



Published in final edited form as:

Gastroenterology. 2019 September ; 157(3): 884–887.e3. doi:10.1053/j.gastro.2019.05.058.

Noninvasive Detection of High-risk Adenomas Using Stool-derived Eukaryotic RNA Sequences as Biomarkers

Erica K. Barnell^{1,2,*}, Yiming Kang^{2,3,*}, Elizabeth M. Wurtzler², Malachi Griffith^{1,4,5,6}, Aadel A. Chaudhuri^{3,6,7,+}, Obi L. Griffith^{1,4,5,6,+}, Andrew Barnell² [Geneoscopy Scientists], Katie Campbell^{1,2} [Geneoscopy Scientists], Kimberly R. Kruse² [Geneoscopy Scientists]

¹McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA

²Geneoscopy LLC, Division of Gastroenterology and Hepatology, St. Louis, MO, USA

³Department of Computer Science and Engineering, Washington University, St. Louis, MO, USA

⁴Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

⁵Division of Oncology, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

⁶Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA

⁷Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO, USA

Keywords

precision oncology; early cancer detection; high-risk adenoma; stool RNA

*Correspondence: Obi L Griffith, PhD, 4444 Forest Park Ave, St. Louis, MO, 63108, Campus Box 8501, obigriffith@wustl.edu, Aadel A Chaudhuri, MD PhD, 4511 Forest Park Ave, St. Louis, MO, 63108, 314-273-2931, aadel@wustl.edu.

Author Contributions

EB and YK contributed to the study concept and design; EB, YK, AB, KC, and EW contributed to the acquisition, analysis, and interpretation of data; EB, YK, AB, KC, KK, and EW contributed to drafting of the manuscript; MG, AC, and OG contributed to critical revision of the manuscript; EB, YK, AB, and KC contributed to statistical analysis; EB, YK, and AB obtained funding; MG, AC, and OG contributed to study supervision. All authors have seen and approved the final draft.

*These authors contributed equally to this work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Potential Competing Interests

These authors disclose the following: Erica Barnell, Yiming Kang, Andrew Barnell, Katie Campbell, and Elizabeth Wurtzler are inventors of the intellectual property owned by Geneoscopy. Erica Barnell, Yiming Kang, and Andrew Barnell are owners of Geneoscopy. Aadel Chaudhuri is a scientific advisor for Geneoscopy. Andrew Barnell, Elizabeth Wurtzler, and Kimberly Kruse are employees of Geneoscopy. Aadel Chaudhuri is a scientific advisor/consultant for Roche Sequencing Solutions and Tempus Labs, has received speaker honoraria and travel support from Varian Medical Systems, Roche Sequencing Solutions and Foundation Medicine, receives research support from Roche Sequencing Solutions, has served as a consultant for Tempus Labs and for Oscar Health, and is an inventor of intellectual property licensed to Biocognitive Labs. The remaining authors disclose no conflicts.

Introduction

Colorectal cancer (CRC) mortality¹ is often attributable to patient noncompliance with screening guidelines. While many noninvasive tests have been developed to address compliance issues, none compare to the diagnostic accuracy of colonoscopy.² Currently, the most accurate noninvasive diagnostic on the market (Cologuard DNA-FIT, Exact Sciences) cites a CRC sensitivity of 92%.³ However, the high-risk adenoma (HRA) detection rate is only 42%.³ Accurate detection of HRAs would permit preemptive excision of dysplastic tissue prior to carcinogenesis, thus reducing CRC incidence and associated mortality.⁴ Here we describe a method to reliably extract and evaluate stool-derived eukaryotic RNA (seRNA) transcripts for development of an algorithm that can noninvasively, sensitively, and specifically detect HRAs in a screening population. Full development of an assay that leverages seRNA biomarkers could facilitate noninvasive detection of HRAs and prevention of CRC.

Methods

Stool samples were prospectively collected from patients prior to preparing for and undergoing CRC screening via colonoscopy. In total, 26 patients had HRAs, 37 patients had medium-risk adenomas (MRAs), 61 patients had low-risk adenomas (LRAs), 50 patients had benign polyps, and 90 patients had no findings on colonoscopy (Supplementary Table 1). Isolated seRNA was subjected to targeted amplification using a custom panel of 639 amplicons (TruSeq Targeted RNA Custom Panel; Illumina, San Diego, CA) and next-generation sequencing (NextSeq 550; Illumina, San Diego, CA). Normalized expression of 639 amplicons was evaluated for all samples in the training set (n = 154 samples). Ten-fold internal cross-validation of the training set with independent feature selection within each fold was used to assess training model performance (n = 154 samples with 9:1 splits). A cut-off point for positive findings was determined by combining predictions from the subtesting sets into one receiver operating characteristic (ROC) curve and selecting a value to achieve an 85% specificity. Subsequently, final model features were selected using 100-fold bootstrapping of the entire training set (n = 154 samples) and an ordinal regression model was built (Figure 1A). This model was employed on a prospective hold out test set (n=110 unique samples). Hold out test set performance was measured by applying the previously defined cut-off point (Supplementary Methods).

Results

Technical replicates exhibited minimal difference in amplicon expression (Pearson r^2 average = 0.99); replicates subjected to varied enrichment strategies (200ng with 30 PCR cycles vs. 400ng with 28 PCR cycles) demonstrated an average Pearson r^2 correlation of 0.76; replicates subjected to independent sequencing runs demonstrated average Pearson r^2 correlation for expression of 0.73 (see Supplementary Methods). Using 100-fold bootstrapping of the training set (n=154 unique samples, 15 amplicons were identified as differentially expressed (informative in >25% of all bootstrapped splits) (Figure 1A). The 15 differentially expressed amplicons and raw *GAPDH* values were used to develop an ordinal regression model. Initial model performance was assessed through 10-fold internal cross-

validation of the training set. When comparing HRAs to all other findings (i.e., MRAs, LRAs, benign polyps, and no findings on a colonoscopy), model performance for all ten folds of internal cross-validation attained a ROC AUC of 0.70 (Figure 1B). A threshold value of 0.1415 was selected to be the cut-off point for a positive finding.

Model performance was subsequently tested by applying it to the prospective hold out test set (n=110 samples, each from a unique donor). Model output correlated with disease severity (1-way ANOVA; p-value=0.017), which was not provided as a feature for model training (Figure 1C). Upon ROC analysis, the ordinal regression model attained an AUC of 0.77 when comparing HRAs to all other findings. When employing the previously defined cut-off point of 0.1415 to the ROC curve, the model demonstrated a 45% sensitivity for HRAs (n=11 samples; 95%CI = 18.4% to 73.4%), 93% blended specificity for medium- and low-risk adenomas (n=40; 95%CI = 83.4% to 98.6%), 88% specificity for benign polyps (n=24; 95%CI = 72.6% to 97.5%), and an 80% specificity for no findings on a colonoscopy (n=35; 95%CI = 65.5% to 91.3%) (Figure 1D).

Discussion

The seRNA assay described herein attained a 45% sensitivity and 87% specificity for HRA detection. Model performance was increased in the hold out test set (AUC = 0.77) relative to internal cross-validation (AUC = 0.70), however, this difference was within the margin of error defined by the confidence intervals. Regarding our assay, seRNA offers several potential advantages compared to other stool- or blood-based biomarkers.⁵ First, seRNA biomarkers are derived from epithelial cells shed within the gastrointestinal tract. Therefore, the seRNA signal represents a homogenized sampling of perilesional tissue, which can be shed into the lumen and excreted in stool.⁶ Second, seRNA may provide a concentrated and amplified signal that can be observed via multiple transcripts in a single pathway.⁷ Finally, the RNA transcriptome can provide an assessment of the downstream molecular consequence of multiple precancerous variants that converge upon common tumorigenesis pathways. These characteristics enabled a relatively small panel of seRNA biomarkers to sensitively and specifically detect HRAs. HRAs are important to detect and remove due to an annual transition rate of HRA to CRC of 2.6–5.6%⁸, which implies that the cumulative risk for cancer transformation prior to the next screening recommendation is ~12% given a 3-year screening interval and ~40% given a 10-year screening interval.

Limitations of this study include use of a single organization (three geographically distinct endoscopy sites) for sample collection, use of a hold out test set obtained from the same collection sites, and the limited number of HRAs in our hold out test set. Additionally, the low incidence of CRC in a screening population made it challenging to prospectively obtain stool samples from CRC patients. Future research should include evaluation of these markers in a larger independent test set drawn from multiple sites. Nonetheless, these data provide evidence that seRNA biomarkers could significantly improve the ability to noninvasively detect HRAs, with potential to improve screening accuracy and compliance in the millions of Americans who are currently noncompliant with existing screening guidelines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work would not have been possible without dedication and commitment from the Geneoscopy Team: Andrew R. Barnell, Katie M. Campbell, and Kimberly R. Kruse. We would like to thank our scientific advisors: Dave Messina, Ira Kodner, and Phil Needleman. This research would not have been possible without support and assistance from Accelerate St. Louis, Arch Grants, BioSTL, BioGenerator Labs, Sling Health St. Louis, The Global Impact Award, The Skandalaris Center, Cofactor Genomics, The Wharton School, and Washington University School of Medicine.

Financial Contributions

This research was funded by Geneoscopy LLC (St. Louis, MO), which is a startup company developed from technology at the Washington University School of Medicine and grants from Accelerate St. Louis, Arch Grants, BioGenerator Labs, The Global Impact Award, and The Wharton School. Malachi Griffith is supported by the NHGRI under award number K99HG007940. Aadel Chaudhuri is supported by the NCI under award number K08CA238711, the Washington University Pancreas SPORE Career Enhancement Program under grant number P50CA196510 from the NCI, and the Cancer Research Foundation Young Investigator Award. Obi Griffith is supported by the NCI under award number K22CA188163. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Siegel RL, Miller KD & Jemal A Cancer statistics, 2018. *CA Cancer J. Clin* 68, 7–30 (2018) [PubMed: 29313949]
2. Inadomi JM Screening for Colorectal Neoplasia. *N. Engl. J. Med* 376, 149–156 (2017). [PubMed: 28076720]
3. Imperiale TF et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *N. Engl. J. Med* 370,1287–1297 (2014). [PubMed: 24645800]
4. Screening for Colorectal Neoplasia. *N. Engl. J. Med* 376, 1598–1600 (2017).
5. Xi X et al. RNA Biomarkers: Frontier of Precision Medicine for Cancer. *Noncoding RNA* 3, (2017).
6. Diehl F et al. Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. *Gastroenterology* 135, 489–498 (2008). [PubMed: 18602395]
7. Dickinson BT, Kisiel J, Ahlquist DA & Grady WM Molecular markers for colorectal cancer screening. *Gut* 64, 1485–1494 (2015). [PubMed: 25994221]
8. Brenner H et al. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. *Gut* 56, 1585–1589 (2007). [PubMed: 17591622]

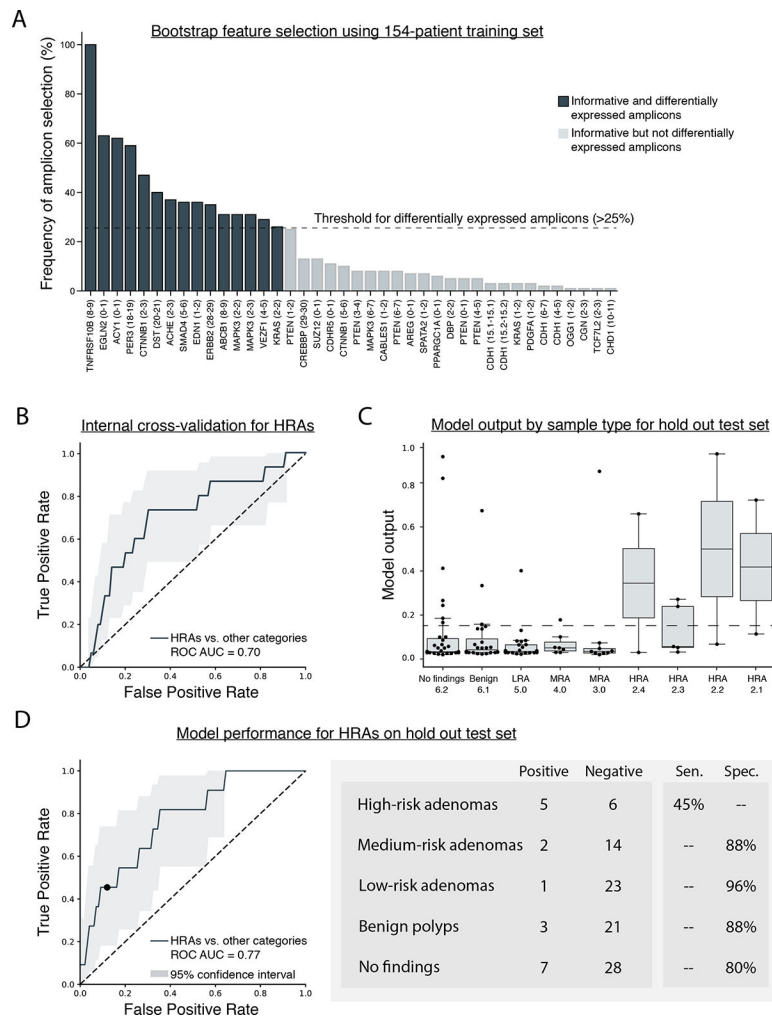


Figure 1. Eligible feature selection using bootstrapping of the training set (n = 154 samples) and model performance for the detection of HRAs based on 10-fold internal cross-validation and performance on a prospective hold out test set (n = 110 samples).

A) Transcripts used in the custom amplicon panel (n = 639 amplicons) were selected based on previously conducted research and differentially expressed amplicons were identified using 100-fold bootstrapping of the 154-patient training set. If an amplicon was observed in at least 25% of all 100 splits (bootstrap threshold), then it was considered differentially expressed and was eligible as a feature for the final model. Each column represents a single amplicon denoted by the HUGO gene name with exon location of forward and reverse probes. In total, 15 amplicons on 14 unique genes were selected as differentially expressed.

B) 10-fold internal cross-validation was performed using the training set (n = 154 samples), 15 differentially expressed amplicons, and raw GAPDH values. The ROC curve shows model performance whereby high-risk adenomas (HRAs) were considered positive and other findings (medium-risk adenomas, low-risk adenomas, benign polyps, no findings on a colonoscopy) were considered negative. **C)** Box plots show model output for each sample, parsed by sample type, for the prospective hold out test set (n = 110 samples). Sample type is ascending based on lesion severity (i.e., No finding 6.2 = least severe, HRA 2.1 = most severe) (see Supplementary Table 1). Each dot represents a single sample employed in the

analysis. The box encases the first and third quartile of the dataset, the bar within the box represents the median value. Whiskers represent 1.5 times the interquartile range and values that extend beyond the length of the whiskers were considered outliers. The dashed line represents the threshold defined by internal cross-validation performance (0.1415) **D**) An ordinal regression model was created using the training set (n = 154 samples) and all 16 eligible features. The ordinal regression model was employed on the prospective hold out test set (n = 110 samples) to determine model performance. High-risk adenomas (HRAs) were considered positive and other findings (medium-risk adenomas, low-risk adenomas, benign polyps, no findings on a colonoscopy) were considered negative. Sensitivity is shown for HRAs and specificity is shown for all other findings. Each sample in the training set (n = 154) and hold out test set (n = 110) was from a unique donor. *Abbreviations: ROC – receiver operator characteristic; AUC – area under the curve; Sen. – sensitivity; Spec. – specificity; LRA – low-risk adenoma; MRA – medium-risk adenoma; HRA – high-risk adenoma.*

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript