

Cytosine Deamination Is a Major Cause of Baseline Noise in Next-Generation Sequencing

Guoli Chen · Stacy Mosier · Christopher D. Gocke ·
Ming-Tseh Lin · James R. Eshleman

Published online: 5 August 2014
© Springer International Publishing Switzerland 2014

Abstract

Background and Objectives As next-generation sequencing (NGS) becomes a major sequencing platform in clinical diagnostic laboratories, it is critical to identify artifacts that constitute baseline noise and may interfere with detection of low-level gene mutations. This is especially critical for applications requiring ultrasensitive detection, such as molecular relapse of solid tumors and early detection of cancer. We recently observed a ~10-fold higher frequency of C:G > T:A mutations than the background noise level in both wild-type peripheral blood and formalin-fixed paraffin-embedded samples. We hypothesized that these might represent cytosine deamination events, which have been seen using other platforms.

Methods To test this hypothesis, we pretreated samples with uracil N-glycosylase (UNG). Additionally, to test whether some of the cytosine deamination might be a laboratory artifact, we simulated the heat associated with polymerase chain reaction thermocycling by subjecting samples to thermocycling in the absence of polymerase. To test the safety of universal UNG pretreatment, we tested known positive samples treated with UNG.

Results UNG pretreatment significantly reduced the frequencies of these mutations, consistent with a biologic

source of cytosine deamination. The simulated thermocycling-heated samples demonstrated significantly increased frequencies of C:G > T:A mutations without other baseline base substitutions being affected. Samples with known mutations demonstrated no decrease in our ability to detect these after treatment with UNG.

Conclusion Baseline noise during NGS is mostly due to cytosine deamination, the source of which is likely to be both biologic and an artifact of thermocycling, and it can be reduced by UNG pretreatment.

Key Points

The baseline noise in normal peripheral blood and formalin-fixed paraffin-embedded samples detected by next-generation sequencing (NGS) is dominated by C:G > T:A mutations, which are signature mutations of cytosine deamination

Consistent with this, treatment of samples with an enzyme designed to remove uracil reduced the frequencies of these mutations, suggesting that one source is biologic. It was also demonstrated that the heat of thermocycling (in the absence of polymerase) can increase the frequencies of these mutations

It was concluded that the major sources of baseline noise in NGS are both biologic and laboratory-induced cytosine deamination

G. Chen · S. Mosier · C. D. Gocke · M.-T. Lin ·
J. R. Eshleman (✉)
Department of Pathology, The Sol Goldman Pancreatic Cancer
Research Center, Johns Hopkins University School of Medicine,
Suite 344, CRB-II, 1550 Orleans Street, Baltimore, MD 21231,
USA
e-mail: jeshlem@jhmi.edu

C. D. Gocke · J. R. Eshleman
Department of Oncology, Johns Hopkins University School of
Medicine, Baltimore, MD, USA

1 Introduction

Next-generation sequencing (NGS) has been revolutionizing biomedical research and clinical practice for the past

9 years, since its introduction [1]. With substantial improvements in accuracy, read length, and depth of coverage, and remarkable reductions in costs, NGS is becoming a major sequencing platform for clinical diagnostic laboratories. In the field of oncology, NGS has allowed detection of all potential driver gene mutations that cause a particular malignancy and thus clinically reports the status of all genes known to predispose to a particular type of cancer [2–7]. To embrace the era of targeted therapy and personalized medicine, NGS will be an essential tool to evaluate all related genes that predict drug response and clinical outcome, and will offer comprehensive genetic profiling, especially if target-based treatment covers mutations in multiple gene/pathway family members [8, 9].

A promising clinical area in which NGS will likely play a role is minimal residual disease (MRD) testing of solid tumors. This will allow clinicians to initiate second- or third-line therapy more quickly on the basis of residual tumor molecules, rather than waiting until the tumor progresses radiographically [10, 11]. On the other hand, a negative MRD test could be used to avoid the risk and expense associated with additional chemotherapy, when the patient is actually in remission.

Because of its extremely high depth of coverage, NGS will likely also be the tool of choice for early detection of cancer through measurement of critical oncogene-activating mutations [12] and structural abnormalities [13, 14]. While NGS holds tremendous promise for clinical use, it also faces challenges associated with technical complexity and result interpretation. Reproducible sequence artifacts constitute the baseline noise of sequencing and may interfere with detection of true gene mutations. This is a critical concern to address when extremely low mutation frequencies need to be detected unambiguously for these and other applications.

It has been demonstrated that cytosine deamination contributes to background noise in DNA sequencing of ancient and formalin-fixed paraffin-embedded (FFPE)-treated DNA [15, 16]. Since it can manifest as either a base substitution of C to T (C > T) on the sense strand or as a G > A mutation on the sense strand (arising from a C > T deamination on the antisense strand), these mutations are collectively designated C:G > T:A. Williams and colleagues [17] observed formalin fixation-induced C:G > T:A transitions. The same artifact was also observed in freshly prepared samples in a recent study validating detection of gene mutations in an Ion AmpliSeq™ Cancer Hotspot Panel v2, using an Ion Torrent Personal Genome Machine® (PGM™) [both from Life Technologies, Carlsbad, CA, USA] in our laboratory [18]. We observed significantly higher C:G > T:A mutation frequencies than background noise levels during NGS in

both normal peripheral blood and FFPE samples, by checking all common *KRAS* [Kirsten rat sarcoma viral oncogene homolog] (codons 12 and 13), *BRAF* [B-Raf proto-oncogene, serine/threonine kinase] (V600E), and *EGFR* [epidermal growth factor receptor] (T790M and L858R) gene point mutations. These C:G > T:A mutations must be either biologic (intrinsic to the sample prior to isolation) or an artifact of the molecular biology, including DNA isolation, polymerase chain reaction (PCR) amplification, and/or sequencing.

During routine NGS, we noted an increase in C:G > T:A transitions in normal samples. In this report, we identify the etiology of NGS baseline noise, explore its contributing factors, and provide a partial solution. To confirm cytosine deamination as the source, we treated DNA from peripheral blood with uracil N-glycosylase (UNG) and demonstrated that it significantly reduced the frequencies of C:G > T:A mutations, by checking all common *KRAS* (codons 12 and 13), *BRAF* (V600E), and *EGFR* (T790M and L858R) gene point mutations. Consistent with a previous finding that prolonged heating may induce deamination of DNA [19], we found that the heat associated with thermocycling induced a significant increase in the C:G > T:A mutation frequency. We next showed that UNG pretreatment of positive control samples does not interfere with the capacity of NGS to detect real mutations. Finally, we attempted to include a thermostable UNG in PCR reactions, but we were unable to identify conditions that would allow both enzymes to work.

2 Materials and Methods

2.1 Materials

This study was conducted with institutional review board approval. The specimens consisted of four peripheral blood specimens from normal donors and seven FFPE samples from patients carrying distinctive *EGFR*, *KRAS*, and *BRAF* gene mutations. DNA was isolated as described previously [20, 21]. DNA concentrations were determined by a Qubit® 2.0 Fluorometer, and UNG enzyme 0.5 µL [1 unit/µL] (both from Life Technologies) was added into each reaction (30 ng DNA, 20 µL for total volumes) and incubated for 30 min at 50 °C prior to thermocycling for library preparation.

2.2 The NGS Platform

NGS was conducted using the Ion AmpliSeq™ Cancer Hotspot Panel v2 for targeted multigene amplification, as described in our previous study [18]. An Ion AmpliSeq™ Library Kit 2.0 was used for library preparation (PCR

thermocycling at 65 °C for 30 min and 95 °C for 2 min, 20 cycles at 95 °C for 15 s and 60 °C for 4 min, with a hold at 10 °C), with an Ion PGM™ Template OT2 200 Kit and an Ion OneTouch™ ES Instrument for emulsion PCR and enrichment, an Ion PGM™ Sequencing 200 Kit v2, Ion 318™ Chips, and the PGM™ sequencing platform for NGS [all from Life Technologies], as recommended by the manufacturers' protocols, without modification. The DNA input for targeted multigene PCR was 30 ng. Eight specimens were barcoded using Ion Xpress™ Barcode Adapters (from Life Technologies), pooled, and run on a single Ion 318™ Chip. For samples treated with extra heat, we cycled them at 99 °C for 20 min, with 40 cycles at 99 °C for 2 min and 60 °C for 4 min, and a hold at 10 °C prior to library preparation.

2.3 Data Analysis

The sequencing data were analyzed using Ion Torrent Suite™ Version 3.2.0 (from Life Technologies). The frequency (percentage) of all of the common *KRAS* (codons 12 and 13), *BRAF* (V600E), and *EGFR* (T790M and L858R) gene point mutations and those at five randomly picked base positions with nucleotide G or C in amplified regions of chromosomes 7 and 12 were calculated. The C:G > T:A mutations include *KRAS* G12D (GGT > GAT), G12S (GGT > AGT), G13D (GGC > GAC), G13S (GGC > AGC); and *EGFR* T790M (ACG > ATG). The non-C:G > T:A mutations include *KRAS* G12A (GGT > GCT), G12C (GGT > TGT), G12R (GGT > CGT), G13A (GGC > GCC), G13C (GGC > TGC), G13R (GGC > CGC); and *BRAF* V600E (GTG > GAG).

3 Results

3.1 C:G > T:A Mutation Frequencies in Peripheral Blood Are Reduced by Uracil DNA Glycosylase

In our previous study, we found that C:G > T:A mutations were significantly more common than other mutations in both peripheral blood and FFPE specimens [18]. As shown in Fig. 1a, the frequencies of C:G > T:A mutations—including *KRAS* G12D, G12S, G13D, and G13S; *EGFR* T790M; and those at five randomly picked positions in PCR amplicons (shown by *black circles*)—were significantly (about 8-fold) higher than other baseline noise levels (shown by *black triangles*) (Fig. 1a, d; $p < 0.01$). To test whether cytosine deamination had occurred prior to library construction, we treated normal peripheral blood specimens with uracil DNA glycosylase (UNG) before we conducted the initial AmpliSeq™ PCR. The glycosylase activity of UNG excises the uracil base from DNA, leaving

the sugar–phosphate backbone intact, thereby functionally removing that strand from the PCR reaction, as the polymerase cannot synthesize across the abasic site. UNG treatment reduced the frequencies of C:G > T:A mutations at all of the above sites (*black circles*), except for two of them (shown by *black arrows* in Fig. 1b, c), demonstrating that some of the C:G > T:A mutations arose from deamination of cytosine to uracil prior to PCR. The overall reduction in C:G > T:A mutation frequencies was approximately 30 % and statistically significant (Fig. 1d; $p < 0.05$), although the C:G > T:A mutation frequencies following UNG treatment were still significantly higher than those of the other mutations (Fig. 1d; $p < 0.01$). In FFPE specimens, we also observed a 22 % reduction in mutation frequencies, but without statistical significance, probably because of higher levels of variation in FFPE samples (data not shown).

3.2 Heat Associated with Thermocycling Induces Deamination

Considering a previous observation that prolonged heat could induce cytosine deamination of DNA [19], we hypothesized that the denaturation phase of thermocycling during library preparation and emulsion PCR might cumulatively cause such an effect. To test this, we thermocycled DNA from peripheral blood specimens (without performing PCR) prior to library preparation. As shown in Fig. 2, this treatment induced additional deamination effects at all susceptible positions (shown by *black circles*), except for two of them (shown by *black arrows* in Fig. 2b, c). Compared with the statistically significant increase in the overall C:G > T:A mutation frequencies caused by additional thermocycling (Fig. 2d; $p < 0.05$), the frequencies of other baseline mutations (shown by *black triangles*) at non-cytosine positions was not obviously affected by the treatment (Fig. 2d; $p > 0.05$).

3.3 UNG Does Not Interfere with Detection of Real Mutations

Since UNG reduces the frequencies of deamination-related mutations, it is a potential tool to lower NGS background noise levels. With this consideration, we wanted to exclude the possibility that UNG may reduce the ability of NGS to detect bona fide gene mutations. We selected seven positive control FFPE samples carrying a wide range of mutant allele frequencies (Table 1) of distinct *EGFR*, *KRAS*, and *BRAF* mutations. We then treated those DNA samples with UNG prior to library preparation. Compared with the results from untreated specimens, the percentages of the mutations detected after UNG treatment were consistent with the previously determined mutation frequencies.

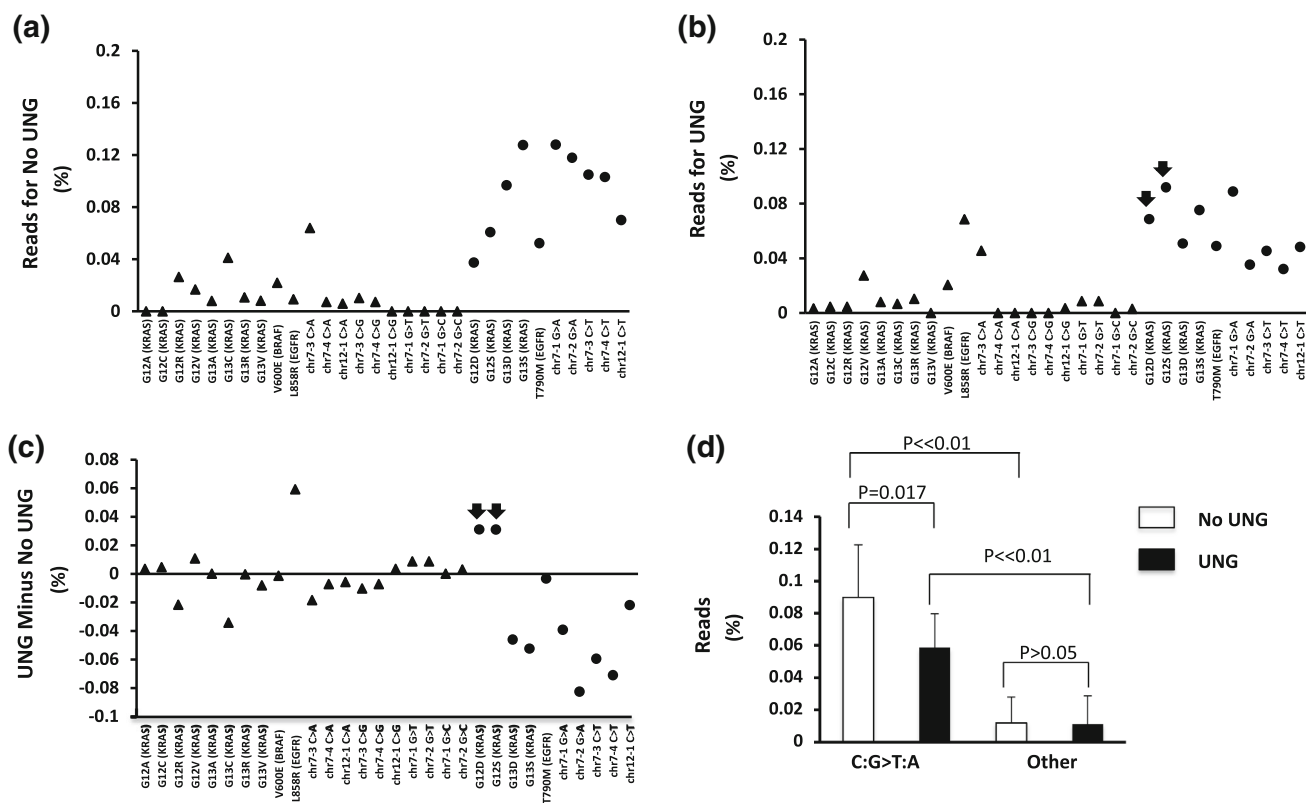


Fig. 1 Uracil N-glycosylase (UNG) reduces the frequencies of C:G > T:A mutations. The background noise of next-generation sequencing for each of the common mutations within *KRAS* [Kirsten rat sarcoma viral oncogene homolog] (codons 12 and 13), *BRAF* [B-Raf proto-oncogene, serine/threonine kinase] (V600E), *EGFR* [epidermal growth factor receptor] (T790M and L858R), and five additional C:G sites in amplified regions of chromosomes 7 and 12 (a) without or (b) with UNG treatment prior to polymerase chain reaction (PCR) amplification was evaluated in normal peripheral blood samples from four healthy subjects. The C:G > T:A mutations are shown by *black circles*, whereas other potential mutations are shown by *black triangles*. (c) The difference between the two groups is shown by subtraction of the “No UNG” values from the “UNG”

values. Note that most C:G > T:A mutation frequencies are reduced with UNG treatment, except at two sites (shown by *black arrows*). The reason for the lack of an effect on the latter two sites is unknown. (d) The C:G > T:A mutation frequencies and the other baseline noise levels are averaged, and the C:G > T:A mutation frequencies are significantly higher than the other baseline noise levels ($P < 0.01$; Student's t test). UNG treatment prior to PCR significantly decreases the C:G > T:A mutation frequency ($P < 0.05$; paired t test), without affecting the frequency of the other mutations ($P > 0.05$; paired t test). Following UNG treatment, the C:G > T:A mutation frequency is still significantly higher than the other baseline noise level ($P < 0.01$; Student's t test). The *error bars* represent the standard deviations

Thus, UNG treatment did not interfere with the capability of NGS to detect clinically important mutations (Fig. 3; Table 1), and this was consistent with the findings of other studies [16].

4 Discussion

In this study, we demonstrated that the baseline noise of NGS is mainly attributable to cytosine deamination and that the source is both biologic and thermocycling induced. UNG eliminates the uracil that results from deamination and thus is a tool to reduce biologic background noise levels in NGS. Another possible method to eliminate PCR-induced mutations should be to avoid heating by employing isothermal amplification technologies. Pretreating

samples with UNG does not inhibit the ability to detect known positive control mutations. Our conclusions are all based on the Ion Torrent platform; however, our findings are consistent with similar work on the MiSeq system analyzing FFPE samples [22].

NGS is a powerful tool to discover novel disease-related genetic variations, to clinically diagnose and predict disease on the basis of comprehensive genetic profiling, and to reveal therapeutic targets. Extremely high depth of coverage allows NGS to be highly sensitive and accordingly suitable for discovery of rare genetic variants, including early detection of cancer and monitoring of MRD in cancer patients. In conducting NGS, reproducible sequence artifacts may produce false positives or interfere with detection of true gene mutations. For early detection of cancer or MRD monitoring, where extremely low mutation

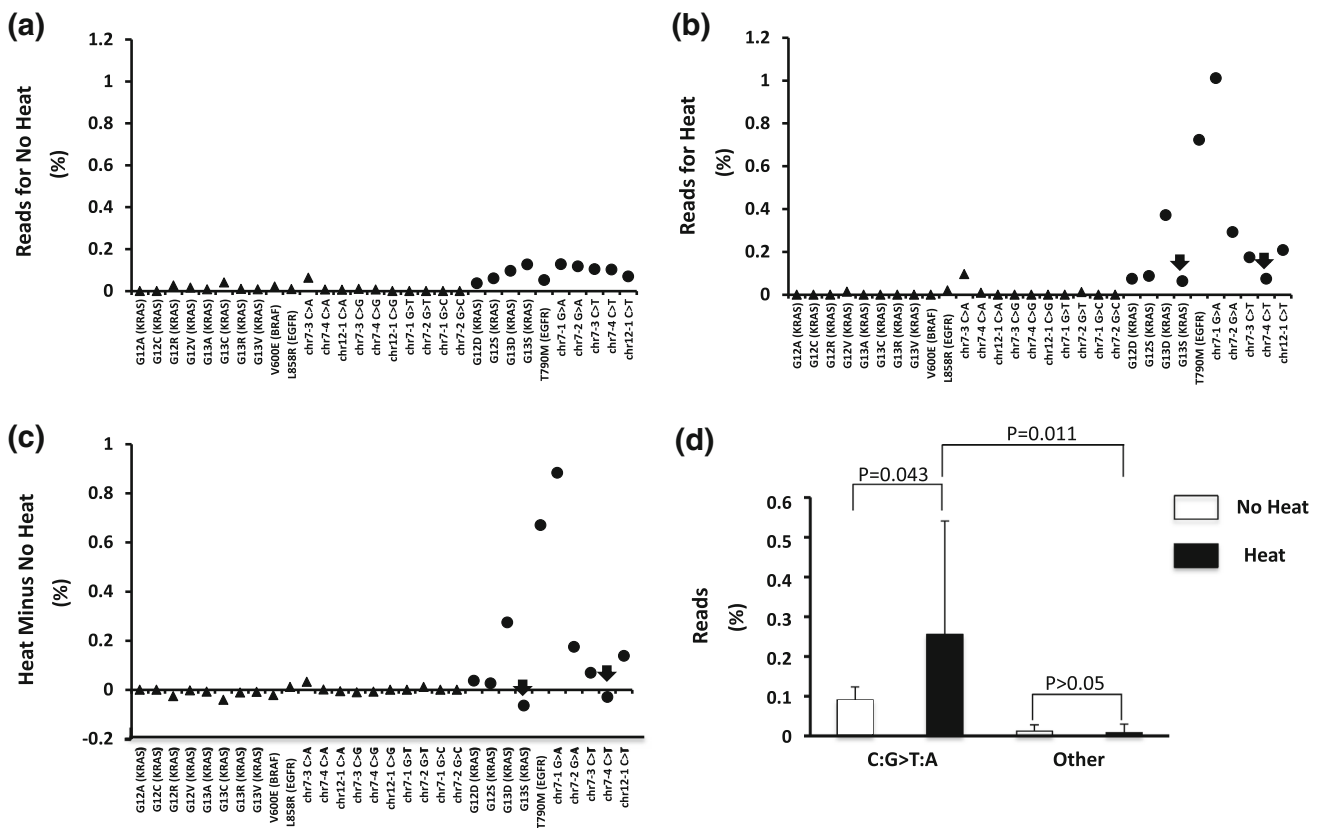


Fig. 2 Thermocycling induces deamination. The background noise of next-generation sequencing for each of the common mutations within *KRAS* [Kirsten rat sarcoma viral oncogene homolog] (codons 12 and 13), *BRAF* [B-Raf proto-oncogene, serine/threonine kinase] (V600E), *EGFR* [epidermal growth factor receptor] (T790M and L858R), and five additional C:G sites in amplified regions of chromosomes 7 and 12 (a) without or (b) with prolonged heat treatment prior to library preparation was evaluated in normal peripheral blood samples from four healthy subjects. (c) The difference between the two groups is shown by subtraction of the “No

Heat” values from the “Heat” values. While most of the values are above zero, there is clearly some variability among the positions in the response to heat. The C:G > T:A mutations are shown by *black circles* and are induced with heat treatment, except at two sites (shown by *black arrows*). Control mutations are shown by *black triangles*. (d) The C:G > T:A mutation frequencies and the other baseline noise levels are averaged, and prolonged heat significantly increases the C:G > T:A mutation frequency ($p < 0.05$; paired t test), without affecting the frequency of the other mutations ($p > 0.05$; paired t test). The *error bars* represent the standard deviations

Table 1 Mutation frequencies with and without uracil N-glycosylase (UNG) treatment

Sample	Gene	Mutation		Mutation frequency (%)	Mutation frequency [UNG] (%)
		Complementary DNA	Protein		
1	<i>KRAS</i>	c.35G>A	p.G12D	9.50	9.23
2	<i>KRAS</i>	c.35G>A	p.G12D	22.50	24.31
3	<i>KRAS</i>	c.35G>C	p.G12A	24	24.37
4	<i>BRAF</i>	c.1799T>A	p.V600E	36.30	33.33
5	<i>EGFR</i>	c.2573T>G	p.L858R	48.20	48.71
6	<i>EGFR</i>	c.2310_2311insAACCCCCAC	p.N771_H773dup	70.70	76.16
7	<i>EGFR</i>	c.2236_2249del15	p.E746_A750del	76.80	74.72

BRAF B-Raf proto-oncogene, serine/threonine kinase, *EGFR* epidermal growth factor receptor, *KRAS* Kirsten rat sarcoma viral oncogene homolog

frequencies need to be identified unambiguously, nonspecific background noise needs to be minimized, if not eliminated.

Cytosine deamination is actually one of the most prevalent point mutations spontaneously occurring in nature, thereby contributing to background noise for sequencing

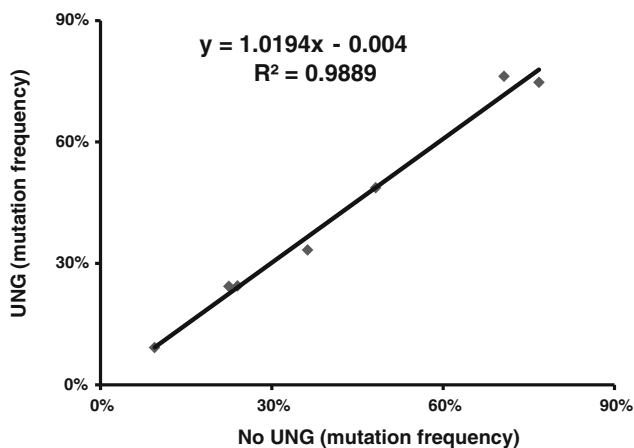


Fig. 3 Uracil N-glycosylase (UNG) does not interfere with detection of real mutations by next-generation sequencing. The mutation frequencies detected in UNG-treated positive control samples (shown on the *y*-axis) are compared with those in untreated ones (shown on the *x*-axis)

[15, 16]. The two major underlying mechanisms include deamination of 5-methylcytosine, resulting in thymine and ammonia. In DNA, this reaction can be corrected by the enzyme thymine-DNA glycosylase prior to passage of the replication fork, otherwise a cytosine to thymine base substitution is generated [23, 24]. The other mechanism of deamination involves hydrolysis of cytosine into uracil. This deamination in DNA is corrected by the DNA glycosylase UNG, which removes the uracil base to generate an abasic site, which is then repaired by adding back a cytosine opposite the guanine. However, if the pro-mutagenic G–U mispair is not repaired prior to the next round of DNA replication, a U:A mutation is generated [23, 24], which results in a T:A during the next round of synthesis.

Deamination can be attributed to multiple factors, including biologic factors (intrinsic to the sample prior to isolation) or an artifact of the molecular biology [23–25]. For studies of ancient DNA, it is a major source of sequencing artifacts [26]. In addition to age, it has been observed that formalin induces C:G > T:A transitions in Sanger sequencing [17]. However, we identified the same artifact in freshly prepared samples in our recent study validating detection of gene mutations using NGS [18]. This adds evidence that deamination may also result from polymerase-induced errors along with the lack of DNA repair, or directly from the heat associated with thermocycling [19]. Nevertheless, it appears that the frequencies of these mutations are higher in FFPE samples than in peripheral blood [18], indicating that the process of specimen fixation also contributes to the noise level.

In the current study, we demonstrated that biologic deamination contributes to C:G > T:A mutations in background noise, given that treating peripheral blood samples

with UNG prior to NGS led to a significant reduction in the C:G > T:A mutation frequency. While we favor a biologic source to explain the reduction by UNG treatment, we cannot eliminate the possibility that deamination is induced during DNA isolation. However, the reduction is only approximately 30 % (Fig. 1d), suggesting that the other 70 % of the mutations are already fixed (fully converted to T:A) prior to DNA isolation, or it is occurring during the process of PCR. In this regard, we found that the heat from the denaturation phase of thermocycling induced a significant increase in this background noise, consistent with the prolonged heating used by Ehrlich et al. [19]. To test whether it was solely an artifact of PCR, we attempted to add thermostable UNG from an extreme thermophile organism, *Archaeoglobus fulgidus* [27], during PCR. However, we were unable to identify conditions where this enzyme and the polymerase were both active (data not shown). An alternative approach may be to use an alternate thermostable UNG [28] or to replace traditional PCR with isothermal amplification.

5 Conclusion

A major cause of baseline noise in NGS is cytosine deamination. This appears to be pre-analytic (i.e., biologic in origin), but it can also be induced by the heat associated with thermocycling. Routine use of UNG pretreatment and isothermal amplification are viable strategies to reduce the background noise level in NGS.

Acknowledgments The authors wish to acknowledge Dr. James Stivers (Johns Hopkins University School of Medicine) for helpful discussions. Grant support was received from the Women’s Board of The Johns Hopkins Hospital (to Ming-Tseh Lin) and through National Institutes of Health grant numbers R21HG005745 (to Christopher D. Gocke) and R21CA164592 (to James R. Eshleman), and the Pancreatic Cancer Action Network Innovation Award (to James R. Eshleman).

Conflicts of Interest Guoli Chen, Stacy Mosier, Christopher D. Gocke, Ming-Tseh Lin, and James R. Eshleman have no conflicts of interest that are directly relevant to the content of this study.

References

1. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembem LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;437:376–80.
2. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. 2008;321:1801–6.
3. Biankin AV, Waddell N, Kassahn KS, Gingras MC, Muthuswamy LB, Johns AL, et al. Pancreatic cancer genomes reveal

- aberrations in axon guidance pathway genes. *Nature*. 2012; 491:399–405.
4. Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. *Sci Transl Med*. 2011;3:92ra66.
 5. Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*. 2011;331:1199–203.
 6. Dal Molin M, Hong SM, Hebbar S, Sharma R, Scrimieri F, de Wilde RF, et al. Loss of expression of the SWI/SNF chromatin remodeling subunit BRG1/SMARCA4 is frequently observed in intraductal papillary mucinous neoplasms of the pancreas. *Hum Pathol*. 2012;43:585–91.
 7. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, et al. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. *J Mol Diagn*. 2012;14:357–66.
 8. van der Heijden MS, Brody JR, Dezentje DA, Gallmeier E, Cunningham SC, Swartz MJ, et al. In vivo therapeutic responses contingent on Fanconi anemia/BRCA2 status of the tumor. *Clin Cancer Res*. 2005;11:7508–15.
 9. Turner NC, Lord CJ, Iorns E, Brough R, Swift S, Elliott R, et al. A synthetic lethal siRNA screen identifying genes mediating sensitivity to a PARP inhibitor. *EMBO J*. 2008;27:1368–77.
 10. 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.
 11. Heitzer E, Auer M, Gasch C, Pichler M, Ulz P, Hoffmann EM, et al. Complex tumor genomes inferred from single circulating tumor cells by array-CGH and next-generation sequencing. *Cancer Res*. 2013;73:2965–75.
 12. Kanda M, Knight S, Topazian M, Syngal S, Farrell J, Lee J, et al. Mutant GNAS detected in duodenal collections of secretin-stimulated pancreatic juice indicates the presence or emergence of pancreatic cysts. *Gut*. 2013;62:1024–33.
 13. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med*. 2010;2:20ra14.
 14. Leary RJ, Sausen M, Kinde I, Papadopoulos N, Carpten JD, Craig D, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med*. 2012;4:162ra54.
 15. Hofreiter M, Jaenicke V, Serre D, von Haeseler A, Paabo S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res*. 2001;29:4793–9.
 16. Do H, Dobrovic A. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil-DNA glycosylase. *Oncotarget*. 2012; 3:546–58.
 17. Williams C, Ponten F, Moberg C, Soderkvist P, Uhlen M, Ponten J, et al. A high frequency of sequence alterations is due to formalin fixation of archival specimens. *Am J Pathol*. 1999; 155:1467–71.
 18. Lin MT, Mosier S, Cope L, Thiess M, Beierl K, Chen G, et al. Clinical validation of KRAS, BRAF, and EGFR mutation detection using next generation sequencing. *Am J Clin Pathol*. 2014;141:856–66.
 19. Ehrlich M, Norris KF, Wang RY, Kuo KC, Gehrke CW. DNA cytosine methylation and heat-induced deamination. *Biosci Rep*. 1986;6:387–93.
 20. Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, et al. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *J Mol Diagn*. 2010;12:425–32.
 21. Lin MT, Tseng LH, Rich RG, Hafez MJ, Harada S, Murphy KM, et al. Delta-PCR, a simple method to detect translocations and insertion/deletion mutations. *J Mol Diagn*. 2011;13:85–92.
 22. Do H, Wong SQ, Li J, Dobrovic A. Reducing sequence artifacts in amplicon-based massively parallel sequencing of formalin-fixed paraffin-embedded DNA by enzymatic depletion of uracil-containing templates. *Clin Chem*. 2013;59:1376–83.
 23. Yonekura S, Nakamura N, Yonei S, Zhang-Akiyama QM. Generation, biological consequences and repair mechanisms of cytosine deamination in DNA. *J Radiat Res*. 2009;50:19–26.
 24. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. *Nature*. 1980;287:560–1.
 25. Bjelland S, Seeberg E. Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutat Res*. 2003;531:37–80.
 26. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Paabo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res*. 2010;38:e87.
 27. Sandigursky M, Franklin WA. Uracil-DNA glycosylase in the extreme thermophile *Archaeoglobus fulgidus*. *J Biol Chem*. 2000;275:19146–9.
 28. Sartori AA, Fitz-Gibbon S, Yang H, Miller JH, Jiricny J. A novel uracil-DNA glycosylase with broad substrate specificity and an unusual active site. *EMBO J*. 2002;21:3182–91.