Variation in Organ-Specific \textit{PIK3CA} and \textit{KRAS} Mutant Levels in Normal Human Tissues Correlates With Mutation Prevalence in Corresponding Carcinomas

Barbara L. Parsons  
\textit{U.S. Food and Drug Administration}, Barbara.parsons@fda.hhs.gov

Karen L. McKim  
\textit{U.S. Food and Drug Administration}

Meagan B. Myers  
\textit{U.S. Food and Drug Administration}

Follow this and additional works at: http://digitalcommons.unl.edu/usfda

Part of the \textit{Dietetics and Clinical Nutrition Commons, Health and Medical Administration Commons, Health Services Administration Commons, Pharmaceutical Preparations Commons}, and the \textit{Pharmacy Administration, Policy and Regulation Commons}

Parsons, Barbara L.; McKim, Karen L.; and Myers, Meagan B., "Variation in Organ-Specific \textit{PIK3CA} and \textit{KRAS} Mutant Levels in Normal Human Tissues Correlates With Mutation Prevalence in Corresponding Carcinomas" (2017). \textit{Food and Drug Administration Papers}. 1.  
http://digitalcommons.unl.edu/usfda/1

This Article is brought to you for free and open access by the U.S. Department of Health and Human Services at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Food and Drug Administration Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Variation in Organ-Specific PIK3CA and KRAS Mutant Levels in Normal Human Tissues Correlates With Mutation Prevalence in Corresponding Carcinomas

Barbara L. Parsons,* Karen L. McKim, and Meagan B. Myers
Division of Genetic and Molecular Toxicology, U.S. Food and Drug Administration, National Center for Toxicological Research, Jefferson, Arkansas

Large-scale sequencing efforts have described the mutational complexity of individual cancers and identified mutations prevalent in different cancers. As a complementary approach, allele-specific competitive blocker PCR (ACB-PCR) is being used to quantify levels of hotspot cancer driver mutations (CDMs) with high sensitivity, to elucidate the tissue-specific properties of CDMs, their occurrence as tumor cell subpopulations, and their occurrence in normal tissues. Here we report measurements of PIK3CA H1047R mutant fraction (MF) in normal colonic mucosa, normal lung, colonic adenomas, colonic adenocarcinomas, and lung adenocarcinomas. We report PIK3CA E545K MF measurements in those tissues, as well as in normal breast, normal thyroid, mammary ductal carcinomas, and papillary thyroid carcinomas. We report KRAS G12D and G12V MF measurements in normal colon. These MF measurements were integrated with previously published ACB-PCR data on KRAS G12D, KRAS G12V, and PIK3CA H1047R. Analysis of these data revealed a correlation between the degree of interindividual variability in these mutations (as log10 MF standard deviation) in normal tissues and the frequencies with which the mutations are detected in carcinomas of the corresponding organs in the COSMIC database. This novel observation has important implications. It suggests that interindividual variability in mutation levels of normal tissues may be used as a metric to identify mutations with critical early roles in tissue-specific carcinogenesis. Additionally, it raises the possibility that personalized cancer therapeutics, developed to target specifically activated oncogenic products, might be repurposed as prophylactic therapies to reduce the accumulation of cells carrying CDMs and, thereby, reduce future cancer risk.

Key words: biomarker; cancer susceptibility; tumor heterogeneity; personalized medicine; clonal selection

INTRODUCTION

Large-scale cancer genome sequencing projects have been valuable in terms of identifying the prevalent cancer driver mutations (CDMs) associated with particular types of cancers [Ciriello et al., 2013; Kandoth et al., 2013; Watson et al., 2013; Stover and Wagle, 2015]. Cancer “drivers” have been defined as genetic events associated with tumor initiation or progression [Alizadeh et al., 2015] and as somatic mutations that increase the fitness of a cell [Fisher et al., 2013]. More broadly, a cancer driver is “a cell autonomous or non-cell autonomous alteration that contributes to the tumor evolution at any stage—including initiation, progression, metastasis, and resistance to therapy—by promoting a variety of functions including proliferation, survival, invasion, or immune evasion” [Alizadeh et al., 2015]. Some of the most impactful cancer drivers are hotspot point mutations. Hotspot point mutations merit investigation because of their high prevalence in cancer, their established roles as

Grant sponsor: National Center for Toxicological Research, U.S. Food and Drug Administration.
Additional Supporting Information may be found in the online version of this article.
*Correspondence to: Barbara L. Parsons, Division of Genetic and Molecular Toxicology, National Center for Toxicological Research, U.S. Food and Drug Administration, HFT-120, 3900 NCTR Rd., Jefferson, AR 72079, USA. E-mail: Barbara.parsons@fda.hhs.gov
Received 24 February 2017; provisionally accepted 10 May 2017; and in final form 14 June 2017
DOI 10.1002/em.22110
Published online 29 July 2017 in Wiley Online Library (wileyonlinelibrary.com).

© 2017 This article is a U.S. Government work and is in the public domain in the USA. Environmental and Molecular Mutagenesis published by Wiley Periodicals, Inc. on behalf of Environmental Mutagen Society
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
early events in carcinogenesis, and evidence that the mutant proteins can be therapeutic targets and/or determinants of the efficacy of specific anticancer therapies (i.e., biomarkers to inform personalized approaches to cancer treatment) [Lu et al., 2014; Ryan et al., 2015]. There are known hotspot base substitution mutations in the PIK3CA and KRAS genes. The prevalence of these mutations in carcinomas of the breast, colon, lung and thyroid are provided in Supporting Information Table S1 (data from Catalogue of Somatic Mutations in Cancer, COSMIC) [COSMIC, 2016].

Individual tumors are heterogeneous in terms of histology, genetics, epigenetics, phenotypic markers, and gene expression [Bonavia et al., 2011; Marusyk et al., 2012]. It has been established that cancers exhibit spatial and temporal genetic diversity, meaning different mutations have been detected in different sectors of a cancer and that abundance of clonal subpopulations can evolve over time [Martelotto et al., 2014; Renovanz and Kim, 2014]. Investigations into the genetic heterogeneity of several cancers has enabled clinicians to identify patients who will benefit from particular therapies (i.e., personalized cancer treatment) and, conversely, demonstrated that mutant cell subpopulations can derail the process and drive the development of acquired resistance to therapy [Fisher et al., 2013; Burrell and Swanton, 2014; Stover and Wagle, 2015]. Given the extent of tumor heterogeneity, the prevalence of mutant subpopulations in advanced cancers and the difficulties associated with treating such cancers, it has been suggested that the development of cancer preventative modalities may be a more effective approach for reducing cancer deaths [Maresso et al., 2015].

Because tumor heterogeneity is a consequence and a reflection of the mechanisms driving tumor initiation and progression, understanding the origins of tumor heterogeneity is critical for designing therapeutic strategies to block the development of resistance to treatment [Hiley et al., 2014]. In addition, information about the earliest stages of carcinogenesis is needed as a foundation for experimental approaches to abrogate or delay the development of cancer [Kensler et al., 2016]. For these reasons, there is growing interest in describing the mutations present in normal and pre-neoplastic tissues and tracking genetic changes during tumor progression and treatment [Hiley et al., 2014; Jamal-Hanjani et al., 2014; Nakamura et al., 2014; Roberts and Gordenin, 2014; Francioli et al., 2015; Martincorena and Campbell, 2015].

It has been asserted that a majority of cancers (65%, with 95% CI of 39–81%) arise due to spontaneous mutations, a claim based on the correlation between the incidence of cancer in different tissue types and the number of stem cell divisions in those tissues [Tomasetti and Vogelstein, 2015]. Others have claimed that intrinsic factors account for <10 to 30% of lifetime cancer risk and “rates of endogenous mutation accumulation due to intrinsic processes are not sufficient to account for the observed cancer rates” [Wu et al., 2016]. Clearly, additional information about the frequencies of spontaneous mutations in cancer driver genes and the selective advantages they may confer is necessary to clarify how spontaneous mutations contribute to cancer risk and whether reducing spontaneous CDM frequencies through medical intervention is a viable approach for reducing cancer deaths [Martincorena and Campbell, 2015].

Our laboratory uses a high sensitivity method called Allele-specific Competitive Blocker PCR (ACB-PCR), to quantify specific CDMs in normal and cancerous tissue samples derived from particular organs [Parsons et al., 2010; Myers et al., 2014a, 2015, 2016]. By comparing ACB-PCR amplification of unknown samples to that of a standard curve, ACB-PCR can determine the ratio of mutant: wild-type sequence in a DNA sample, measuring one specific base substitution mutation at a time, as long as the mutant to wild-type ratio is \( \geq 10^{-5} \). The ACB-PCR limit of detection is considered to be \( 1 \times 10^{-5} \) because this is the lowest MF standard that can be discriminated experimentally from a wild-type only control. An ACB-PCR MF of \( 10^{-5} \) equates to the detection of three mutant molecules in a total of 300,000. ACB-PCR is robust in the measurement of MFs between \( 10^{-1} \) and \( 10^{-5} \) (ACB-PCR may underestimate MFs between \( 10^{-1} \) and 1). ACB-PCR can identify robustly which samples have MFs below \( 10^{-5} \) but cannot measure MFs below \( 10^{-5} \). This sensitivity is unlikely to be sufficient to detect neutral mutations in normal human tissues. For some CDMs, however, this sensitivity has been shown to be sufficient for the mutant quantification in normal rodent and human tissues, likely due to clonal expansion of cells carrying CDMs that confer a positive selective advantage [Parsons et al., 2005, 2010, 2012; McKinzie et al., 2006; Meng et al., 2010a,b, 2011; McKinzie and Parsons, 2011; Banda et al., 2013, 2015]. Information about ultralow frequency mutant cell populations can complement existing characterizations of tumor samples by next-generation sequencing (NGS, which generally detects variant alleles only at frequencies \( >1\% \) [Cotrell et al., 2014]), to yield a more complete description of the genetic changes present in cancers. Here we report quantification of levels of the PIK3CA E545K and H1047R hotspot point mutations in normal colonic mucosa, colonic adenocarcinomas, normal lung, and lung adenocarcinomas, as well as levels of the PIK3CA E545K mutation in normal breast, mammary ductal carcinoma, normal thyroid, and papillary thyroid carcinoma (PTC). These mutations were selected because they have the potential to provide a positive selective advantage. The PIK3CA E545K (guanine to adenine mutation at codon 545) results in a glutamine to lysine amino acid change in the p110α kinase binding domain. The PIK3CA H1047R (adenine to guanine mutation at codon 1047) results in a histidine to arginine amino acid change in the p110α
kinase binding domain. Both PIK3CA mutations increase activation of PI3K and its binding to the cell membrane, thereby impacting oncogenic transformation, cell proliferation, cell growth, survival, differentiation and altering invasion potential [Myers et al., 2016].

KRAS is a downstream mediator of growth factor receptor signaling, with critical roles in cell proliferation, survival, and differentiation. The KRAS G12D (GGT to GAT at codon 12 resulting in a glycine to aspartic acid amino acid change) and KRAS G12V (GGT to GTT at codon 12 resulting in a glycine to valine amino acid change) mutations alter the phenotype of KRAS, providing a context-dependent selective advantage [Parsons and Myers, 2013]. Here, we also report additional measurements of KRAS G12D and G12V in normal colonic mucosa. These new data, integrated with that published previously, creates a dataset of four hotspot point mutations (KRAS G12D, KRAS G12V, PIK3CA E545K, and PIK3CA H1047R) in normal and malignant tissues from four different organs. Analysis of these data demonstrates that interindividual variability in cancer driver gene mutant fraction (MF) within normal tissues is correlated with the tissue-specific prevalence of carcinomas carrying the corresponding mutations.

**MATERIALS AND METHODS**

**Sample Collection**

Procedures for the acquisition and analysis of anonymous human tissues were reviewed by the FDA’s IRB (Research Involving Human Subjects Committee, FWA 00006196). The National Disease Research Interchange (NDRI, Philadelphia, PA) and the National Cancer Institute’s Cooperative Human Tissue Network (CHTN) were identified as the primary sources of normal and malignant tissues, respectively. Fresh-frozen, normal breast, normal colonic mucosa, normal lung, and normal thyroid tissue samples, as well as one fresh-frozen papillary thyroid carcinoma (PTC), were purchased from the NDRI [Parsons et al., 2010; Myers et al., 2014a, 2015, 2016]. Fresh-frozen, primary breast ductal carcinomas (DCs), colonic adenomas, colonic adenocarcinomas, lung adenocarcinomas, and PTCs, along with normal colonic mucosa samples, were provided by the CHTN [Parsons et al., 2010; Parsons and Myers 2013; Myers et al., 2014a, 2015, 2016]. These samples were collected at autopsy from tissue donors (one sample per donor) who died from causes unrelated to cancer or diseases affecting the relevant target organ. The numbers of samples analyzed for the different mutations are given in Supporting Information Table S2. All tumor specimens were histologically evaluated and classifications were confirmed by board certified pathologists. The extent of information provided regarding the smoking history of the subjects was inconsistent across subjects (and in some instances absent), only allowing for stratification of subjects as ever- and never-smokers.

**DNA Isolation and First-Round PCR**

DNA isolation was performed as described previously [Parsons et al., 2010; Myers et al., 2014a, 2015, 2016], using sufficient tissue to ensure that the characteristic cellular content of each tissue type would be represented proportionally in the genomic DNA. The average weights of the tissue samples processed for breast, colonic mucosa (separated from submucosa), lung and thyroid were 4.0, 0.4, 2.3, and 1.9 g, respectively. A high fidelity, first-round PCR amplification of a 306 bp sequence encompassing human PIK3CA codon 545 was performed for each DNA sample, using 1 μg of EcoRI-digested genomic DNA in a 200 μl PCR reaction mix of: 10 mM KCl, 10 mM (NH₄)₂SO₄, 20 mM Tris-HCl (pH 8.75), 2 mM MgSO₄, 0.1% Triton X-100, 0.1 mg/ml bovine serum albumin, 0.2 mM dNTPs, and 10 units of cloned PfuUltra Hotstart DNA Polymerase (Agilent Technologies, Santa Clara, CA). The primers were 5'-GGGGAAATGACAAAGGAACAG-3' and 5'-AATGTGCCAAACTACCAATGTG-3'. The thermocycler reaction conditions were 2 min at 94°C, followed by 31 cycles of 1 min at 94°C, 2 min at 56°C, and 1 min at 72°C, followed by a final 7 min at 72°C extension and 4°C hold.

A high fidelity, first-round PCR amplification of a 265 bp sequence encompassing human PIK3CA codon 1047 was performed for each DNA sample, using 1 μg of EcoRI-digested genomic DNA in a 200 μl PCR reaction mix of: 10 mM KCl, 10 mM (NH₄)₂SO₄, 20 mM Tris-HCl (pH 8.75), 2 mM MgSO₄, 0.1% Triton X-100, 0.1 mg/ml bovine serum albumin, 0.2 mM dNTPs, and 10 units of cloned PfuUltra Hotstart DNA Polymerase (Agilent Technologies). The primers were 5'-CCAAGCTTCCCTACTTACTTATAG-3' and 5'-GAAGATCAATACCTATTTGTTG-3'. Thermocycler reaction conditions were 2 min at 94°C, followed by 33 cycles of 1 min at 94°C, 1 min at 52°C, and 1 min at 72°C, followed by a final 6 min at 72°C extension and 4°C hold.

The high fidelity, first-round PCR of a 384 bp sequence encompassing human KRAS codon 12 was described previously [Myers et al., 2014a]. Each first-round PCR employed one intron-specific primer to avoid amplification of pseudogenes.

**Synthesis of PIK3CA and KRAS WT and Mutant Standards**

Plasmids carrying the mutant or wild-type sequences (for the two different ampiclons of the PIK3CA gene) were used as template to synthesize mutant and wild-type standards. Plasmids carrying G12D mutant KRAS, G12V mutant KRAS, and wild-type KRAS sequences were used as template to synthesize standards. Each of the standards was synthesized using conditions identical to those used to amplify the unknown genomic DNAs, except that the amount of input plasmid DNA was calibrated prior to PCR to ensure approximately equivalent amplification in the synthesis of standards and unknowns.

**Purification and Quantification of Standards and Unknowns**

Purification of PIK3CA E545K WT, mutant, and unknown PCR products was accomplished by ion-pair reverse phase chromatography using non-denaturing conditions and a Transgenomic WAVE Nucleic Acid Fragment Analysis and Collection system (Omaha, NE). PIK3CA H1047R PCR products and KRAS PCR products were gel purified as described previously [Myers et al., 2016]. The DNA concentration of each sample was determined using an Epoch Spectrophotometer Model (Biotek, Winooski, VT), calculated as the average of three measurements that varied by ≤10% from the group mean.

**ACB-PCR**

Mutant (either PIK3CA E545K, PIK3CA H1047R, KRAS G12D, or KRAS G12V) and WT reference DNA samples were mixed to generate standards with mutant:WT ratios (i.e., MFs) of 10⁻¹, 10⁻², 10⁻³, 10⁻⁴, 10⁻⁵, and 0 (containing only the WT sequence), at a concentration of 5 × 10⁸ copies/μl for the PIK3CA E545K mutation or 1 × 10⁹ copies/μl for the PIK3CA H1047R mutation. Each ACB-PCR reaction employed 10 μl of each DNA mixture, for a total of 5 × 10⁸ or 1 × 10⁹ copies per reaction for the E545K and H1047R assays, respectively. Each MF standard was analyzed in duplicate, along with a no-DNA control. ACB-PCR reactions were performed in 96-well PCR plates (Thermo Fisher Scientific, Waltham, MA) using three primers: a fluorescein-labeled
mutant-specific primer (MSP), a non-extendable blocker primer (BP), to reduce background amplification of WT PIK3CA, and an upstream/downstream primer [Myers et al., 2014b]. The ACB-PCR measurement of KRAS G12D and G12V was performed as described previously [Myers et al., 2015].

For PIK3CA E545K mutation, each 50 μl ACB-PCR reaction contained: 1× Universal Taq Buffer (New England BioLabs, Inc. (NEB), Beverly, MA), 1.5 mM MgCl₂, 0.1 mg/ml gelatin, 1.0 mg/ml Triton X-100, 80 μM dNTPs, 400 nM each primer [MSP, 5'-fluorescein-TCTCTCTCTGAAAATCACCAG-3'; BP, 5'-TCTCTCTCTGAAAATCACCAG-3' (i.e., 3'-deoXyG modification); and downstream primer, 5'-CAGAGAATCTCCATTTAGC-3'], 3 ng/μl ET SSB (NEB), 2.8 mM/μl Perfect Match PCR Enhancer (Stratagene, La Jolla, CA), and 0.08 μl Hemo KlenTaq DNA polymerase (NEB). Cycling conditions were 2 min at 94°C, followed by 36 cycles of 94°C for 30 sec, 41°C for 90 sec, and 68°C for 1 min. The PIK3CA E545K ACB-PCR product is 77 bp in length.

For PIK3CA H1047R mutation, each 50 μl ACB-PCR reaction contained: 1× Universal Taq Buffer, 0.75 mM MgCl₂, 0.1 mg/ml gelatin, 1.0 mg/ml Triton X-100, 40 μM dNTPs, 270 nM MSP (5'-fluorescein-TGTTGTCCACGCCAGATGTC-3'), 500 nM BP (5'-TGTGTTGTCCACGCCAGATGTC-3'), 1.5 μM downstream primer (5'-GATGCTTGGCTCTGAGAATCTCCATTTAGC-3'), 1,2 μl Perfect Match PCR Enhancer, and 0.3 μl Hemo KlenTaq DNA polymerase. Cycling conditions were 2 min at 94°C, followed by 41 cycles of 94°C for 30 sec, 49°C for 45 sec, and 72°C for 1 min. The PIK3CA H1047R ACB-PCR product is 148 bp in length. Blocker primers were purchased from Sigma Genosys (The Woodlands, TX), all others were purchased from Integrated DNA Technologies, Inc. (Coralville, IA).

**Vertical Polyacrylamide Gel Electrophoresis, Image Analysis, and Data Collection**

Following ACB-PCR, 10 μl of 6× ficoll loading buffer/dye were added to each 50 μl reaction, and 10 μl of each ACB-PCR product were analyzed on 8% non-denaturing, polyacrylamide gels. A PharosFX Molecular Imager with an external blue laser (Bio-Rad) was used to visualize the fluorescent ACB-PCR products. Quantity One software (Bio-Rad), with a locally averaged background correction, was used to quantify the pixel intensities of the correct-sized bands. For the PIK3CA E545K, log-log plots relating MF to fluorescence were constructed and fit with a power function. For the PIK3CA H1047R, KRAS G12D, and G12V mutations, log-linear plots relating MF to fluorescence were constructed and fit with a logarithmic function. Using the function of the standard curve and the pixel intensities of the ACB-PCR products, the MF (ratio of mutant to wild-type sequence) of each unknown sample was calculated.

**Statistical Analyses**

For each sample, MF was calculated as the arithmetic average of three independent ACB-PCR measurements. The average MF measurement for each sample was log₁₀-transformed. For each mutational target, the geometric mean MF was calculated as the average log₁₀-transformed MF measured for a particular tissue type. Statistical analyses were performed using log₁₀-transformed data.

Log-transformed datasets were examined for normality using the D’Agostino and Pearson omnibus normality test. Nonparametric approaches were applied to log-transformed data that were not normally distributed. The Mann-Whitney rank sum test was used when a quantitative ranking of MFs was possible. Contingency analyses were used to test for statistical significance when datasets contained multiple samples with MFs below the limit of accurate ACB-PCR quantification (i.e., <10⁻⁵). Specifically, the numbers of samples with MFs >10⁻⁵ and <10⁻⁵ were examined using χ² or Fisher’s exact test. Pearson correlation analyses were performed on log-transformed, normally-distributed data (Spearman’s rank correlation analyses were performed when data were not normally-distributed). Two-tailed P values <0.05 were considered significant. All statistical analyses were performed using GraphPad Prism 5 Software (GraphPad Software, Inc., La Jolla, CA).

**RESULTS**

The goals of this study were to elucidate the role of spontaneous CDMs in tissue-specific carcinogenesis and to define the prevalence of low frequency CDMs within cancers. Therefore, unlike many studies that examine tumor tissue and normal tumor adjacent sample, this study examined the levels of hotspot CDM in “normal” autopsy samples from men (except breast) and women without cancer or disease in the relevant organ.

Levels of PIK3CA E545K mutation (codon 545 GAG→AAG) were measured in normal breast, colonic mucosa, lung, and thyroid DNA samples, as well as in DNA isolated from ductal carcinomas, colonic adenomas, colonic adenocarcinomas, lung adenocarcinomas, and papillary thyroid carcinomas. Each unknown was quantified in three independent experiments, generating a data-set of 441 PIK3CA E545K MF measurements. The numbers of samples analyzed for each tissue type are presented in Supporting Information Table S2. Following ACB-PCR, products were run on vertical polyacrylamide gels. Replicate examples of PIK3CA E545K ACB-PCR output are provided in Supporting Information Figure S1. MF quantification was achieved by interpolation of the fluorescent intensities of unknown samples with that of a standard curve constructed using samples with defined ratios of mutant:WT alleles (i.e., duplicate 10⁻¹, 10⁻², 10⁻³, 10⁻⁴, 10⁻⁵, and 0 standards). A representative example of a PIK3CA E545K ACB-PCR standard curve is shown in Supporting Information Figure S2. The average coefficient of determination (r²) for the PIK3CA E545K standard curves were as follows: breast, 0.9730 (range 0.9662–0.9828); colon, 0.9628 (range 0.9334–0.9831); lung, 0.9595 (range 0.9316–0.9726); and thyroid, 0.9670 (range 0.9595–0.9724). The measured MFs are given in Supporting Information Table S3. The average coefficient of variation for the triplicate PIK3CA E545K MF measurements obtained from individual breast, colon, lung, and thyroid samples were 0.36, 0.55, 0.35, and 0.43, respectively.

PIK3CA H1047R (codon 1047 CAT→CGT) MF was measured in normal colonic mucosa and lung DNA samples, as well as in DNA isolated from colonic adenomas, colonic adenocarcinomas, and lung adenocarcinomas (see Supporting Information Table S2). Each unknown was quantified in three independent experiments (276 PIK3CA H1047R MF measurements). Replicate examples of PIK3CA H1047R ACB-PCR output are provided in Supporting Information Figure S1. A representative example of a PIK3CA H1047R ACB-PCR standard curve is shown...
Malignant tissues were observed between the levels of PIK3CA and tumors is given in Table I. No significant differences were observed between the levels of KRAS G12D and G12V ACB-PCR output are provided in Supporting Information Figure S1. Representative examples of KRAS published ACB-PCR analysis of six colonic mucosa samples are given in Supporting Information Table S4. The average coefficient of variation for the triplicate PIK3CA H1047R MF measurements obtained from individual colon and lung samples were 0.40 and 0.16, respectively.

Fifteen normal colonic mucosa samples were analyzed for KRAS G12D (codon 12 GGT→GAT) and G12V (codon 12 GGT→GTT) MFs as described above (90 ACB-PCR MF measurements), adding to the previously published ACB-PCR analysis of six colonic mucosa samples [Parsons et al., 2010]. Replicate examples of the KRAS G12D and G12V ACB-PCR output are provided in Supporting Information Figure S1. Representative examples of KRAS G12D and G12V standard curves are shown in Supporting Information Figure S2. The average coefficient of determination ($r^2$) for the standard curves used to measure PIK3CA H1047R MF in colon and lung were 0.9849 (range 0.9732–0.9928) and 0.9802 (range 0.9738–0.9840), respectively. The PIK3CA H1047R MFs are given in Supporting Information Table S4. The average coefficient of variation for the triplicate PIK3CA H1047R MF measurements obtained from individual colon and lung samples were 0.40 and 0.16, respectively.

Fifteen normal colonic mucosa samples were analyzed for KRAS G12D (codon 12 GGT→GAT) and G12V (codon 12 GGT→GTT) MFs as described above (90 ACB-PCR MF measurements), adding to the previously published ACB-PCR analysis of six colonic mucosa samples [Parsons et al., 2010]. Replicate examples of the KRAS G12D and G12V ACB-PCR output are provided in Supporting Information Figure S1. Representative examples of KRAS G12D and G12V standard curves are shown in Supporting Information Figure S2. The average coefficient of determination ($r^2$) for the standard curves used to measure KRAS G12D and G12V MF in colon were 0.9895 (range 0.9735–0.9969) and 0.9889 (range 0.9744–0.9973), respectively. The KRAS G12D and G12V MFs are given in Supporting Information Table S5. The average coefficient of variation for the triplicate KRAS G12D and G12V MF measurements obtained from individual colon samples were 0.52 and 0.17, respectively.

Comparison of PIK3CA E545K, PIK3CA H1047R, KRAS G12D, and KRAS G12V MFs Between Normal and Malignant Tissues

A summary of MF measurements for normal tissues and tumors is given in Table I. No significant differences were observed between the levels of PIK3CA E545K mutation in normal tissues and tumor samples, for any of the organs examined (see Fig. 1), although the increased MFs in colonic adenomas compared to normal colonic mucosa were of borderline significance (Fisher’s exact test, $P = 0.0741$). No significant differences were observed between the levels of PIK3CA H1047R mutation in normal tissue and tumors of the colon or lung (see Fig. 1). For KRAS G12D and G12V, the 15 ACB-PCR measurements from normal colonic mucosa were compared to those previously published for colonic adenomas and adenocarcinomas [Parsons et al., 2010]. This analysis confirmed an earlier report conducted using just six normal samples. Significant differences in MF were observed among sample types (Kruskal-Wallis test, $P = 0.0013$), with colonic adenomas having significantly greater KRAS G12D mutant levels than normal colonic mucosa. Significant differences in KRAS G12V MFs were also observed ($r^2$ test, $P = 0.0220$).

The data developed in the current study, combined with that previously published [Parsons et al., 2010; Myers et al., 2014a, 2015, 2016], constitutes a data set of more than 1,800 ACB-PCR MF measurements on 606 different DNA samples. The mutant frequency distributions observed in normal/tumor tissue pairs are presented in Figure 1, for each mutation analyzed. Figure 1 illustrates for the first time: (1) the tissue-specific differences in ultralow frequency KRAS and PIK3CA mutation levels, (2) the remarkable prevalence of low-frequency mutant populations, and (3) in some cases, a surprising overlap in the frequency distributions between normal and tumor.

### MF Measurements in Normal Tissues

PIK3CA H1047R MF in normal colonic mucosa and lung, along with PIK3CA E545K MF in normal colonic

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Gene mutation</th>
<th>Tissue</th>
<th>Median MF</th>
<th>Geomean MF</th>
<th>No. samples &gt;10&lt;sup&gt;−5&lt;/sup&gt;</th>
<th>No. samples &lt;10&lt;sup&gt;−5&lt;/sup&gt;</th>
<th>Mutation positive samples (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>PIK3CA H1047R</td>
<td>Colon</td>
<td>1.57 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>1.98 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>20</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Normal</td>
<td>PIK3CA H1047R</td>
<td>Lung</td>
<td>2.45 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>3.96 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>19</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Normal</td>
<td>PIK3CA E545K</td>
<td>Breast</td>
<td>6.06 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>1.78 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>4</td>
<td>6</td>
<td>40</td>
</tr>
<tr>
<td>Normal</td>
<td>PIK3CA E545K</td>
<td>Colon</td>
<td>2.67 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>2.69 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Normal</td>
<td>PIK3CA E545K</td>
<td>Lung</td>
<td>5.35 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>6.75 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>3</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Normal</td>
<td>PIK3CA E545K</td>
<td>Thyroid</td>
<td>6.52 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>7.62 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>5</td>
<td>14</td>
<td>26</td>
</tr>
<tr>
<td>Normal</td>
<td>KRAS G12D</td>
<td>Colon</td>
<td>2.88 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>4.90 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>15</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Normal</td>
<td>KRAS G12V</td>
<td>Colon</td>
<td>7.66 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>1.08 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>6</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>Normal</td>
<td>PIK3CA E545K</td>
<td>Colon</td>
<td>1.73 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>4.56 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>8</td>
<td>1</td>
<td>89</td>
</tr>
<tr>
<td>Adenoma</td>
<td>PIK3CA E545K</td>
<td>Colon</td>
<td>2.15 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>2.38 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>2</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>PIK3CA H1047R</td>
<td>Colon</td>
<td>1.63 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>3.46 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>19</td>
<td>1</td>
<td>95</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>PIK3CA H1047R</td>
<td>Lung</td>
<td>2.54 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>3.48 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>24</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>PIK3CA E545K</td>
<td>Breast</td>
<td>3.92 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>8.36 × 10&lt;sup&gt;−5&lt;/sup&gt;</td>
<td>7</td>
<td>2</td>
<td>78</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>PIK3CA E545K</td>
<td>Colon</td>
<td>3.07 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>7.89 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>3</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>PIK3CA E545K</td>
<td>Lung</td>
<td>4.17 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>7.13 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>4</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>PIK3CA E545K</td>
<td>Thyroid</td>
<td>8.39 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>8.78 × 10&lt;sup&gt;−6&lt;/sup&gt;</td>
<td>7</td>
<td>11</td>
<td>39</td>
</tr>
</tbody>
</table>

The median and geomean MFs shown in italics are below the limit of accurate ACB-PCR quantification (10<sup>−5</sup>).
mucosa, lung, breast, and thyroid, were examined for correlations with age, smoking status, and gender. No statistically significant associations with age or smoking status were observed (comparing ever smoker vs. never-smokers). For colonic mucosa, however, the \textit{PIK3CA} H1047R mutation was significantly more abundant in DNA from men than from women (Fig. 2). This novel finding is consistent with the higher age-adjusted incidence rates for colorectal cancer among men than among women [Abotchie et al., 2012]. In addition, \textit{KRAS} G12D MF in normal colonic mucosa was positively correlated with age, although this correlation was statistically significant at only the 90% confidence level (Spearman $r = 0.4553$, $P = 0.0576$).

Table I and Figure 3 show there are differences in mutation prevalence and variability across normal tissue types. For example, all colonic mucosa samples had \textit{PIK3CA} H1047R and \textit{KRAS} G12D MFs $>10^{-5}$, whereas none of the normal colonic mucosa DNA samples had \textit{PIK3CA} E545K MFs $>10^{-5}$. MF measurements in normal tissues were analyzed statistically for tissue-specific differences (see Table I and Fig. 3). Because many \textit{PIK3CA} E545K mutations were below the limit of accurate ACB-PCR quantification ($10^{-5}$), statistical differences between normal tissue types were examined using a $\chi^2$ test (comparing the numbers of samples with MFs greater than and less than $10^{-5}$ across tissues). This analysis detected a significant difference in levels of \textit{PIK3CA} E545K mutation across the normal tissue types ($P = 0.0328$). Current \textit{PIK3CA} H1047R MF measurements in normal tissues were combined with those published previously. Because all \textit{PIK3CA} H1047R MFs were $>10^{-5}$, analysis of variance was performed using a
Kruskal-Wallis test, with a Dunn’s Multiple Comparison post-test. This analysis identified significant differences in PIK3CA H1047R MF across tissue types ($P = 0.0001$), and established that PIK3CA H1047R MFs in breast DNA samples were significantly greater than those of DNA from colonic mucosa.

This dataset was used to investigate relationships between tissue-specific MF measurements and the importance of particular mutations in tissue-specific carcinogenesis. Frequencies of the PIK3CA E545K, PIK3CA H1047R, KRAS G12D, and KRAS G12V mutations in carcinomas of the breast, large intestine, lung, and thyroid were collected from the expertly curated data in the COSMIC database (v76) [COSMIC, 2016] (see Supporting Information Table S1). We determined that the log 10 geo-mean MFs from the 16 sets of ACB-PCR MF measurements collected from normal tissues (PIK3CA E545K, PIK3CA H1047R, KRAS G12D, and KRAS G12V in breast, colonic mucosa, lung, and thyroid) were not correlated with the frequencies with which these mutations are reported in carcinomas from each organ (Spearman $r = 0.3676$, $P = 0.1612$).

Visual inspection of the data, however, suggested that mutations reported to occur frequently in particular cancers showed large inter-individual variability in the ACB-PCR MFs in normal tissues (e.g., PIK3CA H1047R for breast, KRAS G12D for colon, see Figs. 1 and 3) [COSMIC, 2016]. Therefore, the relationship between interindividual variability (as log$_{10}$ MF standard deviation) and the frequency of mutation within the various target organs was examined. Calculated MFs were used, except when the calculated MFs were below the theoretical limit of detection $3.33 \times 10^{-6}$ (a log$_{10}$ MF of $-5.477$), which is equivalent to one mutant molecule in a background of $3 \times 10^5$ WT molecules. In order to standardize the analysis of standard deviation across the data set and avoid overestimation of MF standard deviation due to non-detects, samples with a MF below $3.33 \times 10^{-6}$ were ascribed a MF of $3.00 \times 10^{-6}$ (log$_{10}$ MF $-5.481$, less than 1 mutant molecule per 300,000). The non-detects included 2/10 breast PIK3CA E545K measurements, 13/20 colon PIK3CA E545K measurements, 3/19 lung PIK3CA E545K measurements, 1 thyroid PIK3CA E545K measurement, and 1 breast KRAS G12D measurement. A strong correlation was observed between the log$_{10}$ MF standard deviation and the frequencies of the mutations in carcinomas from the corresponding organs (Spearman $r = 0.7265$, $P = 0.0014$). The correlation was significant with and without the correction for non-detects, but the correlation was stronger with the correction [Spearman $r$ (without correction) = $0.6382$, $P = 0.0078$]. Furthermore, the correlation was not driven by a single data point. Even when the breast PIK3CA H1047R data point (log$_{10}$ MF standard deviation = 1.2418; COSMIC tumor prevalence = 13.5%) was removed from the...
PIK3CA demonstrates (see Supporting Information, Table S6). The current study report for the first time a significant correlation exists between the degree of interindividual variability in PIK3CA and KRAS MFs in normal tissues and the tissue-specific prevalence of the mutations in carcinomas. Conversely, the log_{10} geomean MF did not correlate with the corresponding tumor mutation frequency. Because standard deviation is influenced by the magnitude of the measured variable, the variability among individuals samples combined with the magnitude of the tissue-specific normal MF measurements may be contributing to the observed correlation (Fig. 4). These observations suggest that some level of mutagenesis (an ACB-PCR measurable background MF)
combined with tissue-specific clonal expansion (inter-sample variability in MF measurements) are characteristics of cancer driver mutations within a particular tissue type. These observations have important implications in terms of establishing the relative importance of different mutations as biomarkers of susceptibility for particular human tissues. Further, this suggests that interindividual variation in MF in normal tissues can be used as a metric to assess the context-dependent, selective advantage of cells carrying oncogenic point mutations.

According to Rozhok and DiGregori, “oncogenic mutations should have vastly different fitness effects on somatic cells dependent on the tissue microenvironment in an age-dependent manner” [Rozhok and DiGregori, 2015]. ACB-PCR MF measurements provide evidence that PIK3CA and KRAS mutations impact mutant cell fitness in a manner that is context dependent. In the current study, we observed significant differences in PIK3CA E545K and H1047R mutation abundance across the different normal tissue types. We also observed a gender difference in PIK3CA H1047R MF in colonic mucosa. The significantly greater levels of PIK3CA H1047R mutation in male versus female colonic mucosa are consistent with men having a greater risk of developing colon cancer than women. PIK3CA H1047R mutation increases in abundance with age in normal breast, suggesting it confers a positive selective advantage in breast [Myers et al., 2016]. KRAS G12D and G12V mutations are more abundant in colonic adenomas, as compared to normal colonic mucosa, suggesting these mutations confer a selective advantage early in carcinogenesis. However, there is a significant decrease in KRAS G12V during adenoma to adenocarcinoma progression (see Fig. 1) [Parsons et al., 2010], and there is a significant inverse correlation between KRAS G12V MF and maximum tumor dimension for colon tumors and papillary thyroid carcinomas [Parsons and Myers, 2013], but not for lung adenocarcinomas [Myers et al., 2015]. These data suggest that KRAS G12V is neutral or selected against in some large/advanced cancers. This negative selection in large/advanced cancers could explain the prevalence of pre-existing KRAS mutant tumor cell subpopulations in the colorectal cancers that develop resistance to anti-EGFR monoclonal antibody therapy [Diaz et al., 2012; Misale et al., 2012].

There are a few reports of KRAS mutations being detected in normal colon and lung tissues [Ronai et al., 1994; Yakubovskaya et al., 1995; Sudo et al., 2006; Parsons et al., 2010]. Recently, Gao et al. [2017] reported detecting KRAS mutations in 52/156 histologically normal bronchial biopsies from former lung cancer patients. To our knowledge, there is only one report of PIK3CA mutation in normal tissue (breast) [Myers et al., 2016], although PIK3CA mutations have been detected in breast hyperplasia [COSMIC, 2016] and ductal hyperplasia with columnar cell change [Ang et al., 2012]. Data from the current and previous ACB-PCR analyses [Parsons et al., 2010; Myers et al., 2014a, 2015, 2016], allow us to conclude that cells carrying PIK3CA and KRAS hotspot point mutations are prevalent in normal tissues. The prevalence of PIK3CA and KRAS mutations in normal tissues is consistent with the mathematically-derived conclusion that “half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation” [Tomasetti et al., 2013, 2017].

There is an ongoing debate about the degree to which spontaneous mutation determines human cancer risk. Recent publications by Tomasetti and Vogelstein highlight the importance of spontaneous mutation in tissue-specific carcinogenesis [Tomasetti et al., 2013]; they suggest that two-thirds of cancers are attributable to random mutations in normal stem cells due to replication errors [Tomasetti and Vogelstein, 2015; Tomasetti et al., 2017]. Others contend that “cancer risk is heavily influenced by extrinsic factors,” with only 10 to 30% of cancers due to intrinsic factors [Wu et al., 2016]. Our results refute arguments that spontaneous mutation is insufficiently frequent to have an impact on carcinogenesis, as the combination of spontaneous mutation plus subsequent cell selection must give rise to the relatively high levels of mutations that we observed in normal tissues.

The concept of hotspot transdriver mutations has the potential to unify the opposing schools of thought regarding the importance of intrinsic versus extrinsic factors in carcinogenesis. Spontaneous transdriver mutations can be viewed as substrates that contribute to carcinogenesis, potentially though interaction with exogenously-induced genetic or epigenetic changes. Our analysis does not discriminate spontaneous from exogenously-induced mutation. Given the deficits in the smoking histories for the normal tissue samples, the lack of significant associations between MF and smoking history should be interpreted cautiously and revisited in the future. Nevertheless, given the relative insensitivity of epidemiological studies and the relatively small sample sizes analyzed by ACB-PCR, it seems unlikely that the significant correlation we observed between inter-individual variability in cancer driver MF and tissue-specific prevalence of the corresponding mutations in carcinomas is due solely to exogenous exposures. That said, the correlation depicted in Figure 4 may provide some insights into which cancers are impacted by exogenous exposures. The three data points where the cancer incidence is greater than the linear regression expectation correspond to PIK3CA E545K, KRAS G12D, and KRAS G12V in colorectal cancer, a cancer whose incidence is considered to have an extrinsic exposure component.

The remarkable prevalence of PIK3CA and KRAS mutations in DNA from normal tissues suggests that there may be significant opportunities for cancer prevention. Therapies that target pre-existing oncogenes in
normal tissues have the potential to decrease future cancer risk. Progress is being made in developing therapeutics that target mutant KRAS [Patricelli et al., 2016]. Therapies that target mutant KRAS or PIK3CA proteins could also be investigated as prophylactic therapies in individuals with Noonan and Cowden syndromes, who have high cancer risk because they carry germline mutations in these genes.

ACKNOWLEDGMENTS

The authors thank Drs. Page McKinzie, Todd Townsend, and Robert Hefflich for their helpful review of the manuscript. Tissue samples were provided by the Cooperative Human Tissue Network which is funded by the National Cancer Institute. Other investigators may have received specimens from the same subjects. The information in these materials is not a formal dissemination of information by FDA and does not represent agency position or policy.

AUTHOR CONTRIBUTIONS

B.L.P. and M.B.M. designed the study. K.L.M., M.B.M., and B.L.P. performed the experiments. B.L.P. and M.B.M. designed the study. K.L.M., and B.L.P. performed the experiments. B.L.P. prepared figures, tables and manuscript. All authors approved the final manuscript.

ABBREVIATIONS

ACB-PCR putative competitive blocker-polymerase chain reaction
CDMs cancer driver mutations
DC ductal carcinoma
MF mutant fraction
NGS next-generation sequencing
PTC papillary thyroid carcinoma
WT wild-type.

REFERENCES


Tomasetti C, Vogelstein B. 2015. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 347:78–81.


