



Published in final edited form as:

Clin Chem. 2012 March ; 58(3): 580–589. doi:10.1373/clinchem.2011.176198.

ENABLING CLINICAL CANCER GENOMICS FOR RARE MUTATIONS: COLD-PCR MAGNIFIES MUTATIONS PRIOR TO TARGETED AMPLICON RE-SEQUENCING

Coren A. Milbury¹, Mick Correll³, John Quackenbush³, Renee Rubio³, and G. Mike Makrigiorgos^{1,2,*}

¹Division of DNA Repair and Genome Stability, Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA

²Division of Medical Physics and Biophysics, Department of Radiation Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA

³Center for Cancer Computational Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA

Abstract

Despite widespread interest in the application of next-generation-sequencing (NGS) to the mutation profiling of individual cancer specimens, the onset of personalized clinical genomics is currently stalled due in part to technical hurdles. As tumors are genetically-heterogeneous and often mixed with normal/stromal cells, the resulting low-abundance DNA somatic mutations often produce ambiguous results or fall below the current NGS detection limit, thus hindering mutation calling that abides to clinical sensitivity/specificity standards. Here we examine the feasibility of applying COLD-PCR, a form of PCR that magnifies selectively the mutations, to boost the detection of unknown rare somatic mutations prior to applying NGS-based amplicon re-sequencing to clinical samples. We amplified DNA from serially-diluted mutation-containing human cell-lines into wild-type (WT) DNA, as well as lung adenocarcinoma and colorectal cancer specimens using COLD-PCR or conventional PCR for comparison. Following individual amplification of *TP53*, *KRAS*, *IDH1*, and *EGFR* regions, PCR products were barcoded, pooled for library preparation and sequenced on the Illumina-HiSeq2000 platform. Regardless of sequencing depth, sequencing errors dictated a mutation-detection limit of ~1–2% mutation abundance in conventional PCR amplicons analyzed by NGS. In contrast, COLD-PCR amplicons enabled genuine mutations to exceed the sequence noise levels, thus allowing reliable identification of

*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 Street, Boston, MA 02115. mmakrigiorgos@iroc.harvard.edu.

CONFLICT OF INTEREST STATEMENT: COLD-PCR is a technology owned by the Dana Farber Cancer Institute and partly licenced to Transgenomic Inc. The manuscript has not been published previously and is not being considered concurrently by another journal. All authors and acknowledged contributors have read and approved the manuscript.

This is an un-copyedited authored manuscript copyrighted by the American Association for Clinical Chemistry (AACC). This may not be duplicated or reproduced, other than for personal use or within the rule of 'Fair Use of Copyrighted Materials' (section 107, Title 17, U.S. Code) without permission of the copyright owner, AACC. The AACC disclaims any responsibility or liability for errors or omissions in this version of the manuscript or in any version derived from it by the National Institutes of Health or other parties. The final publisher-authenticated version of the article will be made available at <http://www.clinchem.org> 12 months after its publication in *Clinical Chemistry*.

STATEMENT OF AUTHOR CONTRIBUTIONS

CAM and MM developed the project design and analyzed data. CAM carried out COLD-PCR and conventional PCR amplifications. MC and JQ contributed to study design and sequence data analysis. RR advised and carried out experiments related to library preparation and sample sequencing. All authors were involved in writing the paper and had final approval of the submitted and published versions.

mutation abundances of ~0.04%. Sequencing depth was not a significant factor in the identification of COLD-PCR-magnified mutations. The analyzed clinical specimens revealed several *TP53* and *KRAS* missense mutations that could not be called following NGS of conventional amplicons, yet were clearly detectable in COLD-PCR amplicons. Extensive tumor heterogeneity in the *TP53* gene was revealed in some samples. As cancer care shifts toward personalized intervention, based on the unique genetic abnormalities in each patient's tumor genome, we anticipate that COLD-PCR-NGS will elucidate the role of rare mutations in tumors, enable NGS-based analysis of diverse clinical specimens and the broad inter-phasing of NGS with clinical practice.

Keywords

COLD-PCR; mutation enrichment; low-abundance mutations; next generation sequencing; cancer

INTRODUCTION

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 cost of nucleotide resolution oncogenomics, making whole transcriptomes, exomes and genomes readily achievable [1,2]. At present, the pace of acquisition of genomic data in cancer patients far outstrips the utility of that information to oncologists in choosing specific therapeutic avenues for their patients in most tumor types [1]. Focused amplicon re-sequencing is another alternative that provides a balance between the amount of information obtained, affordability, and the ability to include mutation profiling of the most meaningful genes [3–5]. The opportunity afforded by clinical genomics is unique, as clinical decision-making for patients diagnosed with a growing number of tumor types will increasingly be driven by the status of mutant cancer genes [1]. Whether these new approaches will impact routine clinical practice and the treatment of disease is no longer debatable, but how precisely this will happen is a source of ongoing speculation and development [1].

Yet, as comprehensively described in a recent review by Taylor and Ladanyi [1] there are both conceptual and technical challenges on how to assess the wealth of information obtained using next generation sequencing (NGS) technologies for individual tumors. For example, the majority of validated somatic mutations are unique, private mutations specific to the sequenced cancer genome [1]. While most can be passenger mutations in the classical sense, it is not likely that they are universally functionally insignificant in that patient's cancer, i.e. some of these mutations can be driving the individual tumor. Further, the clinical significance of the frequently encountered minor alleles in key cancer genes, often appearing at abundances of <10% in the tumor cell population appears to be case-dependent. Low-abundance *TET2* mutant clones in chronic myelomonocytic leukemia patients confers no prognostic value [6]. In contrast, detection of low-level *KRAS* mutations in metastatic colorectal cancer (CRC) enhances the prediction of anti-EGFR (epidermal growth factor receptor) monoclonal antibody resistance [7]. Mutations in 1–5% of cells in primary breast tumors can be found at prevalent (clonal) status in the secondary metastasis consistent with obtaining 'driver' status in the micro-environment of the metastatic site [8]. Clearly a fraction of the low-abundance genetic alterations encompass precious clinical information that one should be able to capture.

Before we can even begin identifying which low-abundance DNA variations revealed by NGS are clinically meaningful, the question of confidence in the generated data must be addressed. While reliable NGS for DNA with high-prevalence tumor somatic mutations has

been demonstrated [3,8,9], the required depth of sequence interrogation remains problematic [10] and detection of low-prevalence somatic mutations at levels below ~2–5% in tumors with heterogeneity, stromal contamination or in bodily fluids is fraught with false-positives irrespective of coverage [5,11]. Because physicians are not likely to make clinical decisions based on DNA mutation signals that are close to background levels, it is important to develop procedures that enable separation of signals coming from low-abundance mutations versus background noise, thereby boosting the confidence in NGS results.

Herein we evaluate the use of COLD-PCR, a newly developed methodology from our laboratory [12,13] to enhance mutation detection via massively parallel sequencing using the Illumina HiSeq2000. This approach enables genuine low-abundance mutations to be magnified by enrichment prior to NGS-based amplicon re-sequencing, thereby enabling a clear distinction of mutations from background sequencing noise. COLD-PCR modifies thermocycling parameters such that minor variant alleles (such as low-abundance mutations) are selectively enriched throughout the course of PCR, by an average of two orders of magnitude (~100-fold; e.g. a 0.1% mutation becomes ~10% mutation) or more, thus greatly improving the mutation detection sensitivity of downstream technologies. We demonstrate that the combination of COLD-PCR with next generation sequencing improves the detection limit of targeted amplicon resequencing of low-abundances of mutations. This enables rare mutations in clinical lung adenocarcinoma and CRC specimens, which were previously undetectable, to be clearly revealed and provides an experimental solution to overcome the problem of mutations in genetically heterogeneous tumors.

MATERIALS AND METHODS

DNA template and mutant serial dilutions

Human cell-line DNA was obtained from the American Type Culture Collection (Manassas, VA) and Dana-Farber Cancer Institute (DFCI, Boston, MA). Frozen tissue was obtained from clinical glioblastoma, lung, and colon tumor specimens following Internal Review Board (IRB) approval. Genomic DNA was isolated using the DNeasy™ tissue kit (Qiagen Inc., Valencia, CA); quality and concentration were determined using the NanoDrop1000 (Thermo Scientific, Wilmington, DE). All evaluated DNA is presented in Supplementary Table 1.

Genomic DNA from cell-lines and one previously evaluated glioblastoma specimen [14] was serially diluted into human male wild-type genomic DNA (item# G1471; Promega Corp., Madison, WI) to generate pre-amplification mutant abundances, as follows: 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.

In a proof-of-concept evaluation, clinical lung tumor (TL) and colorectal tumor (CT) specimens samples containing naturally occurring medium and low-level somatic mutations previously documented by at least one more independent method in addition to COLD-PCR [12,15–17] were analyzed in parallel with their paired, putatively normal specimen, obtained during tumor surgery. Naturally occurring mutation abundances in the clinical specimens varied from <1% to heterozygous status.

Target amplicon regions

Nine regions in four frequently mutated oncogenes were evaluated: exons 5 through 10 of tumor suppressor protein 53 (*TP53*), exon 2 of v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*), exon 4 of isocitrate dehydrogenase subunit 1 (*IDHI*), and exon 20 of epidermal growth factor receptor (*EGFR*). Amplicon locations and primers are presented in Supplementary Table 2.

Amplification strategies

COLD-PCR (CO-amplification of major and minor alleles at Lower Denaturation temperature) is a recently-developed PCR-based approach to enrich 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 during PCR amplification. The critical denaturation temperature (T_c) is lower than standard denaturation temperatures such that it 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, dHPLC/Surveyor, MALDI-TOF, Pyrosequencing, real-time TaqMan analyses, SSCP, PCR-based mutation-specific restriction enzyme digestion, and high resolution melting [12,14,15,18–21]. The combination and impact of COLD-PCR on next generation sequencing has not been previously reported.

Detailed thermocycling conditions and amplification protocols applied herein are presented in Supplementary Table 3. Two platforms of COLD-PCR were evaluated herein, *fast*-COLD-PCR and *ice*-COLD-PCR. COLD-PCR amplifications were performed using a critical denaturation temperature (T_c) to preferentially denature and enrich allelic variants containing a lower melting temperature (T_m) as previously described [13–17,22–26]. PCR reactions were performed on SmartCycler II thermocyclers (Cepheid Inc., Sunnyvale, CA). To reduce the incidence of PCR errors, all reactions (25 μ l final volume) were performed using Phusion™ polymerase (New England Biolabs, Ipswich, MA), which possesses a very high replication fidelity. Conventional and COLD-PCR reactions were performed using 1X manufacturer-supplied HF buffer, 0.2 mM each dNTP, 0.3 μ M primers, 1.0X LCGreen+® dye, and 0.02 U/ μ l Phusion™ polymerase. *Ice*-COLD-PCR reactions were performed using 1X manufacturer-supplied HF buffer, 0.2 mM each dNTP, 0.9 μ M primers, 1.0X LCGreen+® dye, 0.02 U/ μ l Phusion™ polymerase, and 25nM reference sequence oligonucleotide.

Sanger sequencing confirmation

PCR products from wild type and from 1% and 0.2% pre-amplification mutant abundances were digested with exonuclease I (New England Biolabs) and shrimp alkaline phosphatase (Affymetrix Inc., Cleveland, OH). Products were processed for Sanger sequencing at the DFCI Molecular Biology Core Facility (primers listed in Supplementary Table 2).

Amplicon library preparation for Illumina NGS

Amplicons were purified using the QIAquick™ PCR purification kit (Qiagen Inc.) and quantified on the NanoDrop1000 (Thermo Scientific, Wilmington, DE). Purified PCR products were pooled in equivalent concentrations across the serially-diluted mutant abundances. Final amplicon mixtures (~1–2 μ g) were precipitated in ethanol and sodium acetate (0.3M), washed in 70% ethanol, dried, and resuspended in 30 μ l water.

Library preparation for paired-end next generation sequencing on the Illumina HiSeq2000 was performed at the DFCI Center for Cancer Computational Biology (CCCB). PCR products underwent end-repair and A-tailing following outlined protocols. Products were purified using the Agencourt AMPure® XP Bead system. Paired-end adaptors (TruSeq Kit, Illumina Inc., San Diego, CA) were ligated to the products following the multiplex paired-end protocols as outlined. Ligation products were purified via the AMPure system, and target bands (ranging 220–270 bp) were collected and purified (MinElute Gel Extraction Kit, Qiagen Inc.). Using Phusion polymerase, PCR was performed to enrich the Adapter-Modified products. In each case, twelve barcodes were multiplexed per pool of amplicons.

Library validation for quality and concentration was performed on the Bioanalyzer (Agilent Inc., Santa Clara CA) prior to immobilization on the flow cell and sequencing on the HiSeq2000 (Illumina Inc.). Products were paired-end sequenced with 100 cycles and an index read.

Data analysis

Primary analysis, including base calling, read filtering, and de-multiplexing were performed using the standard Illumina processing pipeline. Sequence read pairs were mapped independently to the human genome assembly GRh37/hg19 (build 37.2, Feb 2009) using Bowtie [27], allowing for up to three mismatches across the entire length of the read, and reporting only reads that were uniquely aligned ($-v3 -m1$). Read depth and nucleotide frequencies were calculated for each position of the amplicons using SAMtools [28] and custom perl scripts.

RESULTS

Mutation Detection in Serially-diluted mutation-containing DNA

Sanger Sequencing Confirmation—Prior to library preparation for next generation sequencing, amplicons generated from the 1%, 0.2%, and 0% (WT) pre-amplification mutant abundances were analyzed via Sanger sequencing; estimates of resulting mutational abundances are presented in Supplementary Table 4. As anticipated, 1% mutational abundances amplified by conventional PCR amplicons fell below the sensitivity of Sanger sequencing and could not be detected. However, after COLD-PCR enrichment, mutations were pronounced and clearly evident in the 1% pre-amplification abundances of each amplified region. The 1% pre-amplification abundance demonstrated an overall average enrichment of ~57-fold (± 11), while the 0.2% pre-amplification abundance demonstrated an average enrichment of ~146-fold (± 72). The mutational enrichment for the 0.2% pre-amplification mutant abundance is variable among the amplicons and potentially illustrates the lower limits of detection ability via COLD-PCR-Sanger sequencing analysis.

Targeted Amplicon Resequencing on the Illumina HiSeq2000: Serial dilutions

—Amplicon regions produced by both conventional and COLD-PCR were sequenced (paired-end) on the Illumina HiSeq2000 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 WT sequence, were plotted for each nucleotide position of each amplicon. As such, “variant and noise plots” were developed to display the mutant calls, in addition to the background signals. For clarity, the WT sequence calls were not plotted but are used in the calculation of sequence depth and nucleotide frequency. Representative variant and noise plots are presented for the serial dilution study for *TP53* exon 10 (Figure 1). Following conventional PCR, the exon 10 mutation has a limit of detection of about 2%, in view of the sequencing ‘noise’. In contrast, following COLD-PCR the enriched mutation is evident down to 0.02% pre-amplification abundance, despite the noise. In another example, variant and noise plots are presented following *ice*-COLD-PCR (Supplementary Figure 1). *Ice*-COLD-PCR involves a more elaborate PCR cycling than *fast*-COLD-PCR, but can enrich all possible mutations. Genomic DNA from two cell lines with *TP53* exon 8 mutations (HCC1008, T_m -equivalent mutation, G>C, and PFSK-1, T_m -increasing mutation, T>G) were mixed and both mutations were simultaneously evaluated in serial dilutions within wild type. The mutation enrichment enabled by *ice*-COLD-PCR allows for reliable detection sensitivity down to 0.2% pre-amplification abundance. Detailed results for the serially-diluted mutated DNA are summarized in Supplementary Table 5. The observed nucleotide frequency at the mutant position being evaluated is presented for each of the pre-amplification mutational

abundances. Supplementary Table 5 also presents the maximum observed background noise along the amplicon sequence, and the sequence depth. Serially-diluted mutation abundances <2% in most cases was not possible to discern in the conventional-PCR amplicons. However, sequencing COLD-PCR amplicons demonstrated a median detection limit of 0.04% mutation abundance, with an overall range of 0.02% to 0.2%. Accordingly, the ability to detect the mutations is improved by an average of 50-fold following COLD-PCR.

Observed frequencies of the mutated nucleotide after amplification by conventional PCR and COLD-PCR were plotted and compared. Representative plots of mutant nucleotide frequency are depicted in Supplementary Figure 2. Observed nucleotide frequencies following conventional PCR are consistent with the prepared serial dilutions, while observed mutation frequencies following COLD-PCR reflect the achieved enrichment. In some cases, such as *TP53* exon 7 and exon 10, pre-amplification mutation abundances <1% are enriched to greater than heterozygous status (Supplementary Table 4) representing mutation enrichments by COLD-PCR up to 300-fold.

The influence of amplification method and sequence interrogation depth on noise was also assessed. From the WT replicates, average background noise of each given amplicon was determined and plotted against sequence depth (Supplementary Figure 3). An average noise estimate, across all amplicons, was calculated at 0.08% ($\pm 0.03\%$) for the conventional-PCR amplicons, and at 0.15% ($\pm 0.06\%$) for the COLD-PCR amplicons. No decrease in noise by increase in depth of interrogation was observed, indicating that the observed noise is not due to sampling error, but rather due to sequencing errors or upstream preparation (polymerase misincorporations). Indeed, a higher noise was associated with COLD-PCR amplification relative to conventional-PCR, possibly reflecting polymerase errors generated at early COLD-PCR cycles and enriched thereafter.

Comparison of the data in Supplementary Figures 2 and 3 indicate that genuine mutations are enriched by COLD-PCR much more than polymerase errors and that their magnification also overcomes the second source of errors (sequencing errors). Accordingly, the overall signal-to-noise ratio increases sharply following COLD-PCR.

Mutation Detection via Targeted Amplicon Resequencing on the Illumina HiSeq2000: Clinical specimens

Lung adenocarcinoma and colorectal cancer clinical specimens were picked for the present evaluation study in view of having been shown previously to contain low-level mutations [12,15–17]. Using the Illumina NGS system, variant and noise plots were generated in these specimens. Naturally occurring low-abundance mutations and heterogeneity was identified in COLD-PCR amplicons, which was not detectable through NGS sequence analysis of conventional PCR amplicons. For example, lung adenocarcinoma sample TL8 contains a <1% abundance missense mutation that cannot be called using the conventional-PCR amplicons (Figure 2). Conversely, the mutation is enriched by COLD-PCR to almost 30% abundance and is easily detectable both via Illumina and Sanger sequencing (Figure 2). Similarly, colorectal cancer specimen CT20 contains three mutations within *TP53*: two low-level mutations in exons 8 and 9 and a heterozygous mutation in exon 5 (Figure 3). While the heterozygous mutation in exon 5 is clearly evident in conventional PCR amplicons, the two low-level mutations are borderline (~3% abundance in exon 9) to non-detectable (exon 8). However, after amplification by COLD-PCR, the 3% exon 9 mutation was enriched to nearly 76%, and the exon 8 mutation has been enriched to over 50% by COLD-PCR (Figure 3). Widespread intra-tumoral heterogeneity is evident in sample CT20.

Further demonstration of the improved mutation detection limit when using COLD-PCR is presented for clinical specimens TL6, CT2, TL121, TL22, and TL119 (see Table 1 and

Supplementary Figures 4–8, respectively). Sequencing revealed a common, yet previously undocumented mutation in specimen TL22 (p.Cys176Phe, c.527G>T) (Table 1, Supplementary Figure 7). As two rounds of COLD-PCR enriches this mutation to approximately 10%, it remains below the sensitivity level of COLD-PCR-Sanger sequencing, and has thus remained undetected by our previous analyses [15]. The improved sensitivity enabled by the Illumina sequencing has allowed for the identification of this variant. Thus, this combination of COLD-PCR enrichment with Illumina HiSeq2000 sequencing allowed each of the low-abundance mutations evaluated herein to be detected with confidence.

DISCUSSION

Interest in the clinical application of next generation sequencing (NGS) on individual cancer specimens for personalized treatment-guidance, prognosis, and therapy follow-up is burgeoning exponentially [1,29,30]. However, clinical tumor samples often come in form that challenges the technical limits of NGS-based molecular diagnostics: genetically heterogeneous tumors with subclones of widely different potential clinical impact; infiltrating, diffuse-type tumor specimens; sub-optimally micro-dissected tumor samples; DNA from circulating nucleic acid, circulating cells, sputum or other bodily fluids; and testing of tumor margins. The excessive concentrations of wild type cells and DNA in all these specimens disable the reliable identification of low-level tumor mutations that can have profound clinical implications on disease progression, the development of metastasis, choice of treatment, or early detection strategies [1].

In agreement with previous reports [5,6,31–33] using amplicon-based NGS we find that the current limit of mutation detection for NGS-based sequencing analysis is limited to ~1–2% mutation abundance, primarily due to the background noise generated by system errors. Such errors are typically the result of library preparation steps, the sequencing reaction, and the processing of the sequence calls. However, if NGS is to be widely adopted in the clinical pathology arena, it is imperative to ensure the reliable detection of low-abundance genetic variations that can have profound clinical significance. For example, *KIF1C* and *USP28* mutations that are found clonally expanded in metastasis pre-exist at levels 1% or less in primary breast tumors [8]; clinical resistance-causing *KIT*(GIST) and *EGFR* (lung adenocarcinoma) mutations can be present in tumors at levels significantly less than 1% [34,35]; and the assessment of crucial mutations down to 0.01% mutation abundance can be important for the accurate assessment of biomarker-mutations throughout disease progression [31,36].

Importantly, increasing sequencing depth, which unavoidably impacts NGS throughput, is not required for obtaining reliable results when using COLD-PCR. Magnifying the low-abundance mutations via COLD-PCR enables NGS mutation detection limits as low as 0.02% with just 28 aligned reads (Figure 1). In general, despite the widespread perception [3], Supplementary Figure 3 shows that increasing the number of reads does not improve the signal-to-noise ratio and lowest detection limit for either COLD-PCR or conventional PCR.

Within this preliminary evaluation of the combination of COLD-PCR-NGS, we have applied fairly basic analyses that focus solely upon read alignments, applying sufficient stringency to avoid errors introduced by mis-mapped reads, and calculation of the resulting nucleotide frequencies at each position in order to assess the baseline error rate of Illumina sequencing. Here we present raw data without extensive bioinformatic processing, acknowledging that more sophisticated methods could further reduce some sources of error, but would still however still benefit from an improved signal-to-noise ratio. As a result, any systematic errors that have occurred are presented in the plotted nucleotide calls. It is evident that some artifacts of the paired-end sequencing occur in a reproducible and

systematic manner such as in *TP53* exon 8 (Figure 3 and Supplementary Figure 6) where elevated G calls are observed at the end of the sequence read both for conventional and COLD-PCR amplification. When such artifacts occur systematically, computational methods can be applied to remove them. Several methods are available for removing error from sequence calls in attempt to increase the sensitivity of low-abundance variant base-calling [31,37–41]. Another recently developed approach [29] applies a method for upstream sample processing that reduces sequencing errors and improves the identification of true variants using a sophisticated target-barcoding approach followed by downstream computational analysis. In the approach described herein, we observed that the mutational status after COLD-PCR exceeded these errors and artifacts and identified genuine mutations down to very low abundances, allowing accurate detection of low-abundance clinically-relevant mutations without the need for extensive data manipulation (which itself can introduce its own biases). In the future, we can also envision a serial combination of computational algorithms and barcoding approaches with the enrichment provided by COLD-PCR in order to vastly increase the power and accuracy of mutational profiling of diverse clinical cancer specimens via next generation sequencing.

One hot topic currently in debate is how does one go about verifying subsets of mutation data derived by next generation sequencing technology [1]. As in the early days of microarray development, when conflicting data resorted to using real-time PCR for validation of uncertain calls, it is desirable to have an accepted validation method for mutations. Since 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 20% mutation abundance, COLD-PCR-Sanger provides a straightforward approach to address this question.

In summary, we have shown that unavoidable noise resulting from a number of sources limits the ability of NGS to detect mutations lower than about 2% abundance, and that magnification of mutations via COLD-PCR prior to NGS enhances the signals by genuine mutations and improves the detection limit to 0.04%. Thereby boosting the confidence in the detection of rare mutations in clinical samples and providing a solution to help interphase of NGS with clinical pathology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Fieda Aberazzaq and Howie Goodell from the CCCB for their assistance in sample processing and bioinformatic analysis. This work was supported by the JCRT Foundation, by T32-CA009078 from the National Cancer Institute and National Institutes of Health grant CA-111994. The contents of this manuscript are the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

References

1. Taylor BS, Ladanyi M. Clinical cancer genomics: how soon is now? *The Journal of Pathology*. 2011; 223:319–327.
2. Shendure J, Ji H. Next-generation DNA sequencing. *Nature Biotechnology*. 2008; 26:1135–1145.
3. Thomas RK, Nickerson E, Simons JF, et al. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nature Medicine*. 2006; 12:852–855.

4. Zito CI, Riches D, Kolmakova J, et al. Direct resequencing of the complete ERBB2 coding sequence reveals an absence of activating mutations in ERBB2 amplified breast cancer. *Genes Chromosomes Cancer*. 2008; 47:633–638. [PubMed: 18418848]
5. Qin W, Kozlowski P, Taillon B, et al. Ultra deep sequencing detects a low rate of mosaic mutations in tuberous sclerosis complex. *Human Genetics*. 2010; 127:573–582. [PubMed: 20165957]
6. Smith AE, Mohamedali AM, Kulasekararaj A, 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; 116:3923–3932. [PubMed: 20693430]
7. Molinari F, Felicioni L, Buscarino M, et al. Increased Detection Sensitivity for KRAS Mutations Enhances the Prediction of Anti-EGFR Monoclonal Antibody Resistance in Metastatic Colorectal Cancer. *Clinical Cancer Research*. 2011; 17:4901–4914. [PubMed: 21632860]
8. Shah SP, Morin RD, Khattra J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. 2009; 461:809–813. [PubMed: 19812674]
9. Reis-Filho J. Next-generation sequencing. *Breast Cancer Research*. 2009; 11:S12. [PubMed: 20030863]
10. Gray J. Cancer: Genomics of metastasis. *Nature*. 2010; 464:989–990. [PubMed: 20393550]
11. Tang S, Huang T. Characterization of mitochondrial heteroplasmy using a parallel sequencing system. *Biotechniques*. 2010; 48:287–296. [PubMed: 20569205]
12. Li J, Wang L, Mamon H, et al. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nature Medicine*. 2008; 14:579–584.
13. Milbury CA, Li J, Makrigiorgos GM. Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations. *Nucleic Acids Research*. 2011; 39:e2. [PubMed: 20937629]
14. Milbury CA, Chen CC, Mamon H, et al. Multiplex amplification coupled with COLD-PCR and High Resolution Melting enables identification of low-prevalence mutations in cancer samples with low DNA content. *Journal of Molecular Diagnostics*. 2011; 13:220–232. [PubMed: 21354058]
15. Li J, Milbury CA, Li C, et al. Two-round coamplification at lower denaturation temperature-PCR (COLD-PCR)-based sanger sequencing identifies a novel spectrum of low-level mutations in lung adenocarcinoma. *Human Mutation*. 2009; 30:1583–1590. [PubMed: 19760750]
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; 55:2130–2143. [PubMed: 19815609]
17. Song C, Milbury CA, Li J, et al. Rapid and sensitive detection of KRAS mutation after COLD-PCR enrichment and high resolution melting analysis. *Diagnostic Molecular Pathology*. 2011; 20:81–89. [PubMed: 21532494]
18. Li J, Wang L, Janne PA, et al. Coamplification at lower denaturation temperature-PCR increases mutation-detection selectivity of TaqMan-based real-time PCR. *Clin Chem*. 2009; 55:748–756. [PubMed: 19233916]
19. Maekawa M, Taniguchi T, Hamada E, et al. Efficiency of COLD-PCR for enrichment of K-ras mutation; a proof by use of SSCP analysis. *Clin Chem*. 2009; 55:A219–A219.
20. Delaney D, Diss TC, Presneau N, et al. GNAS1 mutations occur more commonly than previously thought in intramuscular myxoma. *Mod Pathol*. 2009
21. Kristensen LS, Daugaard IL, Christensen M, et al. Increased sensitivity of KRAS mutation detection by high-resolution melting analysis of COLD-PCR products. *Hum Mutat*. 31:1366–1373. [PubMed: 20848649]
22. Milbury CA, Li J, Liu P, et al. COLD-PCR: improving the sensitivity of molecular diagnostics assays. *Expert Review of Molecular Diagnostics*. 2011; 11:159–169. [PubMed: 21405967]
23. Olivier M, Eeles R, Hollstein M, et al. The IARC TP53 database: New online mutation analysis and recommendations to users. *Human Mutation*. 2002; 19:607–614. [PubMed: 12007217]
24. Sjoblom T, Jones S, Wood LD, et al. The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science*. 2006; 314:268–274. [PubMed: 16959974]
25. Soussi T, Lozano G. p53 mutation heterogeneity in cancer. *Biochemical and Biophysical Research Communications*. 2005; 331:834–842. [PubMed: 15865939]

26. Lee W, Jiang Z, Liu J, et al. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature*. 2010; 465:473–477. [PubMed: 20505728]
27. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 2009; 10:R25. [PubMed: 19261174]
28. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
29. Robison K. Application of second-generation sequencing to cancer genomics. *Briefings in Bioinformatics*. 2010; 11:524–534. [PubMed: 20427421]
30. Cronin M, Ross JS. Comprehensive next-generation cancer genome sequencing in the era of targeted therapy and personalized oncology. *Biomarkers in Medicine*. 2011; 5:293–305. [PubMed: 21657839]
31. Kinde I, Wu J, Papadopoulos N, et al. Detection and quantification of rare mutations with massively parallel sequencing. *Proceedings of the National Academy of Sciences*. 2011
32. Quail MA, Kozarewa I, Smith F, et al. A large genome center's improvements to the Illumina sequencing system. *Nat Meth*. 2008; 5:1005–1010.
33. He Y, Wu J, Dressman DC, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature*. 2010; 464:610–614. [PubMed: 20200521]
34. Engelman JA, Mukohara T, Zejnullahu K, et al. Allelic dilution obscures detection of a biologically significant resistance mutation in EGFR -amplified lung cancer. *J Clin Invest*. 2006
35. Wakita S, Yamaguchi H, Miyake K, et al. Importance of c-kit mutation detection method sensitivity in prognostic analyses of t(8;21)(q22;q22) acute myeloid leukemia. *Leukemia*. 2011
36. Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. *Nature Medicine*. 2008; 14:985–990.
37. Erlich Y, Mitra PP, delaBastide M, et al. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat Meth*. 2008; 5:679–682.
38. Druley TE, Vallania FLM, Wegner DJ, et al. Quantification of rare allelic variants from pooled genomic DNA. *Nat Meth*. 2009; 6:263–265.
39. Kircher M, Stenzel U, Kelso J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biology*. 2009; 10:R83. [PubMed: 19682367]
40. Zagordi O, Geyrhofer L, Roth V, et al. Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. *Journal of Computational Biology*. 2010; 17:417–428. [PubMed: 20377454]
41. Zagordi O, Klein R, Däumer M, et al. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. *Nucleic Acids Research*. 2010; 38:7400–7409. [PubMed: 20671025]

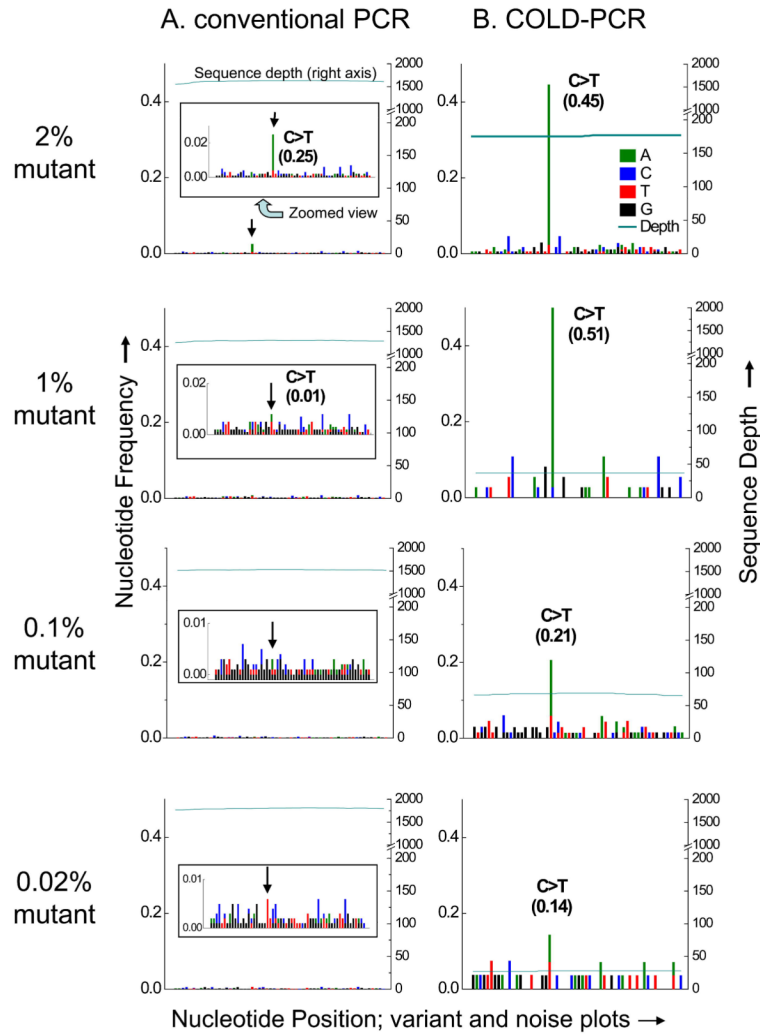


Figure 1. Variant and noise frequency plots for *TP53* exon 10 serial dilutions of mutant into wild type DNA, amplified with conventional PCR (A) or COLD-PCR (B) and followed by Illumina NGS. NGS sequence plots and alignments are presented for the anti-sense strand. The terminating mutation in human cell-line UACC-893 (c.1024C>T, p.Arg342X) is evident in the NGS sequencing calls of the conventional PCR amplicons (A) for the 2% mutational abundance, though cannot be called confidently for the 1% mutational abundance or less, due to the background noise. The inset panel provides a zoomed-in view. Conversely, it is evident that much lower abundances of mutations can be detected due to the enrichment by COLD-PCR (B). In COLD-PCR, even the 0.02% pre-amplification mutational abundance is visible above the noise as it is enriched to ~14%. Sequence depth (right axis) presents the number of sequence reads aligning to the amplicon region.

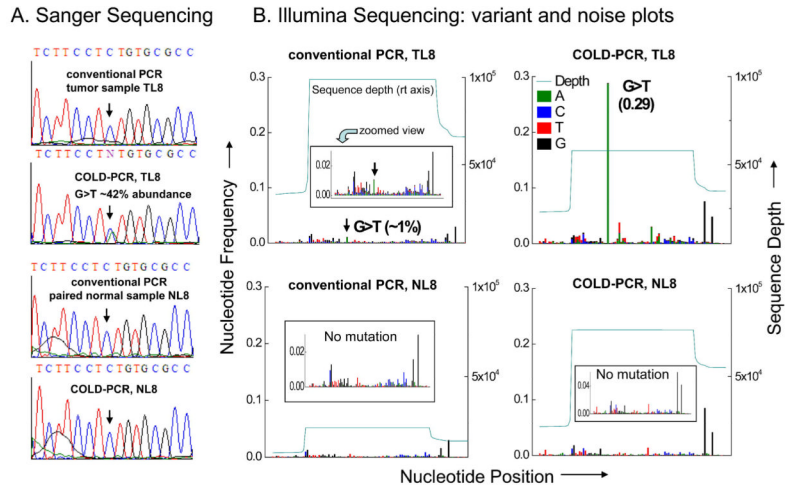


Figure 2. Detection of low-level *TP53* mutation in lung tumor specimen TL8. (A) Anti-sense strand Sanger sequence chromatograms for *TP53* exon 8 in TL8 and its matched normal specimen NL8, amplified via conventional PCR or COLD-PCR strategies. An estimate of allelic abundance is presented for the enriched mutation. (B) NGS variant and noise frequency plots (anti-sense strand alignment) are presented for conventional PCR or COLD-PCR amplicons. The ~1% terminating mutation in TL8 (c.853G>T, p.Glu285X) cannot be confidently called in the NGS sequencing calls of the conventional PCR amplicons due to the background noise; however, after COLD-PCR, the mutational abundance has been enriched to ~29% and is clearly detectable above the noise. The mutation was not detected in the matched normal specimen after amplification by either approach.

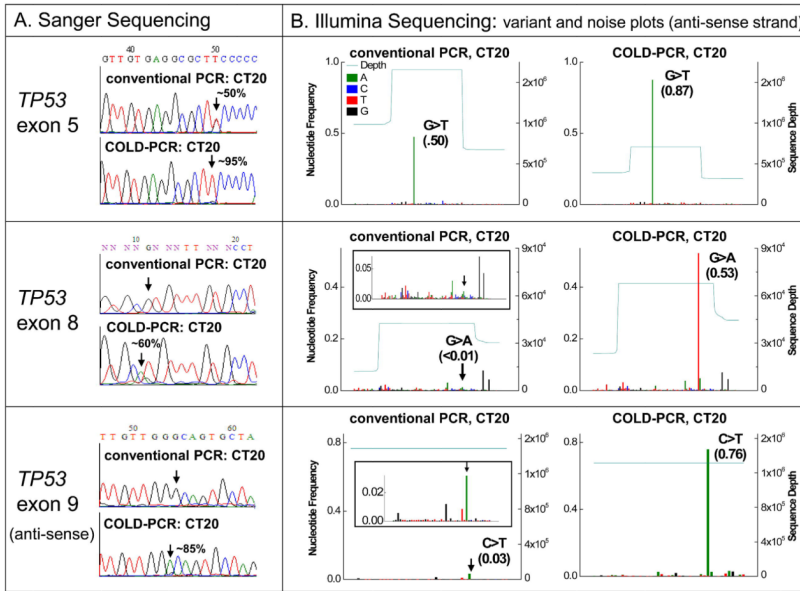


Figure 3. Detection of simultaneous high and low-level mutations in three *TP53* exons of tumor specimen CT20. (A) Sanger sequence chromatograms amplified via conventional or COLD-PCR (exons 5 and 8 sense strand; exon 9 anti-sense strand). (B) Illumina NGS variant and noise frequency plots (anti-sense strands) for NGS sequencing of conventional or COLD-PCR amplicons. Exon 5 contains a heterozygous non-synonymous mutation (c.527G>T, p.Cys176Phe) detectable after amplification by either approach. The <1% non-synonymous mutation in exon 8 (c.818G>A, p.Arg273His) cannot be detected in the sequencing calls of the conventional PCR amplicons due to the background noise; however, after COLD-PCR, the mutational abundance has been enriched to ~53%. Similarly, the ~3% non-synonymous mutation in exon 9 (c.925C>T, p.Pro309Ser) is difficult to call with confidence in the NGS sequencing calls of the conventional PCR amplicons due to the background noise; however, after COLD-PCR, the mutational abundance has been enriched to ~76%. Tumor mutational heterogeneity is evident in this specimen.

Table 1

Analysis of mutations observed in clinical tumor specimens via Illumina next generation targeted amplicon sequencing, after PCR amplification using conventional or COLD-PCR approaches. Observed mutational abundances are presented for conventional PCR and COLD-PCR approaches for each tumor specimen and their respective matched normal counterparts.

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

Specimens with an asterisk (*) indicates when a mutation could not be detected, or due to noise, could not be reliably scored in the conventional PCR, yet COLD-PCR clearly presents the mutational event.

§ As denoted, a previously undocumented mutation was detected in TL22 with the increased combined sensitivity of COLD-PCR-NGS