, S., Di Nicolantonio, F., Solit, D. & Bardelli, A. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532-536, (2012).

162 Di Nicolantonio, F., Martini, M., Molinari, F., Sartore-Bianchi, A., Arena, S., Saletti, P., De Dosso, S., Mazzucchelli, L., Frattini, M., Siena, S. & Bardelli, A. Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol* **26**, 5705-5712, (2008).

163 Loupakis, F., Ruzzo, A., Cremolini, C., Vincenzi, B., Salvatore, L., Santini, D., Masi, G., Stasi, I., Canestrari, E., Rulli, E., Floriani, I., Bencardino, K., Galluccio, N., Catalano, V., Tonini, G., Magnani, M., Fontanini, G., Basolo, F., Falcone, A. & Graziano, F. KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *Br J Cancer* **101**, 715-721, (2009).

164 Peeters, M., Oliner, K. S., Parker, A., Siena, S., Van Cutsem, E., Huang, J., Humblet, Y., Van Laethem, J. L., Andre, T., Wiezorek, J., Reese, D. & Patterson, S. D. Massively parallel tumor multigene sequencing to evaluate response to panitumumab in a randomized phase III study of metastatic colorectal cancer. *Clin Cancer Res* **19**, 1902-1912, (2013).

165 Montagut, C., Dalmases, A., Bellosillo, B., Crespo, M., Pairet, S., Iglesias, M., Salido, M., Gallen, M., Marsters, S., Tsai, S. P., Minoche, A., Somasekar, S., Serrano, S., Himmelbauer, H., Bellmunt, J., Rovira, A., Settleman, J., Bosch, F. & Albanell, J. Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. *Nat Med* **18**, 221-223, (2012).

166 Tougeron, D., Cortes, U., Ferru, A., Villalva, C., Silvain, C., Tourani, J. M., Levillain, P. & Karayan-Tapon, L. Epidermal growth factor receptor (EGFR) and KRAS mutations during chemotherapy plus anti-EGFR monoclonal antibody treatment in metastatic colorectal cancer. *Cancer Chemother Pharmacol*, (2013).

167 Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S. M., Riggins, G. J., Willson, J. K., Markowitz, S., Kinzler, K. W.,

Vogelstein, B. & Velculescu, V. E. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554, (2004).

168 Duffy, M. J. Clinical uses of tumor markers: a critical review. *Crit Rev Clin Lab Sci* **38**, 225-262, (2001).

169 Saltz, L. B. Biomarkers in colorectal cancer: added value or just added expense? *Expert Rev Mol Diagn* **8**, 231-233, (2008).

170 Sawyers, C. L. The cancer biomarker problem. *Nature* **452**, 548-552, (2008).

171 Li, M., Diehl, F., Dressman, D., Vogelstein, B. & Kinzler, K. W. BEAMing up for detection and quantification of rare sequence variants. *Nat Methods* **3**, 95-97, (2006).

172 Leary, R. J., Kinde, I., Diehl, F., Schmidt, K., Clouser, C., Duncan, C., Antipova, A., Lee, C., McKernan, K., De La Vega, F. M., Kinzler, K. W., Vogelstein, B., Diaz, L. A., Jr. & Velculescu, V. E. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med* **2**, 20ra14, (2010).

173 Stott, S. L., Hsu, C. H., Tsukrov, D. I., Yu, M., Miyamoto, D. T., Waltman, B. A., Rothenberg, S. M., Shah, A. M., Smas, M. E., Korir, G. K., Floyd, F. P., Jr., Gilman, A. J., Lord, J. B., Winokur, D., Springer, S., Irimia, D., Nagrath, S., Sequist, L. V., Lee, R. J., Isselbacher, K. J., Maheswaran, S., Haber, D. A. & Toner, M. Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. *Proc Natl Acad Sci U S A* **107**, 18392-18397, (2010).

174 Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N. & Perucho, M. Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell* **53**, 549-554, (1988).

175 Brooks, J. D. Translational genomics: the challenge of developing cancer biomarkers. *Genome Res* **22**, 183-187, (2012).

176 Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szabo, S., Diaz, L. A., Jr., Goodman, S. N., David, K. A., Juhl, H., Kinzler, K. W. & Vogelstein, B. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A* **102**, 16368-16373, (2005).

177 De Roock, W., Claes, B., Bernasconi, D., De Schutter, J., Biesmans, B., Fountzilas, G., Kalogeras, K. T., Kotoula, V., Papamichael, D., Laurent-Puig, P., Penault-Llorca, F., Rougier, P., Vincenzi, B., Santini, D., Tonini, G., Cappuzzo, F., Frattini, M., Molinari, F., Saletti, P., De Dosso, S., Martini, M., Bardelli, A., Siena, S., Sartore-Bianchi, A., Tabernero, J., Macarulla, T., Di Fiore, F., Gangloff, A. O., Ciardiello, F., Pfeiffer, P., Qvortrup, C., Hansen, T. P., Van Cutsem, E., Piessevaux, H., Lambrechts, D., Delorenzi, M. & Tejpar, S. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol* **11**, 753-762, (2010).

178 Misale, S., Arena, S., Lamba, S., Siravegna, G., Lallo, A., Hobor, S., Russo, M., Sartore-Bianchi, A., Bencardino, K., Amatu, A., Siena, S. & Bardelli, A. Blockade of EGFR and MEK intercepts genetically heterogeneous mechanisms of acquired resistance to anti-EGFR therapies in colorectal cancer. . *Sci Transl Med* **6**, (2014).

179 Vaughn, C. P., Zobell, S. D., Furtado, L. V., Baker, C. L. & Samowitz, W. S. Frequency of KRAS, BRAF, and NRAS mutations in colorectal cancer. *Genes Chromosomes Cancer* **50**, 307-312, (2011).

180 Douillard, J. Y., Oliner, K. S., Siena, S., Tabernero, J., Burkes, R., Barugel, M., Humblet, Y., Bodoky, G., Cunningham, D., Jassem, J., Rivera, F., Kocakova, I., Ruff, P., Blasinska-Morawiec, M., Smakal, M., Canon, J. L., Rother, M., Williams, R., Rong,

A., Wiezorek, J., Sidhu, R. & Patterson, S. D. Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med* **369**, 1023-1034, (2013).

181 Bidard, F. C., Fehm, T., Ignatiadis, M., Smerage, J. B., Alix-Panabieres, C., Janni, W., Messina, C., Paoletti, C., Muller, V., Hayes, D. F., Piccart, M. & Pierga, J. Y. Clinical application of circulating tumor cells in breast cancer: overview of the current interventional trials. *Cancer Metastasis Rev* **32**, 179-188, (2013).

182 Leary, R. J., Sausen, M., Kinde, I., Papadopoulos, N., Carpten, J. D., Craig, D., O'Shaughnessy, J., Kinzler, K. W., Parmigiani, G., Vogelstein, B., Diaz, L. A., Jr. & Velculescu, V. E. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med* **4**, 162ra154, (2012).

183 Jones, S., Zhang, X., Parsons, D. W., Lin, J. C., Leary, R. J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S. M., Fu, B., Lin, M. T., Calhoun, E. S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., Smith, D. R., Hidalgo, M., Leach, S. D., Klein, A. P., Jaffee, E. M., Goggins, M., Maitra, A., Iacobuzio-Donahue, C., Eshleman, J. R., Kern, S. E., Hruban, R. H., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V. E. & Kinzler, K. W. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801-1806, (2008).

184 Bettegowda, C., Agrawal, N., Jiao, Y., Sausen, M., Wood, L. D., Hruban, R. H., Rodriguez, F. J., Cahill, D. P., McLendon, R., Riggins, G., Velculescu, V. E., Oba-Shinjo, S. M., Marie, S. K., Vogelstein, B., Bigner, D., Yan, H., Papadopoulos, N. & Kinzler, K. W. Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* **333**, 1453-1455, (2011).

185 Bettegowda, C., Agrawal, N., Jiao, Y., Wang, Y., Wood, L. D., Rodriguez, F. J., Hruban, R. H., Gallia, G. L., Binder, Z. A., Riggins, C. J., Salmasi, V., Riggins, G. J., Reitman, Z. J., Rasheed, A., Keir, S., Shinjo, S., Marie, S., McLendon, R., Jallo, G., Vogelstein, B., Bigner, D., Yan, H., Kinzler, K. W. & Papadopoulos, N. Exomic sequencing of four rare central nervous system tumor types. *Oncotarget* **4**, 572-583, (2013).

186 Bettegowda, C., Sausen, M., Leary, R. J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B. R., Wang, H., Luber, B., Alani, R. M., Antonarakis, E. S., Azad, N. S., Bardelli, A., Brem, H., Cameron, J. L., Lee, C. C., Fecher, L. A., Gallia, G. L., Gibbs, P., Le, D., Giuntoli, R. L., Goggins, M., Hogarty, M. D., Holdhoff, M., Hong, S. M., Jiao, Y., Juhl, H. H., Kim, J. J., Siravegna, G., Laheru, D. A., Lauricella, C., Lim, M., Lipson, E. J., Marie, S. K., Netto, G. J., Oliner, K. S., Olivi, A., Olsson, L., Riggins, G. J., Sartore-Bianchi, A., Schmidt, K., Shih l, M., Oba-Shinjo, S. M., Siena, S., Theodorescu, D., Tie, J., Harkins, T. T., Veronese, S., Wang, T. L., Weingart, J. D., Wolfgang, C. L., Wood, L. D., Xing, D., Hruban, R. H., Wu, J., Allen, P. J., Schmidt, C. M., Choti, M. A., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Papadopoulos, N. & Diaz, L. A., Jr. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* **6**, 224ra224, (2014).

Isaac A. Kinde

ik@jhmi.edu

Education

2015 (expected)	M.D.	Johns Hopkins University School of Medicine
2015 (expected)	Ph.D.	Johns Hopkins University School of Medicine Program in Cellular and Molecular Medicine Mentor: Bert Vogelstein, M.D.
2005	B.S.	University of Maryland, Baltimore County Biological Sciences Phi Beta Kappa *magna cum laude*

Research Experience

2007 – 2013 Ludwig Center for Cancer Genetics and Therapeutics
Position: Ph.D. candidate
Mentor: Bert Vogelstein, M.D.
Thesis: Rational cancer diagnostics using rare mutation detection technology with massively parallel sequencing
Location: Johns Hopkins University School of Medicine

2006 High Throughput Biology Center
Position: Pre-doctoral rotation student
Mentor: Jef Boeke, Ph.D.
Project: The role of retrotransposons in *S. cerevisiae* (baker's yeast) DNA repair.
Location: Johns Hopkins University School of Medicine

2005 Central Autonomic Regulation Laboratory
Position: Pre-doctoral rotation student
Mentors: Lawrence Schramm, Ph.D. and Ronald Schnaar, Ph.D.
Project: Organic synthesis of a specific inhibitor of glucosylceramide synthesis
Location: Johns Hopkins University School of Medicine

2002 – 2005 Howard Hughes Medical Institute
Position: Undergraduate researcher
Mentor: Michael Summers, Ph.D.
Project: Structure-based design of HIV-1 capsid assembly inhibitors
Location: University of Maryland, Baltimore County

2003 Howard Hughes Medical Institute
Position: Undergraduate researcher
Mentor: Louis Reichardt, Ph.D.
Project: Antioxidant effect in rat hippocampal dendritic spine morphogenesis
Location: University of California, San Francisco

Awards & Honors

2005 – 2015	Medical Scientist Training Program M.D./Ph.D. Fellowship National Institutes of Health
2014	Rising Star Award, Outstanding Alumni of the Year University of Maryland, Baltimore County
2014	The Hans Joaquim Prochaska Research Award, Young Investigators' Day Johns Hopkins University School of Medicine
2013	Medalist, National Collegiate Inventors Competition Invent Now
2013	30 Under 30, Science and Healthcare Forbes Magazine
2011	Selected Young Scientist Participant, United States Delegation The Council for the 61st Lindau Nobel Laureate Meeting (Lindau, Germany)
2009 – 2010	Graduate Research Fellowship Merck and United Negro College Fund
2006	Summer Research Fellowship American Federation for Aging Research
2005	Emerald Honor Award in Student Leadership Science Spectrum Magazine
2005	Faculty Award of Excellence in the Biological Sciences University of Maryland, Baltimore County
2005	Certificate of General Honors The Honors College of University of Maryland, Baltimore County
2002 – 2005	Undergraduate Research Scholar Howard Hughes Medical Institute
2003 – 2005	Minority Access to Research Careers (MARC) Undergraduate Scholar National Institute of General Medical Sciences
2001 – 2005	Meyerhoff Undergraduate Scholar University of Maryland, Baltimore County
2004	Summer Undergraduate Research Fellowship Pfizer

Peer-Reviewed Publications

1) Erickson BK, **Kinde I**, Dobbin ZC, Wang Y, Martin JY, Alvarez RD, Conner MG, Huh WK, Roden RB, Kinzler KW, Papadopoulos N, Vogelstein B, Diaz LA Jr, Landen CN Jr. Detection of Somatic TP53 Mutations in Tampons of Patients With High-Grade Serous Ovarian Cancer. *Obstetrics and Gynecology* 124, 881-885 (2014).

2) Bettegowda C*, Sausen M*, Leary RJ*, **Kinde I***, Wang Y, Agrawal N, Bartlett B, Wang H, Luber B, Alani R, Antonarakis E, Azad N, Bardelli A, Brem H, Cameron J, Clarence L, Fecher L, Gallia GL, Gibbs P, Le D, Giuntoli R, Goggins M, Hogarty M, Holdhoff M, Hong SM, Jiao Y, Juhl H, Kim J, Siravegna G, Laheru D, Lauricella C, Lim M, Lipson E, Marie SK, Netto G J, Oliner K, Olivi A, Olsson L, Riggins GJ, Sartore-Bianchi A, Schmidt K, Shih I, Shinjo SM, Siena S, Theodorescu D, Tie J, Harkins T, Veronese S, Wang T, Weingart J, Wolfgang C, Wood LD, Xing D, Hruban RH, Wu J, Allen P, Schmidt M, Choti MA, Velculescu VE, Kinzler KW, Vogelstein B, Papadopoulos N, Diaz LA Jr. Detection of circulating tumor DNA in early and late stage human malignancies. *Science Translational Medicine* 6, 224ra24 (2014).

3) Joseph CG, Darrah E, Shah AA, Skora AD, Casciola-Rosen LA, Wigley FM, Boin F, Fava A, Thoburn C, **Kinde I**, Jiao Y, Papadopoulos N, Kinzler KW, Vogelstein B, Rosen A. Association of the Autoimmune Disease Scleroderma with an Immunologic Response to Cancer. *Science* 343, 152-157 (2014).

4) Joseph CG, Hwang H, Jiao Y, Wood LD, **Kinde I**, Wu J, Mandahl N, Luo J, Hruban RH, Diaz LA Jr, He TC, Vogelstein B, Kinzler KW, Mertens F, Papadopoulos N. Exomic analysis of myxoid liposarcomas, synovial sarcomas, and osteosarcomas. *Genes Chromosomes Cancer* 53, 15-24 (2014).

5) **Kinde I**, Munari E, Faraj SF, Hruban RH, Schoenberg M, Bivalacqua T, Allaf M, Springer S, Wang Y, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N, Netto GJ. TERT Promoter Mutations Occur Early in Urothelial Neoplasia and Are Biomarkers of Early Disease and Disease Recurrence in Urine. *Cancer Research* 73, 7162-7167 (2013).

6) **Kinde I***, Bettegowda C*, Wang Y*, Wu J, Agrawal N, Shih IeM, Kurman R, Dao F, Levine DA, Giuntoli R, Roden R, Eshleman JR, Carvalho JP, Marie SK, Papadopoulos N, Kinzler KW, Vogelstein B, Diaz LA Jr. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Science Translational Medicine* 5, 167ra4 (2013).

7) Leary RJ*, Sausen M*, **Kinde I***, Papadopoulos N, Carpten JD, Craig D, O'Shaughnessy J, Kinzler KW, Parmigiani G, Vogelstein B, Diaz LA Jr, Velculescu VE. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Science Translational Medicine* 4, 162ra154 (2012).

8) **Kinde I**, Papadopoulos N, Kinzler KW, Vogelstein B. FAST-SeqS: a simple and efficient method for the detection of aneuploidy by massively parallel sequencing. *PLoS One* 7, e41162 (2012).

9) Diaz LA Jr, Williams RT, Wu J, **Kinde I**, Hecht JR, Berlin J, Allen B, Bozic I, Reiter JG, Nowak MA, Kinzler KW, Oliner KS, Vogelstein B. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* 486, 537-540 (2012).

10) **Kinde I**, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences of the USA* 7, 9530-9535 (2011).

11) Leary RJ, **Kinde I**, Diehl F, Schmidt K, Clouser C, Duncan C, Antipova A, Lee C, McKernan K, De La Vega FM, Kinzler KW, Vogelstein B, Diaz LA Jr, Velculescu VE. Development of personalized tumor biomarkers using massively parallel sequencing. *Science Translational Medicine* 2, 20ra14 (2010).

12) Kelley BN, Kyere S, **Kinde I**, Tang C, Howard BR, Robinson H, Sundquist WI, Summers MF, Hill CP. Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein. *Journal of Molecular Biology* 373, 355-366 (2007).

13) Tang C, Loeliger E, Luncsford P, **Kinde I**, Beckett D, Summers MF. Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proceedings of the National Academy of Sciences of the USA* 101, 517-522 (2004).

14) Tang C, Loeliger E, **Kinde I**, Kyere S, Mayo K, Barklis E, Sun Y, Huang M, Summers MF. Antiviral inhibition of the HIV-1 capsid protein. *Journal of Molecular Biology* 327, 1013-1020 (2003).

Patents

1) Kinde I, Vogelstein B, Kinzler K, Papadopoulos, N. Safe sequencing system. *Patent pending.*

2) Kinde I, Vogelstein B, Kinzler K, Papadopoulos, N. Rapid aneuploidy detection. *Patent pending.*

3) Kinde I, Vogelstein B, Hoang M, Kinzler K, Papadopoulos N. Reducing errors in massively parallel sequencing by the joint analysis of both strands. *Patent pending.*

4) Kinde I, Velculescu V, Diaz L, Leary R, Parmigiani G, Sausen M, Vogelstein B. Direct Detection of Chromosomal Alterations in the Circulation of Cancer Patients Using Whole Genome Sequencing. *Patent pending.*

5) Kinde I, Vogelstein B, Bettegowda C, Diaz L, Kinzler K, Papadopoulos N, Wang Y. Evaluation of DNA from the Papanicolaou Test to Detect Ovarian and Endometrial Cancers. *Patent pending.*

6) Kinde I, Vogelstein B, Diaz L, Hruban R, Kinzler K, Netto G, Papadopoulos N. TERT Promoter Mutations for the Early Detection and Monitoring of Bladder Cancers. *Patent pending.*

Presentations & Posters

2014 Time=Lives: The Race to the Starting Line
Type: Invited lecture
Host: Partnering for Cures, a *FasterCures* event
Location: New York, NY

2014 Next Generation DNA Diagnostics
Type: Invited lecture
Host: Molecular Diagnostics Laboratory, Johns Hopkins University
Location: Baltimore, MD

2014 Showcasing the Bioeconomy: The Future is Now
Type: Invited panelist
Host: United States Department of State
Location: Washington, DC

2006 Antiviral Inhibition of the HIV-1 Capsid Protein
Type: Selected poster
Host: Student National Medical Association (SNMA) Annual Conference
Location: Atlanta, GA

2004 Entropic Switch Regulates Myristate Exposure in the HIV-1 Matrix Protein
Type: Selected lecture
Host: Annual Biomedical Research Conference for Minority Students (ABRCMS)
Location: Dallas, TX

2003 Antiviral Inhibition of the HIV-1 Capsid Protein
Type: Selected lecture
Host: Annual Biomedical Research Conference for Minority Students (ABRCMS)
Location: San Diego, CA

Leadership & Service

2013 Science speaker and lab tour guide for local youth of St. Ignatius Academy
Ludwig Center at Johns Hopkins University School of Medicine

2005 – 2009 Founder and tutor, EXCEL Tutoring Program
Morning Star Baptist Church (Catonsville, MD)

2008 – 2009 Student member, Medical School Admissions Committee
Johns Hopkins University School of Medicine

2006 – 2007 Student member, Liaison Committee on Medical Education Accreditation Panel
Johns Hopkins University School of Medicine

2005 Volunteer HIV counselor
Baltimore City Health Department

2005 Instructor, Community Adolescent Sexual Education Program (CASE)
Paul Lawrence Dunbar Middle School (Baltimore, MD)

2004 – 2005	Chapter president, Golden Key International Honour Society University of Maryland, Baltimore County
2002 – 2003	Student member, Institutional Review Board University of Maryland, Baltimore County
2002 – 2005	Tutor, Chemistry Tutorial Center University of Maryland, Baltimore County
2001	Volunteer, Shock Trauma Center University of Maryland Medical System
2001	Tutor, Learning Resources Center University of Maryland, Baltimore County