Clinical Chemistry 58:3
580–589 (2012)



# COLD-PCR Enrichment of Rare Cancer Mutations prior to Targeted Amplicon Resequencing

Coren A. Milbury,[1] Mick Correll,[2] John Quackenbush,[2] Renee Rubio,[2] and G. Mike Makrigiorgos[1,3]*

BACKGROUND: Despite widespread interest in next-generation sequencing (NGS), the adoption of personalized clinical genomics and mutation profiling of cancer specimens is lagging, in part because of technical limitations. Tumors are genetically heterogeneous and often contain normal/stromal cells, features that lead to low-abundance somatic mutations that generate ambiguous results or reside below NGS detection limits, thus hindering the clinical sensitivity/specificity standards of mutation calling. We applied COLD-PCR (coamplification at lower denaturation temperature PCR), a PCR methodology that selectively enriches variants, to improve the detection of unknown mutations before NGS-based amplicon resequencing.

METHODS: We used both COLD-PCR and conventional PCR (for comparison) to amplify serially diluted mutation-containing cell-line DNA diluted into wild-type DNA, as well as DNA from lung adenocarcinoma and colorectal cancer samples. After amplification of *TP53* (tumor protein p53), *KRAS* (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), *IDH1* [isocitrate dehydrogenase 1 ($NADP^+$), soluble], and *EGFR* (epidermal growth factor receptor) gene regions, PCR products were pooled for library preparation, bar-coded, and sequenced on the Illumina HiSeq 2000.

RESULTS: In agreement with recent findings, sequencing errors by conventional targeted-amplicon approaches dictated a mutation-detection limit of approximately 1%–2%. Conversely, COLD-PCR amplicons enriched mutations above the error-related noise, enabling reliable identification of mutation abundances of approximately 0.04%. Sequencing depth was not a large factor in the identification of COLD-PCR–enriched mutations. For the clinical samples, several missense mutations were not called with conventional amplicons, yet they were clearly detectable with COLD-PCR amplicons. Tumor heterogeneity for the *TP53* gene was apparent.

CONCLUSIONS: As cancer care shifts toward personalized intervention based on each patient's unique genetic abnormalities and tumor genome, we anticipate that COLD-PCR combined with NGS will elucidate the role of mutations in tumor progression, enabling NGS-based analysis of diverse clinical specimens within clinical practice.

© 2011 American Association for Clinical Chemistry

Rapidly evolving sequencing technologies have empowered enormous growth in the breadth and depth of cancer genome characterization. Second-generation massively parallel sequencing approaches are increasing the throughput and decreasing the costs of nucleotide-resolution oncogenomics, thereby making the characterization of entire transcriptomes, exomes, and genomes readily achievable *(1, 2)*. Currently, the pace of acquisition of genomic data for cancer patients far outstrips the utility of that information for choosing specific therapeutic avenues for individualized patient care *(1)*. Targeted amplicon resequencing is an alternative that balances the amount of information obtained, affordability, and the ability to include mutation profiling of the most meaningful genes *(3–5)*. The opportunity provided by clinical genomics is unique, because for a growing number of tumor types, clinical decision-making for patients with diagnosed cancers will increasingly be driven by the status of mutated cancer genes *(1)*. Whether these new approaches will affect routine clinical practice and the treatment of disease is no longer debatable, but how that will happen is a source of ongoing speculation and development *(1)*.

There are both conceptual and technical challenges to assessing the wealth of information for individual tumors obtained with next-generation sequencing (NGS)[4] technologies *(1)*. For example, the clinical importance of the minor alleles frequently encountered in key cancer genes and often appearing at abun-

[1] Division of DNA Repair and Genome Stability, Department of Radiation Oncology; [2] Center for Cancer Computational Biology; and [3] Division of Medical Physics and Biophysics, Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA.
* Address correspondence to this author at: Dana-Farber/Brigham and Women's Cancer Center, Brigham and Women's Hospital, Level L2, Radiation Therapy, 75 Francis St., Boston, MA 02115. Fax 617-525-7123; e-mail mmakrigiorgos@lroc.harvard.edu.
Received September 20, 2011; accepted November 17, 2011.
Previously published online at DOI: 10.1373/clinchem.2011.176198
[4] Nonstandard abbreviations: NGS, next-generation sequencing; COLD-PCR, coamplification at lower denaturation temperature PCR; TL, lung tumor; CT, colorectal tumor; $T_c$, critical denaturation temperature; $T_m$, amplicon melting temperature.



Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

Personalis EX2020

dances of <10% in the tumor cell population appears to be case dependent. Low-abundance clones containing *TET2*[5] (tet methylcytosine dioxygenase 2) mutations in chronic myelomonocytic leukemia patients confer no prognostic value *(6)*. In contrast, detection of low-level *KRAS* (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) mutations in metastatic colorectal cancer enhances the prediction of resistance to treatment with an anti–epidermal growth factor receptor monoclonal antibody *(7)*. Mutations in 1%–5% of cells in primary breast tumors can be found at prevalent (clonal) status in the secondary metastasis, a finding consistent with the mutations having obtained "driver" status in the microenvironment of the metastatic site *(8)*. Clearly, a fraction of the low-abundance genetic alterations constitutes precious clinical information that one should be able to capture.

Before we can identify which low-abundance DNA variants revealed by NGS are clinically meaningful, the question of the confidence in the generated data must be addressed. Although NGS has been demonstrated to be reliable for DNA with high-prevalence tumor somatic mutations *(5, 8, 9)*, the required depth of sequence interrogation remains problematic *(10)*, and the detection of low-prevalence somatic mutations at levels below approximately 2%–5% in tumors with heterogeneity, in tumors with stromal contamination, or in bodily fluids, is fraught with false positives, irrespective of the coverage *(4, 11)*. Because physicians are not likely to make clinical decisions on the basis of DNA-mutation signals that are close to background levels, it is important to develop procedures that enable signals arising from low-abundance mutations to be separated from the background noise and thereby boost the confidence in NGS results.

We evaluated the use of COLD-PCR (coamplification at lower denaturation temperature PCR), a newly developed methodology from our laboratory *(12, 13)*, to enhance mutation detection via massively parallel sequencing with the Illumina HiSeq 2000 analyzer. This approach enabled genuine low-abundance mutations to be magnified by enrichment before NGS-based amplicon resequencing, thereby enabling a clear distinction of mutations from the background sequencing noise. The combination of COLD-PCR with NGS improved the detection limit of targeted amplicon resequencing of low-abundance mutations by 50- to 100-fold.

[5] Human genes: *TET2*, tet methylcytosine dioxygenase 2; *KRAS*, v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog; *TP53*, tumor suppressor protein 53; *IDH1*, isocitrate dehydrogenase subunit 1 ($NADP^+$), soluble; *EGFR*, epidermal growth factor receptor; *KIF1C*, kinesin family member 1C; *USP28*, ubiquitin specific peptidase 28; *KIT*, v-kit Hardy–Zuckerman 4 feline sarcoma viral oncogene homolog.

## Materials and Methods

### DNA TEMPLATE AND SERIAL DILUTIONS OF MUTANTS

Human cell line DNA was obtained from the ATCC and the Dana-Farber Cancer Institute (Boston, MA). Frozen tissue was obtained from clinical glioblastoma, lung, and colon tumor specimens according to the approval from the Internal Review Board. Genomic DNA was isolated with the DNeasy™ Blood & Tissue Kit (Qiagen) according to manufacturer instructions. DNA quality and concentration were measured with a NanoDrop 1000 spectrophotometer (Thermo Scientific). Data for all evaluated DNA are presented in Table 1 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol58/issue3.

Genomic DNA from cell lines and 1 previously evaluated glioblastoma sample *(14)* was serially diluted into human male wild-type genomic DNA (Human Genomic DNA: Male; Promega) to generate the following preamplification mutant DNA abundances: 2%, 1%, 0.5%, 0.2%, 0.1%, 0.05%, 0.02%, and 0%. In each PCR reaction, 100 ng of genomic DNA was used to ensure an efficient representation of minor alleles.

Clinical lung tumor (TL) and colorectal tumor (CT) samples containing naturally occurring medium- and low-level somatic mutations that had previously been documented by at least 1 independent method in addition to COLD-PCR *(12, 15–17)* were analyzed in parallel with their paired, putatively normal samples, which were obtained from tumor margins and proximal regions during surgery. Naturally occurring mutation abundances in the clinical samples varied from <1% to heterozygous status.

### TARGET AMPLICON REGIONS

Nine regions in 4 frequently mutated oncogenes were evaluated: exons 5–10 of the *TP53* (tumor suppressor protein 53) gene, exon 2 of the *KRAS* gene, exon 4 of the *IDH1* [isocitrate dehydrogenase subunit 1 ($NADP^+$), soluble] gene, and exon 20 of the *EGFR* (epidermal growth factor receptor) gene. The locations of the amplicons and their primers are presented in Table 2 in the online Data Supplement.

### AMPLIFICATION STRATEGIES

COLD-PCR is a recently developed PCR-based approach for enriching low-abundance DNA mutations and minor allele variants *(12)*. COLD-PCR enriches unknown mutations at any position within the amplicon through the use of a critical denaturation temperature ($T_c$) during the PCR. The $T_c$ is lower than stan-



Personalis EX2020

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

dard denaturation temperatures and preferentially denatures heteroduplexed molecules (those formed by hybridization of mutant and wild-type sequences) and amplicons possessing mutations that lower the amplicon melting temperature ($T_m$), such as G:C>A:T or G:C>T:A. Minor-allele enrichment by COLD-PCR has been demonstrated in combination with several downstream approaches, such as Sanger sequencing, denaturing HPLC/Surveyor, MALDI-TOF, pyrosequencing, real-time TaqMan, single-strand conformation polymorphism, PCR-based mutation-specific restriction enzyme digestion, and high-resolution melting analyses *(12, 14, 15, 18–21)*. The combination of COLD-PCR and NGS and the impact of the former on NGS have not previously been reported.

The details of the thermocycling conditions and amplification protocols we have used are presented in Table 3 in the online Data Supplement. We evaluated 2 COLD-PCR platforms: fast COLD-PCR and ice-COLD-PCR. COLD-PCR amplifications used a $T_c$ to preferentially denature and enrich allelic variants containing a lower $T_m$, as previously described *(13–17, 22–26)*. PCR reactions were performed on SmartCycler II thermocyclers (Cepheid). The incidence of PCR errors was reduced by performing all reactions (25-μL final volume) with Phusion™ polymerase (New England Biolabs), which possesses a very high replication fidelity. Conventional and COLD-PCR reactions were performed with 1× manufacturer-supplied HF Buffer (New England Biolabs), 0.2 mmol/L of each deoxynucleoside triphosphate, 0.3 μmol/L primers, 1.0× LCGreen® Plus+ dye (Idaho Technologies), and 0.02 U/μL Phusion polymerase. Ice-COLD-PCR reactions were performed with the 1× manufacturer-supplied HF buffer, 0.2 mmol/L of each deoxynucleoside triphosphate, 0.9 μmol/L primers, 1.0× LCGreen Plus+ dye, 0.02 U/μL Phusion polymerase, and 25 nmol/L reference sequence oligonucleotide.

#### SANGER SEQUENCING CONFIRMATION

PCR products from the wild type and from samples with preamplification mutation abundances of 1% and 0.2% were digested with exonuclease I (New England Biolabs) and shrimp alkaline phosphatase (Affymetrix). Products were processed for Sanger sequencing at the Molecular Biology Core Facility of the Dana-Farber Cancer Institute (for primer sequences, see Table 2 in the online Data Supplement).

#### AMPLICON LIBRARY PREPARATION FOR ILLUMINA NGS

Amplicons were purified with the QIAquick™ PCR Purification Kit (Qiagen) and quantified on a NanoDrop 1000 spectrophotometer. Purified PCR products were pooled in equivalent concentrations across the serially diluted mutant DNA abundances. Final amplicon mixtures (approximately 1–2 μg) were precipitated in ethanol and 0.3 mol/L sodium acetate, washed in 700 mL/L ethanol, dried, and resuspended in 30 μL water.

Library preparation for paired-end NGS on the Illumina HiSeq 2000 instrument was performed at the Center for Cancer Computational Biology at the Dana-Farber Cancer Institute. PCR products underwent end repair and A-tailing according to outlined protocols (Illumina). Products were purified with the Agencourt AMPure® XP Bead system (Beckman Coulter Genomics). Paired-end adaptors (TruSeq Sample Prep Kit; Illumina) were ligated to the products after the multiplex paired-end protocols, as outlined. Ligation products were purified via the AMPure system, and target bands (220–270 bp) were collected and purified (MinElute Gel Extraction Kit; Qiagen). Phusion polymerase was used in the PCR to enrich the adapter-modified products. In each case, 12 bar codes were multiplexed per pool of amplicons. Library validation for quality and concentration was performed on the Bioanalyzer (Agilent) before immobilization on the flow cell and sequencing on the HiSeq 2000 instrument. Products were paired-end sequenced with 100 cycles and an index read.

#### DATA ANALYSIS

Primary analysis, including base calling, read filtering, and demultiplexing were performed according to the standard Illumina processing pipeline (CASAVA 1.7.1). Sequence read pairs were mapped independently to the human genome assembly GRh37/hg19 (build 37.2, February 2009) with Bowtie *(27)*. We allowed up to 3 mismatches across the entire length of the read and reported only reads that were uniquely aligned (−v3 −ml). SAMtools were used to calculate read depth and nucleotide frequencies for each position of the amplicons *(28)*, and custom PERL scripts were used to calculate the observed frequency of each nucleotide at each base position and to compile the results across samples into an integrated report.

## Results

#### MUTATION DETECTION IN SERIALLY DILUTED MUTATION-CONTAINING DNA

*Sanger sequencing confirmation*. Before library preparation for NGS, amplicons generated from the 1%, 0.2%, and 0% (of wild type) preamplification mutation abundances were analyzed via Sanger sequencing (for estimates of the resulting mutation abundances, see Table 4 in the online Data Supplement). As anticipated, the samples with 1% mutation abundances amplified with conventional-PCR amplicons fell below the analytical sensitivity of Sanger sequencing and could not be detected. After COLD-PCR enrichment, however,



Personalis EX2020

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

mutations were pronounced and clearly evident in the samples with 1% preamplification abundances of each amplified region. The 1% preamplification abundance demonstrated an overall mean enrichment of approximately 57-fold (SD, 11-fold), whereas the 0.2% preamplification abundance demonstrated a mean enrichment of approximately 146-fold (SD, 72-fold). The mutation enrichment for the 0.2% preamplification mutation abundance varies among the amplicons and potentially illustrates the lower limits of detection via COLD-PCR and Sanger sequencing analysis.

*Targeted amplicon resequencing: serial dilutions*. Amplicon regions produced by both conventional and COLD-PCR were sequenced (paired-end) on the Illumina HiSeq 2000 and aligned to the reference genome (GRh37/hg19). Frequency calls were generated for each nucleotide aligned within the amplicon locations. Frequency calls for all discordant nucleotides (relative to the wild-type sequence) were plotted for each nucleotide position of each amplicon. We therefore developed "variant-and-noise plots" to display both the frequencies of mutation calls and the background signals. For clarity, the wild-type sequence calls were not plotted but were reflected by the presentation of sequence depth relative to the nucleotide frequencies of the variant alleles. The overall sequence depth is presented in each plot and is defined by the right-hand *y* axis and denoted by the blue plotted line. Fig. 1 presents representative variant-and-noise plots for the serial-dilution study of *TP53* exon 10. After conventional PCR, the exon 10 mutation had a limit of detection of about 2%, in view of the maximum sequencing "noise." In contrast, the enriched mutation after COLD-PCR was evident down to a preamplification abundance of 0.02%, despite the noise. The noise was defined by the maximum observed aberrant calls, such that an error would not be mistaken for a mutation or variant. In another example, variant-and-noise plots were presented after ice-COLD-PCR (see Fig. 1 in the online Data Supplement). Ice-COLD-PCR involves a more elaborate PCR cycling protocol than fast COLD-PCR but can enrich for all possible mutations. Genomic DNA from 2 cell lines with *TP53* exon 8 mutations (HCC1008, $T_m$-equivalent mutation G>C; PFSK-1, $T_m$-increasing mutation T>G) were mixed, and both mutations were simultaneously evaluated in serial dilutions with wild-type DNA. The mutation enrichment enabled by ice-COLD-PCR allowed reliable detection down to a preamplification abundance of 0.2%. Table 5 in the online Data Supplement summarizes the details of the results for the serially diluted mutated DNA. The observed nucleotide frequency at the mutation position being evaluated is presented for each of the preamplification mutation abundances. Table 5 in the online Data Supplement also presents the maximum observed background noise along the amplicon sequence (determined by the maximum observed frequency of an aberrant nucleotide call), the mean noise (the mean observed frequency of aberrant calls within the amplicon), and the sequence depth (the number of aligned paired-end sequence reads). Serially diluted mutation abundances of <2% were not possible to discern in most cases with the conventional-PCR amplicons. Sequencing COLD-PCR amplicons, however, demonstrated a median detection limit of a 0.04% mutation abundance, with an overall range of 0.02%–0.2%. Accordingly, the ability to detect the mutations was improved by a mean of 50-fold after COLD-PCR.

The observed frequencies of the mutated nucleotide after amplification by conventional PCR and COLD-PCR were plotted and compared. Representative plots of mutation nucleotide frequency are presented in Fig. 2 in the online Data Supplement. Observed nucleotide frequencies after conventional PCR were consistent with the prepared serial dilutions, whereas observed mutation frequencies after COLD-PCR reflected the achieved enrichment. In some cases, such as with *TP53* exons 7 and 10, preamplification mutation abundances of <1% were enriched to >50% (see Table 4 in the online Data Supplement), which represents mutation enrichments by COLD-PCR of up to 300-fold.

The influence of amplification method and sequence-interrogation depth on noise was also assessed. From the wild-type replicates, we evaluated the mean background noise of each amplicon and plotted it against sequence depth (see Fig. 3 in the online Data Supplement). We calculated a mean noise estimate across all amplicons of 0.08% (SD, 0.03%) for the conventional-PCR amplicons and 0.15% (SD, 0.06%) for the COLD-PCR amplicons. We observed no decrease in noise with increasing depth of interrogation, indicating that the observed noise was due not to sampling error but rather to sequencing errors or upstream preparation (polymerase misincorporations). Indeed, higher noise was associated with COLD-PCR amplification compared with conventional PCR, possibly owing to polymerase errors generated and enriched throughout COLD-PCR amplification.

A comparison of the data in Figs. 2 and 3 in the online Data Supplement indicates that genuine mutations are enriched by COLD-PCR much more than polymerase errors and that their enriched abundances overcome the sequencing errors. Accordingly, the overall signal-to-noise ratio increased sharply after COLD-PCR.

### MUTATION DETECTION VIA TARGETED AMPLICON RESEQUENCING: CLINICAL SAMPLES

Lung adenocarcinoma and colorectal cancer clinical samples with previously demonstrated *(12, 15–17)*



Personalis EX2020

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

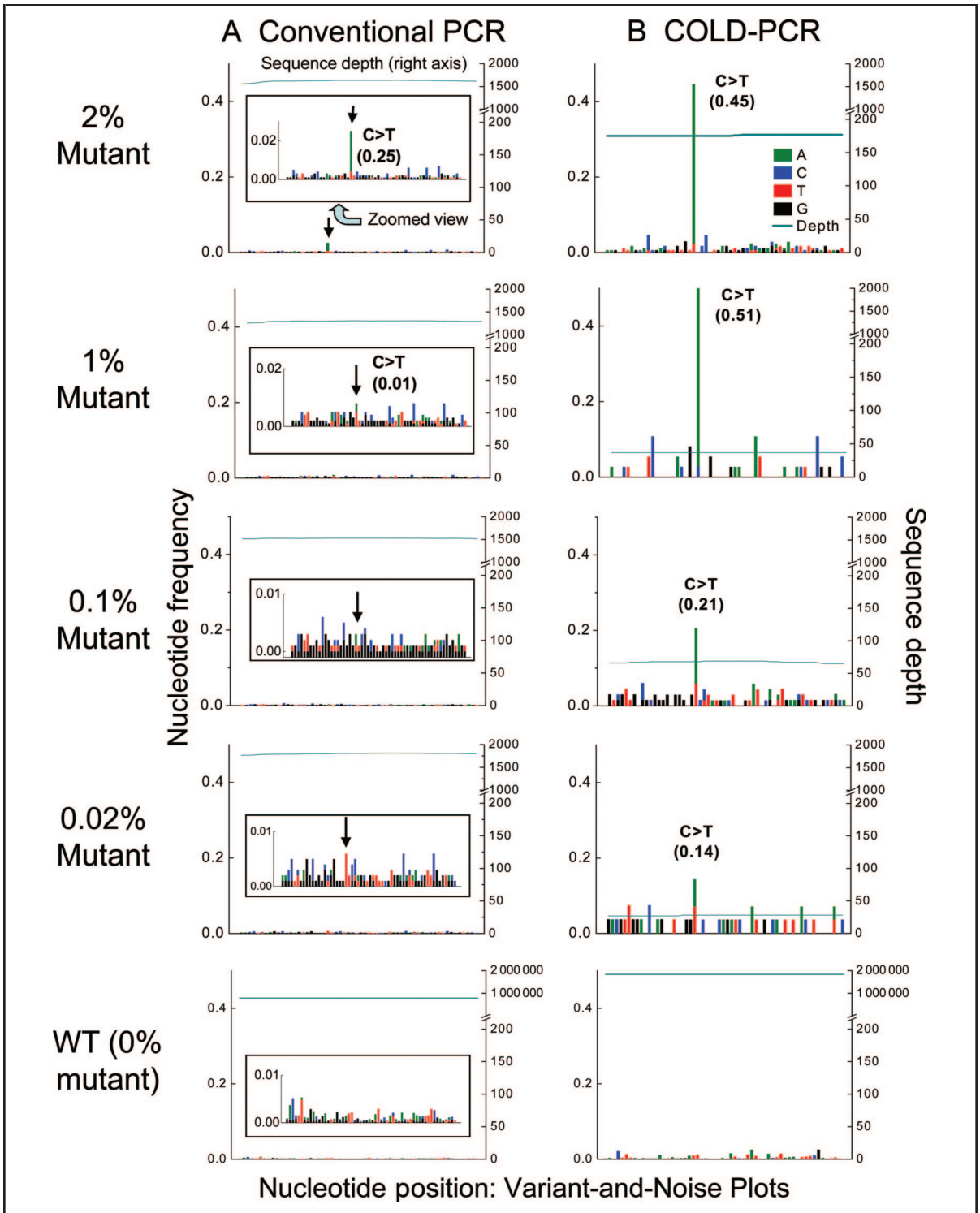


**Fig. 1. Variant/noise frequency plots (antisense strand orientation) for *TP53* exon 10 mutant serial dilutions, as amplified by conventional PCR (A) or COLD-PCR (B).**

The UACC-893 mutation (c.1024C>T) cannot be called in conventional-PCR amplicons at mutation abundances of ≤1% because of background noise. In contrast, after COLD-PCR the preamplification mutational abundance of 0.02% is enriched to approximately 14% and is detectable. Sequence read depth (right *y* axis) is presented as a blue line. WT, wild type.



Personalis EX2020

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

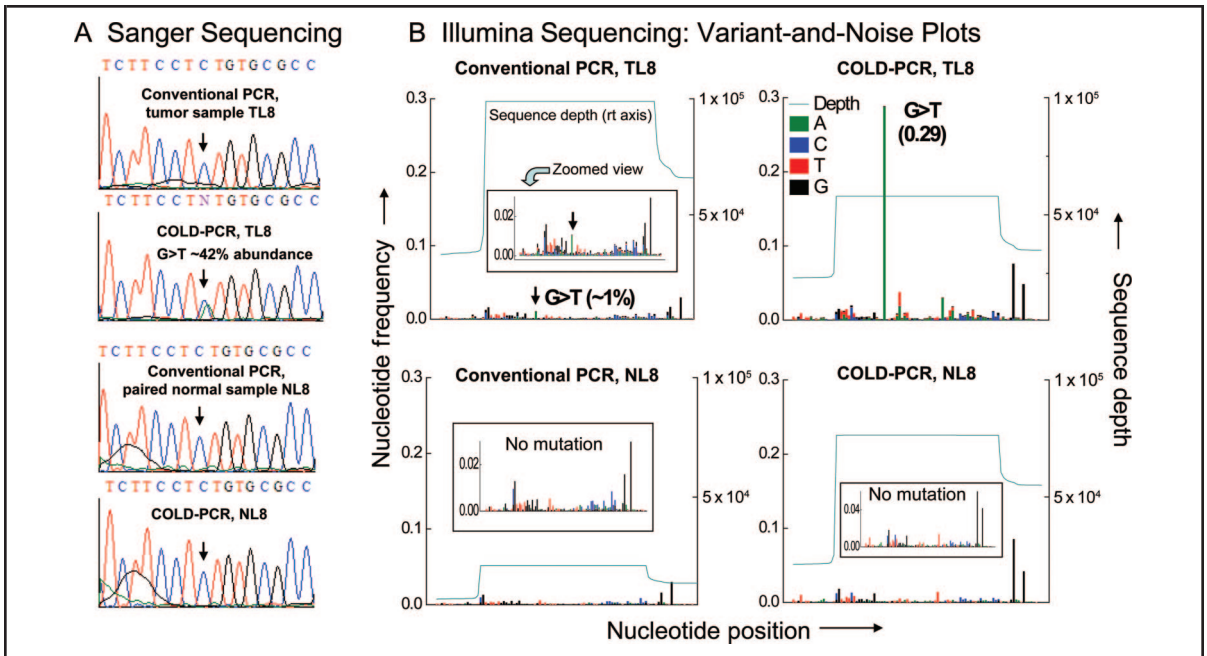


**Fig. 2. Mutation analysis in tumor sample TL8 (*TP53*, c.853G>T) and matched normal sample NL8.**

(A), Antisense Sanger sequences of conventional-PCR or COLD-PCR amplicons, including an estimate of abundance after enrichment. (B), In NGS variant-and-noise frequency plots (antisense), the TL8 mutation at approximately 1% cannot be called in conventional amplicons because of the background noise; however, COLD-PCR has appreciably enriched the mutation (approximately 29%) above the noise.

low-abundance mutations were selected for the present study. Putatively normal samples, which were sampled from the tumor margin and proximal regions, were examined in parallel. We used the Illumina HiSeq 2000 to generate variant-and-noise plots for these samples. For some samples, naturally occurring low-abundance mutations and heterogeneity were identified in COLD-PCR amplicons, although they were not detectable by conventional PCR–based NGS analysis. For example, lung adenocarcinoma sample TL8 contained a missense mutation of <1% abundance that could not be called when conventional-PCR amplicons were used (Fig. 2). Conversely, this mutation was enriched by COLD-PCR to almost 30% abundance and was easily detectable via both Illumina and Sanger sequencing (Fig. 2). Similarly, colorectal cancer sample CT20 contained 3 mutations within *TP53*: 2 low-level mutations in exons 8 and 9 and a heterozygous mutation in exon 5 (Fig. 3). Whereas the heterozygous mutation in exon 5 was clearly evident in conventional-PCR amplicons, the 2 low-level mutations were borderline detectable (approximately 3% abundance in exon 9) to nondetectable (exon 8). COLD-PCR amplification, however, enriched the 3% exon 9 mutation to nearly 76% and enriched the exon 8 mutation to >50% (Fig. 3). Widespread intratumoral heterogeneity was evident in sample CT20. Furthermore, analysis of the matched, putatively normal sample after enrichment by COLD-PCR revealed the *TP53* exon 9 mutation. The final enriched abundance was just 14%, which was below the limit of detection via Sanger sequencing and therefore not previously detected. Given the simultaneous collection of the matched samples, it is possible that the matched normal sample was collected from a tumor margin that incidentally harbored the same mutations found in the tumor, albeit at very low abundance.

Further demonstration of the improved mutation-detection limit with COLD-PCR is presented for clinical samples TL6, CT2, TL121, TL22, and TL119 (Table 1; see Figs. 4–8, respectively, in the online Data Supplement). Sequencing revealed a common *(29, 30)* yet previously undetected mutation in sample TL22 (*TP53* exon 5, c.527G>T) (Table 1; see Fig. 7 in the online Data Supplement) in addition to the 2 previously documented mutations. Because 2 rounds of COLD-PCR enriched this mutation to approximately 10%, it remained below the detection limit of COLD-PCR/Sanger sequencing, and thus remained undetected in our previous analyses *(15)*. The improved detection capability enabled by the Illumina se-



Personalis EX2020

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

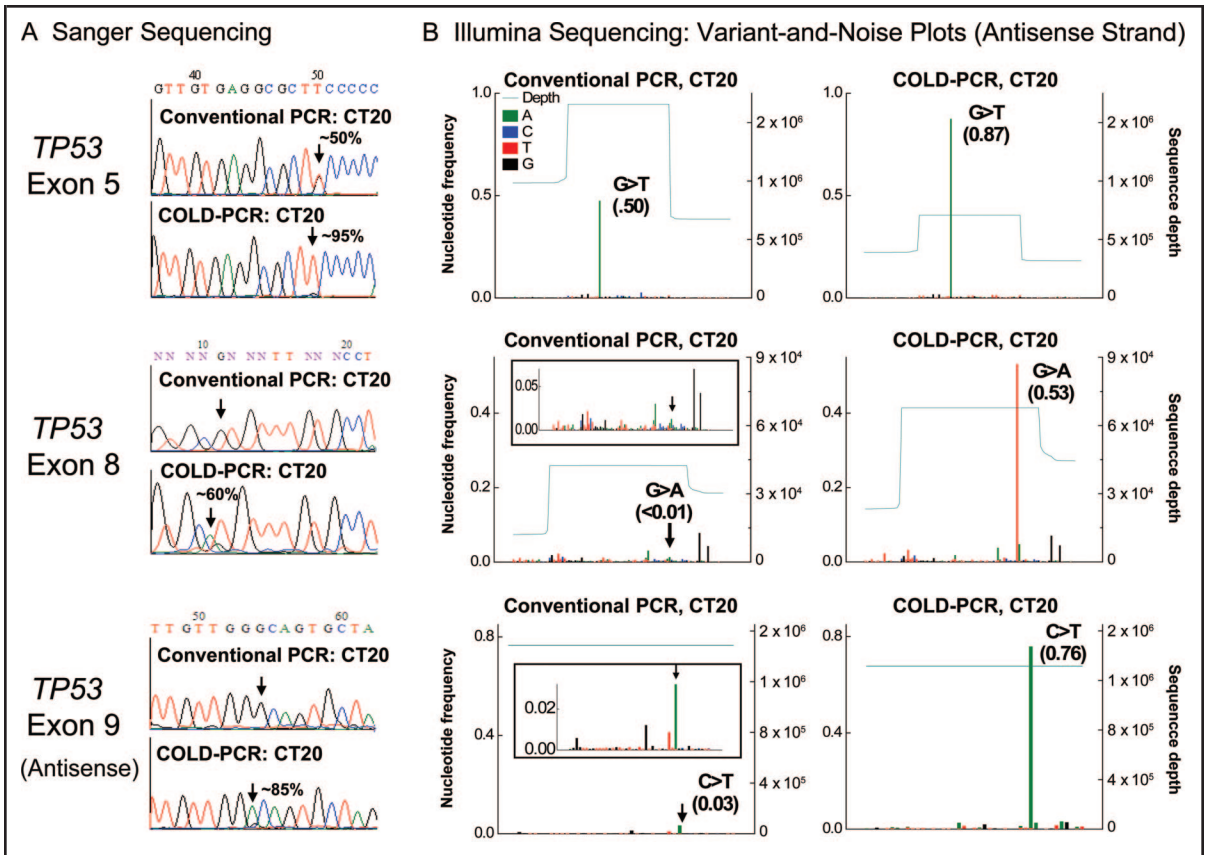


**Fig. 3. Mutational analysis of 3 *TP53* exons for sample CT20.**

Sanger sequencing (A) and conventional-PCR or COLD-PCR NGS variant-and-noise frequency plots (B). A heterozygous mutation (exon 5, c.527G>T) is detectable in both amplicons. The mutation in exon 8 (c.818G>A) occurring at <1% and the mutation in exon 9 (c.925C>T) occurring at approximately 3% are not clear in conventional-PCR amplicons; however, COLD-PCR enriched the exon 8 and exon 9 mutations to approximately 53% and 76%, respectively.

quencing platform allowed identification of this variant. In the majority of the reads (99.6%), the mutated alleles appeared in different allelic strands, indicating the presence of high heterogeneity. As in sample CT20, the presence of the same mutations in the putatively normal matched sample (NL22), albeit at a lower frequency, as in the corresponding tumor sample potentially indicates the effect of the tumor margin within the tissue sample. For each sample we evaluated, this combination of COLD-PCR enrichment with Illumina HiSeq 2000 sequencing allowed the low-abundance mutations to be detected with confidence.

## Discussion

Interest in the clinical application of NGS to individual cancer samples for personalized treatment/guidance, prognosis, and therapy follow-up is burgeoning *(1, 31, 32)*. Clinical tumor samples, however, often come in forms that challenge the technical limits of NGS-based molecular diagnostics. Such samples may be: genetically heterogeneous tumors with subclones of widely differing clinical impact; infiltrating, diffuse-type tumor samples; suboptimally microdissected tumor samples; DNA from circulating nucleic acids, circulating cells, sputum, or other bodily fluids; and samples of tumor margins. The excessive concentrations of wild-type cells and DNA in such specimens hinder the reliable identification of low-level tumor mutations, which can have profound clinical implications on disease progression, the development of metastasis, the choice of treatment, or early-detection strategies *(1)*.

In agreement with previous reports of studies that used amplicon-based NGS *(4, 6, 33–35)*, we found that the current limit of mutation detection for NGS-based sequencing analysis was limited to a mutation abundance of approximately 1%–2%, primarily owing to



Personalis EX2020

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

**Table 1. Analysis of clinical samples with Illumina amplicon resequencing, after conventional PCR and after COLD-PCR amplification.[a]**

| | | | | | Observed mutational abundance in tumor | | Observed mutational abundance in putative normal tissue | |
|---|---|---|---|---|---|---|---|---|
| Sample | Gene | Exon | Protein change | Mutation | Conventional PCR | COLD-PCR | Conventional PCR | COLD-PCR |
| CT2 | *TP53* | 5 | p.Arg175Ser | c.523C>A | Not detected[b] | 28% | Not detected | Not detected |
| CT2 | *TP53* | 7 | p.Asn247Ile | c.739A>T | 50% | 62% | Not detected | Not detected |
| CT20 | *TP53* | 5 | p.Cys176Phe | c.527G>T | 50% | 87% | Not detected | Not detected |
| CT20 | *TP53* | 8 | p.Arg273His | c.818G>A | 1%[b] | 53% | Not detected | Not detected |
| CT20 | *TP53* | 9 | p.Pro309Ser | c.925C>T | 3% | 76% | Not detected[b] | 14% |
| TL6 | *TP53* | 8 | p.Cys277Phe | c.830G>T | 2%[b] | 22% | Not detected | Not detected |
| TL8 | *TP53* | 8 | p.Glu285X | c.853G>T | 1%[b] | 29% | Not detected | Not detected |
| TL22 | *TP53* | 5 | p.Val157Phe | c.469G>T | 15% | 54% | Not detected[b] | 6% |
| TL22 | *TP53* | 5 | p.Arg158Leu | c.473G>T | 5%[b] | 25% | Not detected[b] | 17% |
| TL22 | *TP53* | 5 | p.Cys176Phe | c.527G>T | 3%[b,c] | 10% | Not detected | Not detected |
| TL64 | *TP53* | 8 | p.Arg273His | c.818G>A | 13% | 69% | Not detected | Not detected |
| TL71 | *KRAS* | 2 | p.Gly12Cys | c.34G>T | 17% | 84% | Not detected | Not detected |
| TL96 | *TP53* | 7 | p.Arg249Ser | c.747G>T | 10% | 49% | Not detected | Not detected |
| TL119 | *KRAS* | 2 | p.Gly12Phe | c.34_35GG>TT | 1[b] | 67% | Not detected | Not detected |
| TL119 | *TP53* | 7 | p.Gly244Cys | c.730G>T | 21% | 61% | Not detected | Not detected |
| TL121 | *TP53* | 8 | p.Arg273His | c.818G>A | Not detected[b] | 18% | Not detected | Not detected |
| TL121 | *TP53* | 7 | p.Gly245Ser | c.733G>A | 6% | 58% | Not detected | Not detected |
| TL135 | *TP53* | 6 | p.Val216Met | c.646G>A | 29% | 75% | Not detected | Not detected |

[a] Observed mutational abundances are presented.
[b] Indicates when a mutation could not be detected or reliably scored in the conventional PCR yet could be clearly identified in the COLD-PCR.
[c] A previously undocumented mutation was detected in TL22 because of the increased sensitivity of the COLD-PCR/NGS protocol.

the error-related background noise. Such errors are typically caused by the library-preparation steps, the sequencing reaction, and the processing of sequence calls. This error-related noise can vary across sequence regions because of such factors as priming efficiency, sequence composition, and reagent or DNA quality. As our evidence shows, some inherent variability among the regions analyzed occurs in both conventional-PCR and COLD-PCR amplicons. Only routine application and evaluation of this approach will define the typically observed noise and detection sensitivities in a clinical setting; however, if NGS is to be widely adopted in the clinical pathology arena, it is imperative to ensure the reliable detection of the low-abundance genetic variants that have profound clinical significance. For example, *KIF1C* (kinesin family member 1C) and *USP28* (ubiquitin specific peptidase 28) mutations, which are clonally expanded in metastasis, preexist at levels of ≤1% in primary breast tumors *(8)*, and clinical resistance–causing *KIT* (v-kit Hardy–Zuckerman 4 feline sarcoma viral oncogene homolog) (gastrointestinal stromal tumor) and *EGFR* (lung adenocarcinoma) mutations can be present in tumors at levels <<1% *(36, 37)*. The assessment of crucial mutations down to a mutation abundance of 0.01% can be important for accurately assessing biomarker mutations throughout disease progression *(33, 38)*.

Importantly, increasing sequencing depth, which unavoidably affects NGS throughput, is not required for obtaining reliable results when COLD-PCR is used. Enriching low-abundance mutations via COLD-PCR enables NGS mutation-detection limits as low as 0.02% with just 28 aligned reads (Fig. 1). Despite widespread perceptions to the contrary *(5)*, increasing the number of reads generally does not improve the signal-to-noise ratio or lower the detection limit for either COLD-PCR or conventional PCR (see Fig. 3 in the online Data Supplement).

The enrichment obtained via COLD-PCR varies and depends on the specific mutation and the sequence context; thus, quantification of the original mutation frequency on the basis of only a COLD-PCR result is difficult. The data presented in Fig. 2 in the online Data Supplement does suggest, however, that once a mutation above the noise level has been identified via



Personalis EX2020

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023

COLD-PCR, the frequency of the same mutation as identified via conventional PCR is a good approximation of the true frequency of the mutation in the interrogated sample. Indeed, although conventional PCR amplicons with mutation frequencies of 0.02%–1% cannot be discriminated from the noise (e.g., Fig. 1), these frequencies may still be assessed correctly (see conventional-PCR curves in Fig. 2 of the online Data Supplement). Therefore, performing parallel sequencing of both conventional amplicons and COLD-PCR amplicons may enable mutation identification and quantification.

Within this preliminary evaluation of the combination of COLD-PCR and NGS, we have applied fairly basic analyses that focus solely on read alignments, with sufficient stringency applied to avoid errors introduced by mismapped reads and to calculate the resulting nucleotide frequencies at each position so as to assess the baseline error rate of Illumina sequencing. We present raw data without extensive bioinformatics processing, and we acknowledge that more-sophisticated methods could further reduce some sources of error but that the analysis would still benefit from an improved signal-to-noise ratio. Therefore, any systematic errors that have occurred are presented in the plotted nucleotide calls. It is evident that some artifacts of the paired-end sequencing process are reproducible (such as in *TP53* exon 8; see Fig. 6 in the online Data Supplement and Fig. 3), in which increased G calls are observed at the end of the sequence read for both conventional-PCR and COLD-PCR amplifications. Owing to the nature of the nucleotide ($T_m$-increasing), it is unlikely that this error is an artifact of COLD-PCR. Fast COLD-PCR enriches only $T_m$-reducing variants or errors. When systematic artifacts occur, computational methods can be applied to remove them. Several methods are available for removing error from sequence calls to attempt to increase the sensitivity of base calling for low-abundance variants *(33, 39–42)*. A few recently developed approaches *(33, 43–45)* apply upstream sample processing that reduces sequencing errors and improves the identification of true variants with sophisticated target–bar-coding approaches followed by downstream computational analyses.

In the described approach, we observed that the mutational status after COLD-PCR exceeded errors and artifacts appreciably and enriched genuine low-abundance mutations, thereby allowing accurate detection of clinically relevant mutations without the need for extensive data manipulation. Repeated analyses of multiple amplicon regions with this enrichment approach are necessary to define routine sequencing-depth requirements and to define both the upper limits of noise and the lower limits of mutation-calling sensitivity. In the future, we envision a combination of computational algorithms, bar-coding approaches, and enrichment provided by COLD-PCR that will vastly increase the power and accuracy of mutational profiling of diverse clinical cancer specimens via NGS.

One topic currently being debated is how one can verify subsets of mutation data derived with NGS technologies *(1)*. As in the early days of microarray development, when conflicting data required resorting to the use of real-time PCR for validation of uncertain calls, it is desirable to have an accepted validation method for mutations. Because the combination of COLD-PCR with Sanger sequencing brings the limits of the previously accepted "gold standard" for mutation identification down to 0.2% from the current mutation abundance of 20%, COLD-PCR/Sanger sequencing provides a straightforward approach to address this question.

**Author Contributions:** *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

**Authors' Disclosures or Potential Conflicts of Interest:** *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

**Employment or Leadership:** None declared.
**Consultant or Advisory Role:** None declared.
**Stock Ownership:** None declared.
**Honoraria:** None declared.
**Research Funding:** C.A. Milbury, National Cancer Institute training grant T32-CA009078; J. Quackenbush, Dana-Farber Cancer Institute, NIH (grant no. 1U01CA151118), and the National Science Foundation; G.M. Makrigiorgos, NIH departmental funds, JCRT Foundation, and Innovative Molecular Analysis Technologies Program of the National Cancer Institute (grant nos. R21-CA-151164 and R21-CA-151164).
**Expert Testimony:** None declared.

**Role of Sponsor:** The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

**Acknowledgments:** We thank Fieda Abderazzaq and Howie Goodell for their assistance in sample processing and bioinformatics analysis.

**Disclaimer:** The contents of this report do not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023



Personalis EX2020

## References


1. Taylor BS, Ladanyi M. Clinical cancer genomics: How soon is now? J Pathol 2011;223:319–27.
2. Shendure J, Ji H. Next-generation DNA sequencing. Nature Biotechnol 2008;26:1135–45.
3. Zito CI, Riches D, Kolmakova J, Simons J, Egholm M, Stern DF. Direct resequencing of the complete ERBB2 coding sequence reveals an absence of activating mutations in ERBB2 amplified breast cancer. Genes Chromosomes Cancer 2008;47: 633–8.
4. Qin W, Kozlowski P, Taillon B, Bouffard P, Holmes A, Janne P, et al. Ultra deep sequencing detects a low rate of mosaic mutations in tuberous sclerosis complex. Hum Genet 2010;127:573–82.
5. Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T, Yuza Y, et al. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat Med 2006;12:852–5.
6. Smith AE, Mohamedali AM, Kulasekararaj A, Lim Z, Gäken J, Lea NC, et al. Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value. Blood 2010;116:3923–32.
7. Molinari F, Felicioni L, Buscarino M, De Dosso S, Buttitta F, Malatesta S, et al. Increased detection sensitivity for KRAS mutations enhances the prediction of anti-EGFR monoclonal antibody resistance in metastatic colorectal cancer. Clin Cancer Res 2011;17:4901–14.
8. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 2009;461:809–13.
9. Reis-Filho J. Next-generation sequencing. Breast Cancer Res 2009;11(Suppl 3):S12.
10. Gray J. Cancer: genomics of metastasis. Nature 2010;464:989–90.
11. Tang S, Huang T. Characterization of mitochondrial heteroplasmy using a parallel sequencing system. Biotechniques 2010;48:287–96.
12. Li J, Wang L, Mamon H, Kulke MH, Berbeco R, Makrigiorgos GM. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. Nat Med 2008; 14:579–84.
13. Milbury CA, Li J, Makrigiorgos GM. Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations. Nucleic Acids Res 2011;39:e2.
14. Milbury CA, Chen CC, Mamon H, Liu P, Santagata S, Makrigiorgos GM. Multiplex amplification coupled with COLD-PCR and high resolution melting enables identification of low-prevalence mutations in cancer samples with low DNA content. J Mol Diagn 2011;13:220–32.
15. Li J, Milbury CA, Li C, Makrigiorgos GM. Two-round coamplification at lower denaturation temperature-PCR (COLD-PCR)-based Sanger sequencing identifies a novel spectrum of low-level mutations in lung adenocarcinoma. Hum Mutat 2009;30:1583–90.
16. Milbury CA, Li J, Makrigiorgos GM. COLD-PCR-enhanced high-resolution melting enables rapid and selective identification of low-level unknown mutations. Clin Chem 2009;55:2130–43.
17. Song C, Milbury CA, Li J, Liu P, Zhao M, Makrigiorgos GM. Rapid and sensitive detection of *KRAS* mutation after COLD-PCR enrichment and high resolution melting analysis. Diagn Mol Pathol 2011;20:81–9.
18. Li J, Wang L, Janne PA, Makrigiorgos GM. Coamplification at lower denaturation temperature-PCR increases mutation-detection selectivity of TaqMan-based real-time PCR. Clin Chem 2009; 55:748–56.
19. Maekawa M, Taniguchi T, Hamada E, Takeshita A. Efficiency of COLD-PCR for enrichment of K-ras mutation; a proof by use of SSCP analysis [Abstract]. Clin Chem 2009;55(Suppl):A219.
20. Delaney D, Diss TC, Presneau N, Hing S, Berisha F, Idowu BD, et al. GNAS1 mutations occur more commonly than previously thought in intramuscular myxoma. Mod Pathol 2009;22:718–24.
21. Kristensen LS, Daugaard IL, Christensen M, Hamilton-Dutoit S, Hager H, Hansen LL. Increased sensitivity of *KRAS* mutation detection by high-resolution melting analysis of COLD-PCR products. Hum Mutat 2010;31:1366–73.
22. Milbury CA, Li J, Liu P, Makrigiorgos GM. COLD-PCR: improving the sensitivity of molecular diagnostics assays. Expert Rev Mol Diagn 2011;11: 159–69.
23. Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53 database: new online mutation analysis and recommendations to users. Hum Mutat 2002;19:607–14.
24. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science 2006;314:268–74.
25. Soussi T, Lozano G. p53 mutation heterogeneity in cancer. Biochem Biophys Res Commun 2005; 331:834–42.
26. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. Nature 2010;465:473–7.
27. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 2009;10:R25.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25: 2078–9.
29. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, Olivier M. Impact of mutant p53 functional properties on *TP53* mutation patterns and tumor phenotype: lessons from recent developments in the IARC *TP53* database. Hum Mutat 2007;28:622–9.
30. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011;39:D945–50.
31. Robison K. Application of second-generation sequencing to cancer genomics. Brief Bioinform 2010;11:524–34.
32. Cronin M, Ross JS. Comprehensive next-generation cancer genome sequencing in the era of targeted therapy and personalized oncology. Biomark Med 2011;5:293–305.
33. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 2011;108:9530–5.
34. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods 2008;5:1005–10.
35. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature 2010;464:610–4.
36. Engelman JA, Mukohara T, Zejnullahu K, Lifshits E, Borras AM, Gale CM, et al. Allelic dilution obscures detection of a biologically significant resistance mutation in EGFR-amplified lung cancer. J Clin Invest 2006;116:2695–706.
37. Wakita S, Yamaguchi H, Miyake K, Mitamura Y, Kosaka F, Dan K, Inokuchi K. Importance of c-kit mutation detection method sensitivity in prognostic analyses of t(8;21)(q22;q22) acute myeloid leukemia. Leukemia 2011;25:1423–32.
38. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med 2008;14:985–90.
39. Erlich Y, Mitra PP, delaBastide M, McCombie WR, Hannon GJ. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. Nat Methods 2008;5:679–82.
40. Druley TE, Vallania FLM, Wegner DJ, Varley KE, Knowles OL, Bonds JA, et al. Quantification of rare allelic variants from pooled genomic DNA. Nat Methods 2009;6:263–5.
41. Kircher M, Stenzel U, Kelso J. Improved base calling for the Illumina genome analyzer using machine learning strategies. Genome Biol 2009; 10:R83.
42. Zagordi O, Geyrhofer L, Roth V, Beerenwinkel N. Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. J Comput Biol 2010;17:417–28.
43. Miner BE, Stöger RJ, Burden AF, Laird CD, Hansen RS. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Res 2004;32:e135.
44. McCloskey M, Stöger R, Hansen R, Laird C. Encoding PCR products with batch-stamps and barcodes. Biochemical Genet 2007;45:761–7.
45. Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP. A method for counting PCR template molecules with application to next-generation sequencing. Nucleic Acids Res 2011;39:e81.


Downloaded from https://academic.oup.com/clinchem/article/58/3/580/5620563 by guest on 17 September 2023



Personalis EX2020