

## An ultra-sensitive method to detect mutations in human *RAS* templates

Siqi Li<sup>a,b</sup> and Christopher M. Counter <sup>a</sup>

<sup>a</sup>Department of Pharmacology & Cancer Biology, Duke University Medical Center, Durham, NC, USA; <sup>b</sup>Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

### ABSTRACT

The *RAS* family of small GTPases is mutated in roughly a fifth of human cancers. Hotspot point mutations at codons G<sub>12</sub>, G<sub>13</sub>, and Q<sub>61</sub> account for 95% of all these mutations, which are well established to render the encoded proteins oncogenic. In humans, this family comprises three genes: *HRAS*, *NRAS*, and *KRAS*. Accumulating evidence argues that oncogenic *RAS* point mutations may be initiating, as they are often truncal in human tumours and capable of inducing tumorigenesis in mice. As such, there is great interest in detecting oncogenic mutation in the *RAS* genes to understand the origins of cancer, as well as for early detection purposes. To this end, we previously adapted the microbial ultra-sensitive *Maximum Depth Sequencing* (MDS) assay for the murine *Kras* gene, which was capable of detecting oncogenic mutations in the tissues of mice days after carcinogen exposure, essentially capturing the very first step in tumour initiation. Given this, we report here the adaptation and details of this assay to detect mutations in a human *KRAS* sequence at an analytic sensitivity of one mutation in a million independently barcoded templates. This humanized version of MDS can thus be exploited to detect oncogenic mutations in *KRAS* at an incredible sensitivity and modified for the same purpose for the other *RAS* genes.

### ARTICLE HISTORY

Received 29 March 2022  
Revised 20 April 2022  
Accepted 24 May 2022

### KEYWORDS



Small GTPases; *RAS*; ultra-sensitive sequencing; mutation; mutation detection

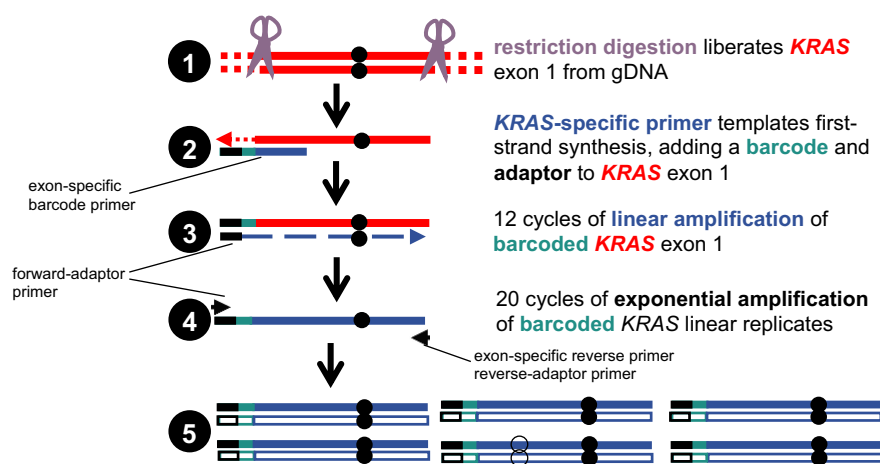
## Introduction

The *RAS* family of small GTPases comprises three genes, namely *HRAS*, *NRAS*, and *KRAS* in humans and *Hras*, *Nras*, and *Kras* in mice [1]. Canonical point mutations – primarily at codon G<sub>12</sub>, less so at Q<sub>61</sub>, and rarely at G<sub>13</sub> – are detected in upwards of a fifth of human cancers [2]. These mutations, through unique but also overlapping effects on the encoded protein, preferentially favour a GTP-loaded active state, which is well established to be oncogenic [3]. Of the three *RAS* genes, *KRAS* is the most commonly mutated in human cancers [2]. Three lines of evidence support oncogenic mutations in *KRAS* being an initiating event. First, these mutations have been detected in early-stage tumours. Second, sequencing efforts support these mutations being truncal. Third, activation of an inducible oncogenic *Kras* allele is able to induce tumorigenesis in mice [4–8]. Being able to identify oncogenic mutations in *KRAS* may therefore allow cancers to be detected early when they are traditionally more therapeutically tractable. However, there are three hotspot mutation sites (G<sub>12</sub>, G<sub>13</sub>, and Q<sub>61</sub>) with six potential amino acid substitutions per site (G<sub>12/13</sub>: V,D,C,S,R,A and Q<sub>61</sub>:L,R,K,H,E,P), which amounts to 18 possible oncogenic mutations [2]. As such, there is value in

capturing all possible mutations by sequencing *KRAS* exons directly. As an additional benefit, other alterations, such as tumour mutation signatures, could also be captured in the same sequencing reaction. However, the detection limit of next-generation sequencing (NGS) is in the range of one mutant per 100 templates or  $1 \times 10^{-2}$  [9]. We therefore sought to develop a more sensitive approach to sequence the *KRAS* locus.

To this end, we turned to the ultra-sensitive bacterial *Maximum Depth Sequencing* (MDS) assay, developed to detect the emerging mutations in a bacterial population challenged with antibiotic [10]. Two features of this assay increase the detection limit of NGS to one in million or  $1 \times 10^{-6}$ : first, directly integrating a barcode into one strand of a gene-of-interest followed by generating multiple linear amplification cycles of that barcoded genomic template, and second, extensively sequencing these products [10]. Detecting the same single-nucleotide variation (SNV) in each copy of the same barcoded template (versus detecting an SNV in only one copy) is then used to differentiate *bona fide* mutations from sequencing and PCR errors (Figure 1). We previously adapted this assay to sequence the first and/or second coding exon of murine *Kras* and *Hras* at an analytic sensitivity of  $5 \times 10^{-7}$ . Using this assay, we

**CONTACT** Christopher M. Counter  [count004@mc.duke.edu](mailto:count004@mc.duke.edu)  Department of Pharmacology & Cancer Biology, Duke University Medical Center, DUMC-3813, Durham, NC, USA 27710



**Figure 1.** MDS for a *KRAS* template. Unique barcodes and an adaptor are introduced into, for example, exon 1 of *KRAS* through restriction digestion (1) and first-strand extension (2), followed by linear amplification to copy the original DNA 12 times (3) and 20 cycles of exponential amplification (4) to expand the library for sequencing (5). Sequencing reads are grouped by barcode and *bona fide* mutations (●) separated from false ones (○) by virtue of being present in all reads sharing the same barcode. Adapted from Refs. [10,11].

were able to capture SNV in these genes as early as 7 days after mice were exposed to the carcinogen urethane. Moreover, by temporal analysis, we could differentiate oncogenic mutations that promoted tumour initiation versus those that did not. Specifically, we demonstrated that the number of independent barcoded templates with a CA<sub>182</sub>A>CT/GA point mutation in *Kras* encoding the oncogenic amino acid substitution Q<sub>61</sub>L/R increased over time, whereas those with the identical CA<sub>182</sub>A>CT/GA mutation encoding the non-oncogenic amino acid substitution Q<sub>70</sub>L/R did not. Finally, by sequencing these exons, MDS also captured the mutational signature of the carcinogen urethane [11,12]. Given that this murine version of MDS was capable of capturing SNVs in two Ras genes days after carcinogen exposure at an ultra-sensitive level of detection, we provide here a detailed description of this assay adapted for human DNA, a detailed description of how to test and validate the sensitivity of the assay for human RAS genes, and evidence that the assay is easily adapted for a human *KRAS* template.

## MDS optimized for a human template

### Overview

Based on the original MDS protocol [10] and modification made thereof for the murine template [11,12], we describe here the steps to apply MDS to a human template. In brief, genomic DNA is extracted from cells/tissues of interest and digested with restriction enzymes

that cut near the selected genomic region-of-interest (ROI). For example, to study mutations in *KRAS* codon 12, restriction enzyme *Stu*I could be used as it generates a 3' end on the transcribed strand of this gene 43 nucleotides away from codon G<sub>12</sub>. Restriction enzyme *Hinf*I could also be used as it generates a 3' end on the non-transcribed strand of *KRAS* gene 39 nucleotides away from codon G<sub>12</sub>. The exposed ROI is barcoded using a DNA barcode primer (exon-specific barcode primer) containing common adaptor, DNA barcode, and PCR primer sequence complementary to sequence beginning at the restriction enzyme-exposed end of ROI. Extra barcode primers are removed using exonuclease I and the barcoded ROI is linearly amplified using a primer annealing to the common adaptor region (forward-adaptor primer). Additional reverse PCR primers containing a common adaptor and PCR primer sequence for the other end of ROI (exon-specific reverse primer and reverse-adaptor primer) are added to the same reaction during the exponential amplification of the linear amplified products to produce the final PCR product containing Illumina sequencing adaptors and barcoded ROI. The resultant MDS library is subjected to NGS, and the frequency of SNV per barcodes to predict nucleotide changes is mapped to the original genomic template.

### Materials

- (1) Cell line/tissue of interest.
- (2) Cell/tissue lysis buffer (100 mM NaCl, 10 mM Tris pH 7.6, 25 mM EDTA pH 8.0, and 0.5%

- SDS in nuclease-free water, supplemented with 20 µg/ml RNase A (Sigma R4642)).
- (3) Proteinase K (NEB P8107S).
  - (4) Phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma P3803).
  - (5) 3 M Sodium acetate solution pH 5.2 (Sigma S7899).
  - (6) 100% Ethanol.
  - (7) Restriction enzyme targeting genomic region of interest with appropriate buffer (NEB).
  - (8) QIAquick PCR Purification Kit (Qiagen 28104).
  - (9) Q5 Hot Start High-Fidelity DNA Polymerase with appropriate buffer (NEB M0493).
  - (10) dNTP mix (NEB N0447).
  - (11) Exon-specific barcode primer (see example below for *KRAS* in the Validation section, Methods: MDS assay and library purification).
  - (12) Primers for amplification (forward-adaptor primer, exon-specific reverse primer, reverse-adaptor primer, see example below for *KRAS* in the Validation section, Methods: MDS assay and library purification).
  - (13) Exonuclease I (NEB M0293).
  - (14) AMPure XP beads (Beckman Coulter A63881).
  - (15) EB buffer (Qiagen 19086).
  - (16) Table-top centrifuge.
  - (17) NanoDrop spectrophotometer (ThermoFisher Scientific).
  - (18) Thermocycler.
  - (19) 0.2 ml PCR tubes.
  - (20) 1.5 ml microcentrifuge tubes.
  - (21) Heat block.
  - (22) Magnetic stand.
  - (23) NovaSeq 6000 system (Illumina).
  - (24) <https://usegalaxy.org/> [13]
  - (25) Microsoft Excel.

## Methods

### Isolation of genomic DNA

- (1) Suspend pelleted mammalian cells in lysis buffer (350 µl per  $5 \times 10^6$  cells). Alternatively, cut tissues into fine pieces using dissecting scissor and resuspend the tissues in lysis buffer (350 µl per 20 mg tissue).
- (2) Incubate the samples at 37°C for 1 h.
- (3) Add proteinase K (NEB P8107S) to each sample to a concentration of 5 units/ml. Vortex and incubate the sample at 55°C overnight.
- (4) Add 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma P3803) and vortex for approximately 20 s.
- (5) Centrifuge at room temperature for 5 min at 16,000 g. Transfer the upper aqueous phase to a fresh tube.
- (6) Repeat steps 4–5 until no protein is visible at the interface of the organic and aqueous phases.
- (7) Add 0.1 volume of sodium acetate (Sigma S7899) and vortex.
- (8) Add 2 volumes of ice-cold ethanol and vortex. Place the tube at –20°C overnight (or –80°C for 1 h) to precipitate the DNA.
- (9) Centrifuge the sample at 4°C for 15 min at 16,000 g.
- (10) Remove the supernatant and resuspend the DNA pellet in 500 µl of 70% ethanol. Centrifuge the sample at 4°C for 2 min at 16,000 g. Remove the supernatant.
- (11) Repeat step 10 once. Remove as much of the remaining ethanol as possible.
- (12) Air-dry the DNA pellet at room temperature for 5–10 min.
- (13) Dissolve the DNA pellet in 200 µl of nuclease-free water by pipetting. Determine the concentration of DNA using NanoDrop.

### MDS

- (1) Digest genomic DNA using desired restriction enzyme(s) chosen based on the ROI (see Overview) with 5 units enzyme per 1 µg DNA per 20 µl reaction at 37°C overnight.
- (2) Purify genomic DNA using QIAquick PCR Purification Kit (Qiagen 28104) following the manufacturer's protocol. Elute DNA using nuclease-free water and determine the concentration of purified DNA using NanoDrop.
- (3) Incubate purified DNA with the appropriate exon-specific barcode primer, as chosen based on the region of interest (see Overview) for 1 cycle of PCR. Use up to 10 µg DNA in a 50 µl reaction with 2.5 µl of 10 µM exon-specific barcode primer, 4 µl of 2.5 mM dNTP, 10 µl of 5X Q5 reaction buffer, and 0.5 µl Q5 Hot Start High-Fidelity DNA Polymerase (NEB M0493). Scale the number of reactions accordingly based on the total amount of genomic DNA obtained from each sample. Use the following PCR condition: 98°C for 1 min, exon-specific

barcode primer annealing temperature for 15 s, and 72°C for 1 min.

- (4) Add 1 µl of exonuclease I (NEB M0293) and 5 µl of 10X exonuclease I buffer to each 50 µl reaction, mix well by pipetting and incubate at 37°C for 1 h and then incubate at 80°C for 20 min.
- (5) Purify the barcoded genomic DNA using QIAquick PCR Purification Kit (Qiagen 28104) following the manufacturer's protocol. Elute DNA using nuclease-free water and determine the concentration of purified DNA using NanoDrop.
- (6) Perform linear amplification using the forward-adaptor primer (see Overview). Use up to 1.5 µg DNA in a 50 µl reaction with 2.5 µl of 10 µM forward-adaptor primer, 4 µl of 2.5 mM dNTP, 10 µl of 5X Q5 reaction buffer, and 0.5 µl Q5 Hot Start High-Fidelity DNA Polymerase (NEB M0493). Scale the number of reactions accordingly based on the amount of DNA obtained from each sample. Use the following PCR condition: 12 cycles of 98°C for 15 s, 70°C for 15 s, and 72°C for 8 s.
- (7) Perform exponential amplification using the exon-specific reverse primer and reverse-adaptor primer, as designed based on the region of interest (see Overview). Add 2.5 µl of 10 µM exon-specific reverse primer and 2.5 µl of 10 µM reverse-adaptor primer to each 50 µl reaction. Use the following PCR condition: 4 cycles of 98°C for 15 s, exon-specific reverse primer annealing temperature for 15 s, 72°C for 8 s, followed by 16 cycles of 98°C for 15 s, 70°C for 15 s, and 72°C for 8 s.

### Library purification

- (1) Add 1.1 volume of AMPure XP beads (Beckman Coulter A63881) to each sample (e.g., 55 µl for a 50 µl reaction), mix well by pipetting up and down at least 10 times, and incubate for 5 min at room temperature.
- (2) Place the tube on magnetic stand for 2 min to separate beads from supernatant. Transfer the supernatant to a new tube. Discard beads that contain the large genomic DNA fragments.
- (3) Add 0.7 volume of AMPure XP beads to each sample (relative to the starting reaction volume at step 1, e.g., 35 µl for a 50 µl reaction), mix well and incubate for 5 min at room temperature.
- (4) Place the tube on magnetic stand for 2 min to separate beads from supernatant. Remove and discard the supernatant.
- (5) Add 200 µl of freshly prepared 80% ethanol to the tube while on the magnetic stand. Incubate at room temperature for 30 s and then discard the supernatant.
- (6) Repeat step 5 once more.
- (7) Remove the residual ethanol and air-dry beads for 5–10 min.
- (8) Elute the PCR product from the beads using 15 µl EB buffer (Qiagen 19086) by pipetting and incubating at room temperature for 1 min.
- (9) Put the tube on the magnetic stand until the solution is clear and transfer the supernatant to a new tube.

### Sequencing and data analysis

- (1) Sequence the library using Illumina NGS platforms such as NovaSeq 6000 S Prime 150 bp PE. To obtain sufficient sequencing depth, aim for no less than 2 million reads for an input material equivalent to 1 million cells (~6.6 µg of genomic DNA).
- (2) Upload raw sequencing data to usegalaxy.org.
- (3) Use 'fastq-join' tool to join read 1 and read 2.
- (4) Use 'Filtered by quality' tool to filter reads based on quality score (90% of bases in sequence have quality score equal to/higher than 20).
- (5) Use 'Barcode Splitter' tool to split reads into files for individual samples based on assigned sample indexes.
- (6) Use 'FASTQ to FASTA' tool to convert files into FASTA format.
- (7) Use 'FASTA-to-Tabular' tool to convert files into Tabular format.
- (8) Use 'Trim' tool to trim the reads for each sample down to the barcode (file 1) and the target genomic region (file 2).
- (9) Use 'Column join' tool to join file 1 and file 2 into one table.
- (10) Use 'Group' tool to group the table by the barcode column with the following three operations on the target genomic region column: 'Count', 'Count Distinct' and 'Concatenate Distinct'.
- (11) Use 'Filter' tool to select rows with the 'Count' column larger than 1 ( $R \geq 2$ ) or 2 ( $R \geq 3$ ) and with the 'Count Distinct' column equal to 1 (sequences have 100% consensus within one barcode family).

- (12) Use 'Group' tool to group the filtered table by the target genomic region column ('Concatenate Distinct' column from step 10) while performing a 'Count' operation on the barcode column.
  - (13) Manually generate a reference Excel file of mutant sequences containing all possible single-nucleotide substitutions in the target genomic region with a separate column annotating these sequences by indicated nucleotide substitutions.
  - (14) Use 'VLOOKUP' function in Excel to compare sequences recovered from step 12 against the above reference to assign nucleotide substitution to sequences. Calculate the frequency of each substitution as the number of barcode families containing the mutation (the number in the 'Count' column) divided by the total number of barcode families (the sum of the numbers in the 'Count' column).
- (3) pcDNA™3.1 (+) plasmid (ThermoFisher Scientific V79020).
  - (4) Exon-specific barcode primer (see below: Validation, Methods: MDS assay and library purification).
  - (5) Primers (forward-adaptor primer, exon-specific reverse primer, reverse-adaptor primer) for amplification (see below: Validation, Methods: MDS assay and library purification).

## Validation assay

### Overview

To illustrate how MDS can be adapted to a human RAS template, we describe here MDS as applied to a panel of human *KRAS*-mutant cDNA templates spiked into human genomic DNA. Briefly, we generated a panel of *KRAS*-mutant plasmids, each comprising *KRAS* cDNA with a unique set of co-occurring double mutations in *KRAS* exon 1. Each was spiked at a specific concentration into genomic DNA isolated from 293T cells to benchmark different levels of sensitivity. As the error rates of PCR and sequencing are unlikely to give the same two exact improper base calls, the actual frequency of mutants present in the sample was estimated by calculating the frequency of barcode families with the pre-engineered co-occurring mutations. The frequency of mutations determined by MDS was then compared against the aforementioned actual frequency.

### Materials

Materials are identical to the above general assay, with the following specific reagents:

- (1) 293 T cell line (Duke CCF) was used as a source of genomic DNA.
- (2) *KRAS*-mutant plasmid cloning primers (see below).

## Methods

### Construction of *KRAS*-mutant plasmids

Mutant *KRAS* cDNA containing co-occurring mutations in *KRAS* exon 1 was made by PCR from plasmid containing wildtype *KRAS4B* cDNA [14] using forward primers containing the desired mutations of target nucleotides and a common reverse primer. The forward primers (mutated positions are in lower case) are as follows:

*KRAS* mutant F1: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGAA-TATAAACTTGTGGTAGTTGGAAcCaGGTG-3'

*KRAS* mutant F2: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGAA-TATAAACTTGTGGTAGTTGGAcCcGGTG-3'

*KRAS* mutant F3: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGAA-TATAAACTTGTGGTAGTTGGAtCgGGTG-3'

*KRAS* mutant F4: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGA-ATATAAACTTGTGGTAGTTGGcGaTGGTG-3'

*KRAS* mutant F5: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGA-ATATAAACTTGTGGTAGTTGGgGgTGGTG-3'

*KRAS* mutant F6: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGAA-TATAAACTTGTGGTAGTTGGtGtTGGTG-3'

*KRAS* mutant F7: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGA-ATATAAACTTGTGGTAGTTaaAGCTGGTG-3'

*KRAS* mutant F8: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGA-ATATAAACTTGTGGTAGTTccAGCTGGTG-3'

*KRAS* mutant F9: 5'-

CCGGGAATTCAGGCCTGCTGAAAATGACTGAA-TATAAACTTGTGGTAGTTtAGCTGGTG-3'

The common reverse primer is as follows:

*KRAS* XhoI R: 5'-

TTAACCTCGAGTTACATAATTACACACTTTGTCT-TTGA-3'.

PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (NEB M0493) with the

following condition: one cycle at 98°C for 30 s, 28 cycles at 98°C for 8 s, 61°C for 15 s, 72°C for 10 s, and one cycle at 72°C for 2 min. Generated PCR amplicon was cloned into pcDNA3.1 (ThermoFisher Scientific V79020) using restriction enzymes EcoRI (NEB R3101S) and XhoI (NEB R0146S).

### **MDS assay and library purification**

The series of plasmids containing *KRAS* cDNA with double mutations was spiked into 40 µg genomic DNA extracted from 293T cells at different concentrations aiming at dilutions from  $1 \times 10^{-7}$  to  $1 \times 10^{-4}$  (see exact dilutions in Table 1). MDS assay was performed as described above using StuI (NEB R0187) or HinFI (NEB R0155) as the restriction enzyme, 60°C as exon-specific barcode primer annealing temperature, and 61°C as the exon-specific reverse primer annealing temperature. Primer sequences are as follows:

- Exon-specific barcode primer: [Forward adaptor] [Index][Barcode][Primer] *where*  
 [Forward adaptor] = 5'-  
 TACGGCGACCACCGAGATCTACACTCTTTC-  
 CCTACACGACGCTCTTCCGATCT-3'  
 [Index] = variable length of known sequences from 0 to 7 nucleotides (specifically, 0 nucleotide, A, GA, CGA, TCGA, ATCGA, GATCGA or CGATCGA)  
 [Barcode] = 14 random nucleotides (NNNNNNNNNNNNNNN)  
 [Primer] for *KRAS* exon 1 non-coding strand StuI = 5'- CCTGCTGAAAATGACTGAA -3'  
 [Primer] for *KRAS* exon 1 coding strand HinFI = 5'- CTGAATTAGCTGTATCGTCAAG -3'
- Forward-adaptor primer: 5'-  
 AATGATACGGCGACCACCGAGAT-3'
- Exon-specific reverse primer: [Reverse adaptor] [Index][Primer] *where*  
 [Reverse adaptor] = 5'-  
 CAAGCAGAAGACGGCATACGAGATGTGACT-  
 GGAGTTCAGA CGTGTGCTCTTCCGATCT-3'  
 [Index] = variable length of known sequences from 0 to 7 nucleotides (specifically, 0 nucleotide, A, GA, CGA, TCGA, ATCGA, GATCGA or CGATCGA)  
 [Primer] for *KRAS* exon 1 non-coding strand StuI = 5'- TAGCTGTATCGTCAAGGC -3'  
 [Primer] for *KRAS* exon 1 coding strand HinFI = 5'-ATGACTGAATATAAACTTGTGGTA GT -3'
- Reverse-adaptor primer: 5'-  
 CAAGCAGAAGACGGCATACGAGA-3'

Library purification was performed as described above.

### **Sequencing and data analysis**

- (1) Sequence the library using Illumina NGS platforms such as NovaSeq 6000 S Prime 150 bp PE. To obtain sufficient sequencing depth, aim for no less than 2 million reads for an input material equivalent to 1 million cells (~6.6 µg of genomic DNA).
- (2) Upload raw sequencing data to usegalaxy.org.
- (3) Use 'fastq-join' tool to join read 1 and read 2.
- (4) Use 'Filtered by quality' tool to filter reads based on quality score (90% of bases in sequence have quality score equal to/higher than 20).
- (5) Use 'Barcode Splitter' tool to split reads into files for individual samples based on assigned sample indexes.
- (6) Use 'FASTQ to FASTA' tool to convert files into FASTA format.
- (7) Use 'FASTA-to-Tabular' tool to convert files into Tabular format.
- (8) Use 'Trim' tool to trim the reads for each sample down to the barcode (file 1) and the two target nucleotides (file 2 and file 3).
- (9) Use 'Column join' tool to join the three files into one table.
- (10) Use 'Merge Columns' tool to merge barcode and target nucleotides into one column.
- (11) Use 'Group' tool to group the table by the merged column with the 'Count' operation on the merged column and 'Concatenate Distinct' operation on the barcode column and the target nucleotide columns. This will produce a file (file 4) with the first column as merged sequence of barcode and the two target nucleotides, the second column as the number of appearances of that merged sequence, the third column as the barcode sequence and the fourth and fifth columns as the two target nucleotides.
- (12) Use 'Group' tool to group file 4 by the barcode column (the third column) with the 'Sum' and 'Concatenate' operations on count column (the second column) and 'Concatenate' operation on the target nucleotide columns (the fourth and fifth columns). This will produce a file (file 5) with the first column as the barcode, the second column as the number of members in that barcode family, the third column as the number of appearances of distinct double nucleotide sequences for the two target nucleotides in

**Table 1.** Frequency of mutations engineered into *KRAS* cDNA templates spiked into 293 T genomic DNA.

Strand targeted	Clone ID	Mutation	Replicate 1		Replicate 2		Replicate 3	
			Expected frequency	Detected frequency	Expected frequency	Detected frequency	Expected frequency	Detected frequency
<b>Transcribed strand (StuI)</b>	1	G31A	4.93E-05	5.29E-05	2.09E-05	2.7E-05	3.65E-05	4.29E-05
		T33A		5.05E-05		2.13E-05		3.57E-05
	2	G31C	0.00021	0.000224	0.000116	0.000118	0.000256	0.000251
		T33C		0.000222		0.000119		0.000253
	3	G31T	2.77E-05	3.12E-05	1.37E-05	1.47E-05	2.63E-05	2.68E-05
		T33G		3.12E-05		1.47E-05		2.68E-05
	4	A30C	6.16E-07	8.01E-07	1E-07	1E-07	1.46E-06	1.79E-06
		C32A		4.01E-06		6.54E-06		3.57E-06
	5	A30G	0.000243	0.000244	0.000146	0.000144	0.000275	0.000278
		C32G		0.000244		0.000144		0.000277
	6	A30T	3.7E-06	3.2E-06	1.31E-06	1.64E-06	5.85E-06	8.04E-06
		C32T		4.01E-06		1.64E-06		7.14E-06
	7	G28A	8.62E-06	1.2E-05	1E-07	5.73E-06	8.04E-06	1.7E-05
		G29A		7.21E-06		1.64E-06		1.07E-05
	8	G28C	0.002256	0.002291	0.001347	0.001348	0.002187	0.002193
		G29C		0.002291		0.001348		0.002194
	9	G28T	1.23E-06	8.01E-07	1E-07	8.18E-07	7.31E-07	1.79E-06
		G29T		8.01E-07		1E-07		8.93E-07
<b>Non-transcribed strand (HinfI)</b>	1	G31A	1.61E-05	2.03E-05	9.81E-06	1.01E-05	1.76E-05	1.86E-05
		T33A		2.13E-05		1.05E-05		2.64E-05
	2	G31C	8.02E-05	8.29E-05	5.92E-05	6.54E-05	8.46E-05	8.9E-05
		T33C		8.66E-05		6.54E-05		8.9E-05
	3	G31T	1.28E-05	1.56E-05	5.7E-06	7.74E-06	1.15E-05	1.57E-05
		T33G		1.15E-05		5.42E-06		1.14E-05
	4	A30C	1E-07	1E-07	5.89E-05	6.47E-05	9.43E-05	9.72E-05
		C32A		3.38E-07		6.47E-05		9.72E-05
	5	A30G	0.00023	8.53E-05	9.49E-07	1.16E-06	4.99E-06	4.64E-06
		C32G		8.56E-05		3.49E-06		5.36E-06
	6	A30T	1.64E-06	2.37E-06	6.33E-07	1E-07	5.88E-07	7.14E-07
		C32T		4.4E-06		1E-07		7.14E-07
	7	G28A	3.28E-06	3.38E-06	1.58E-06	1.94E-06	2.94E-06	2.86E-06
		G29A		3.72E-06		1.55E-06		2.86E-06
	8	G28C	0.00075	0.000792	0.000581	0.000624	0.0008	0.00083
		G29C		0.000792		0.000625		0.00083
	9	G28T	5.46E-07	5.41E-06	6.33E-07	7.74E-07	1.18E-06	4.29E-06
		G29T		1.35E-06		1.16E-06		2.14E-06

that barcode family, and the fourth and fifth columns as the first and second target nucleotide in each of those sequences.

- (13) File 5 was further analysed using Excel to obtain expected mutant frequency and detected mutant frequency. Expected mutant frequency was calculated by dividing the counts of families with engineered double mutations (determined from the third, the fourth and the fifth columns) by the total number of families for families containing  $\geq 1$  reads (determined from the second column) and have at least 90% consensus within the family (determined from the second and the third columns). Detected mutant frequency was calculated for individual nucleotide positions by dividing the counts of families containing engineered mutation at that position (determined from the third, the fourth and the fifth columns) by the total number of families for families containing  $\geq 2$  reads (determined from the second column) and have at least 90% consensus within the family (determined from the second and the third columns).

### Validation with a human *KRAS* cDNA template

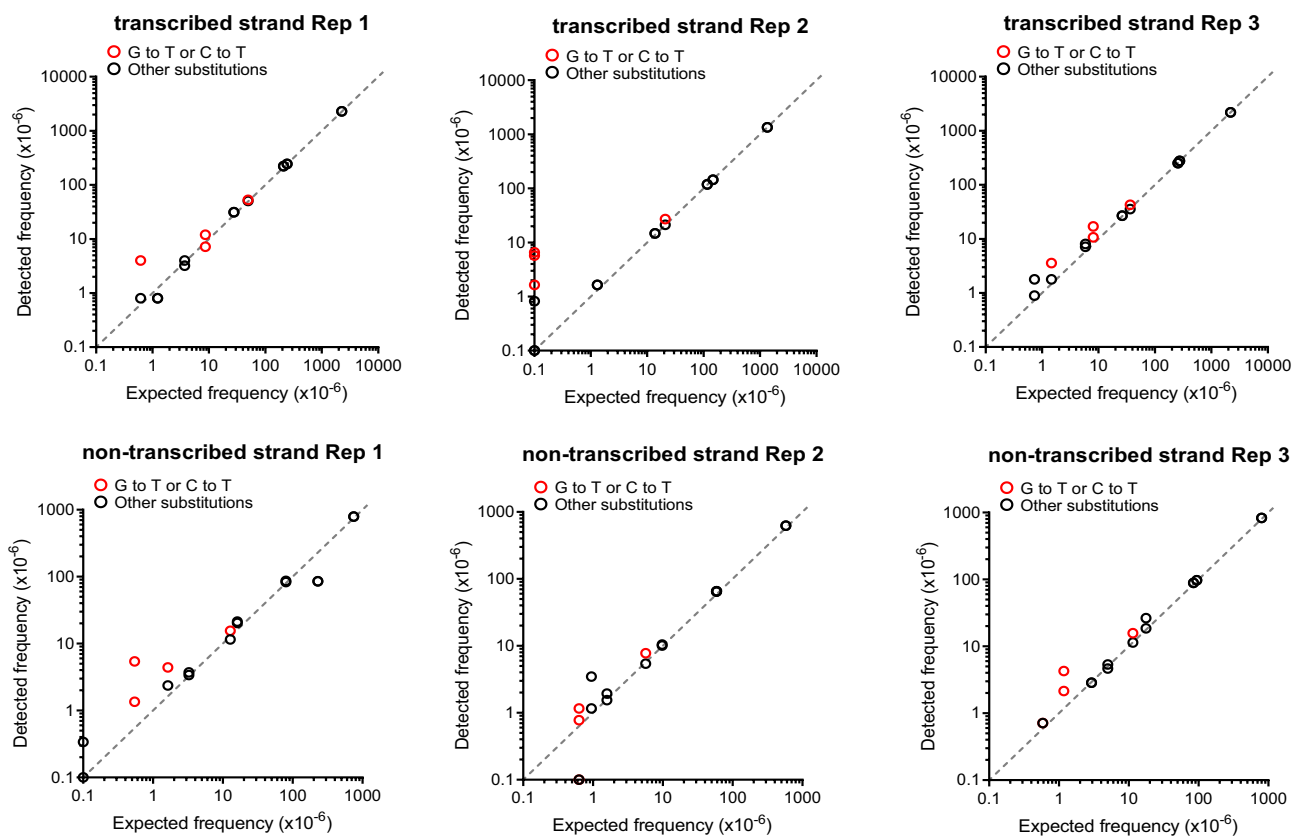
A series of plasmids containing *KRAS* cDNA with double mutations was spiked into 40  $\mu\text{g}$  genomic DNA extracted from 293T cells at different concentrations aiming at dilutions from  $1 \times 10^{-7}$  to  $1 \times 10^{-4}$  (see exact dilutions in Table 1). The genomic DNA mix was processed with the MDS assay using *StuI* or *HinfI* as the restriction enzyme. *StuI* was chosen as it generates a 3' end on the transcribed strand of *KRAS* gene close to codon G<sub>12</sub> (43 nucleotides away). The exon-specific region in the exon-specific barcode primer for *StuI* digested genomic DNA was chosen as it anneals to the exposed 3' end and satisfies conventional PCR primer parameters (19 nucleotides, GC 42%). The exon-specific region in the exon-specific reverse primer for *StuI*-digested genomic DNA was chosen as it anneals to a region close to *KRAS* codon G<sub>12</sub> (15 nucleotides away) and satisfies conventional PCR primer parameters (18 nucleotides, GC 50%). *HinfI* was chosen as it generates a 3' end on the non-transcribed strand of *KRAS* gene close to codon G<sub>12</sub> (39 nucleotides away). The exon-specific region in the exon-specific barcode primer for *HinfI* digested genomic DNA was chosen as it anneals to

the exposed 3' end and satisfies conventional PCR primer parameters (22 nucleotides, GC 41%). The exon-specific region in the exon-specific reverse primer for *HinfI*-digested genomic DNA was chosen as it anneals to a region close to *KRAS* codon G<sub>12</sub> (7 nucleotides away) and satisfies conventional PCR primer parameters (26 nucleotides, GC 31%). The advantage of performing MDS on both transcribed and non-transcribed strand is that mutations caused by DNA damages during library prep processes could be ruled out. Specifically, real mutations that were present in both strands of genomic DNA should be detected from MDS assay targeting both strands, while mutations originate from experimental artefacts, such as G to T mutation due to oxidation of guanine or C to T mutation due to deamination of cytosine or methyl-cytosine [15,16], will only be detected from the assay targeting the strand containing the damaged nucleotide. Comparison of the frequency of mutations determined by MDS against the actual frequency calculated estimated from engineered co-occurring mutations shows that MDS accurately detects *KRAS* cDNA mutants spiked into 293T genomic DNA with a sensitivity down to  $1 \times 10^{-6}$

(Figure 2). G to T or C to T mutations at or below  $1 \times 10^{-6}$  frequency could not be detected accurately on the strand where the G or C is located (Figure 2), presumably reflecting mutations introduced during library prep process due to DNA damage.

### Adapting MDS assay for other ROI

There are two major considerations when adapting MDS assay for the detection of mutations in other ROI. First, selecting suitable restriction enzymes. Restriction enzyme should cut near ROI, release ROI in a fragment purifiable using Qiagen columns (100 bp to 10 kb), and cut genomic DNA efficiently (varies by enzyme and should be validated for individual enzymes). Second, optimizing PCR conditions. There is little flexibility with the design of the exon-specific barcode primers, as they must anneal to the end of ROI exposed by the restriction enzyme. Exon-specific reverse primers, however, could be adjusted based on the desired length of PCR product and other factors commonly considered for PCR primers (e.g., GC content, melting temperature, length, lack of secondary structure, etc.).



**Figure 2.** MDS accurately detects *KRAS* cDNA mutants spiked into 293T genomic DNA. Frequency of single (detected) versus co-occurring (expected) mutations identified by MDS using a dilution series of *KRAS* cDNAs containing two different mutations engineered in *KRAS* exon 1 mixed with genomic DNA from 293T cells. Three biological replicates for MDS assay performed on the transcribed strand of *KRAS* exon 1 are show on the top and three biological replicates for MDS assay performed on the non-transcribed strand of *KRAS* exon 1 are show on the bottom.



PCR annealing temperature could also be adjusted using the temperature provided by NEB Tm calculator for Q5 Hot Start High-Fidelity DNA Polymerase as a starting point. In addition, Q5 High GC Enhancer could be added to the reaction to improve amplification of GC-rich regions. The ultimate goal is to increase the efficiency of each step to maximize the recovery of barcodes and to enable amplification of ROI using a minimum number of cycles of exponential PCR.

## Summary

Here, we describe the protocol to perform the MDS assay in general and an example of how to identify a restriction enzyme and design primers to detect SNV in a human *KRAS* cDNA template, which we show detects engineered mutations in *KRAS* exon 1 with a sensitivity down to  $1 \times 10^{-6}$ . G to T or C to T mutations at or below  $1 \times 10^{-6}$  frequency could not be detected accurately, potentially due to DNA damage introduced during library prep process. This issue could be solved by accurately detecting the frequency of matching C to A or G to A mutations using MDS assay targeting the other strands of DNA. For basic research applications, this approach may be useful in monitoring the emergence or competition between different RAS mutants in experimental systems. Another feature of this assay is that SNVs throughout the sequenced exon can be detected, and as such, MDS may similarly find value in detecting the presence of carcinogen exposure vis-à-vis their defined mutational signatures. Clinically, perhaps MDS may find use in capturing *de novo* mutations in *KRAS* or modified for such detection in other genes to capture expansion of resistant clones to therapeutic treatments or in early detection settings of high-risk patients.

## Disclosure statement

The authors report a patent pending on the application adapting the MDS methodology to detect mutations in exons 2 and 3 of *KRAS*.

## Funding

This work was supported by the National Cancer Institute (USA) under grants R01CA94184, R01CA269272, and P01CA203657.

## ORCID

Christopher M. Counter  <http://orcid.org/0000-0003-0748-3079>

## References

- [1] Colicelli J. Human RAS superfamily proteins and related GTPases. *Sci STKE*. 2004;2004(250):RE13.
- [2] Prior IA, Hood FE, Hartley JL. The frequency of Ras mutations in cancer. *Cancer Res*. 2020;80(14):2969–2974.
- [3] Fernandez-Medarde A, De Las Rivas J, Santos E. 40 years of RAS – a historic overview. *Genes (Basel)*. 2021;12(5):681.
- [4] Li S, Balmain A, Counter CM. A model for RAS mutation patterns in cancers: finding the sweet spot. *Nat Rev Cancer*. 2018;18(12):767–777.
- [5] Jackson EL, Willis N, Mercer K, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev*. 2001;15(24):3243–3248.
- [6] Haigis KM, Kendall KR, Wang Y, et al. Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. *Nat Genet*. 2008;40(5):600–608.
- [7] Braun BS, Tuveson DA, Kong N, et al. Somatic activation of oncogenic Kras in hematopoietic cells initiates a rapidly fatal myeloproliferative disorder. *Proc Natl Acad Sci USA*. 2004;101(2):597–602.
- [8] Hingorani SR, Petricoin EF, Maitra A, et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell*. 2003;4(6):437–450.
- [9] Fox EJ, Reid-Bayliss KS, Emond MJ, et al. Accuracy of next generation sequencing platforms. *Next Gener Seq Appl*. 2014;1:1000106.
- [10] Jee J, Rasouly A, Shamovsky I, et al. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature*. 2016;534(7609):693–696.
- [11] Li S, MacAlpine DM, Counter CM. Capturing the primordial Kras mutation initiating urethane carcinogenesis. *Nat Commun*. 2020;11(1):1800.
- [12] Li S, Counter CM. Signaling amplitude molds the Ras mutation tropism of urethane. *Elife*. 2021;10:e67172.
- [13] Jalili V, Afgan E, Gu Q, et al. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res*. 2020;48(W1):W395–W402.
- [14] Lampson BL, Pershing NL, Prinz JA, et al. Rare codons regulate KRas oncogenesis. *Curr Biol*. 2013;23(1):70–75.
- [15] Schmitt MW, Kennedy SR, Salk JJ, et al. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci USA*. 2012;109(36):14508–14513.
- [16] Arbeithuber B, Makova KD, Tiemann-Boege I. Artfactual mutations resulting from DNA lesions limit detection levels in ultrasensitive sequencing applications. *DNA Res*. 2016;23(6):547–559.