Gastroenterology 2019;157:884–887

BRIEF COMMUNICATIONS

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. Wurtzler,[2] Malachi Griffith,[1,4,5,6] Aadel A. Chaudhuri,[3,4,6,7] Obi L. Griffith,[1,4,5,6] and Geneoscopy Scientists

[1]McDonnell Genome Institute, Washington University School of Medicine, St Louis, Missouri; [2]Division of Gastroenterology and Hepatology, Geneoscopy LLC, St Louis, Missouri; [3]Department of Computer Science and Engineering, Washington University, St Louis, Missouri; [4]Department of Genetics, Washington University School of Medicine, St Louis, Missouri; [5]Division of Oncology, Department of Medicine, Washington University School of Medicine, St Louis, Missouri; [6]Siteman Cancer Center, Washington University School of Medicine, St Louis, Missouri; and [7]Department of Radiation Oncology, Washington University School of Medicine, St Louis, Missouri

Keywords: Precision Oncology; Early Cancer Detection; High-Risk Adenoma; Stool RNA.

Colorectal cancer (CRC) mortality[1] 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, Madison, WI) 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 before 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 before preparing for and undergoing CRC screening via colonoscopy. In total, 26 patients had HRAs, 37 patients had medium-risk adenomas, 61 patients had low-risk adenomas, 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). 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 cutoff 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 1*A*). 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 cutoff point (Supplementary Methods, Supplementary Table 2.

Results

Technical replicates exhibited minimal difference in amplicon expression (Pearson r^2 average = 0.99); replicates subjected to varied enrichment strategies (200 ng with 30 polymerase chain reaction cycles vs 400 ng with 28 polymerase chain reaction [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 1*A*). 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 (ie, medium-risk adenomas, low-risk adenomas, benign polyps, and no findings on a colonoscopy), model performance for all 10 folds of internal cross-validation attained an ROC area under the curve (AUC) of 0.70 (Figure 1*B*). A threshold value of 0.1415 was selected to be the cutoff point for a positive finding.

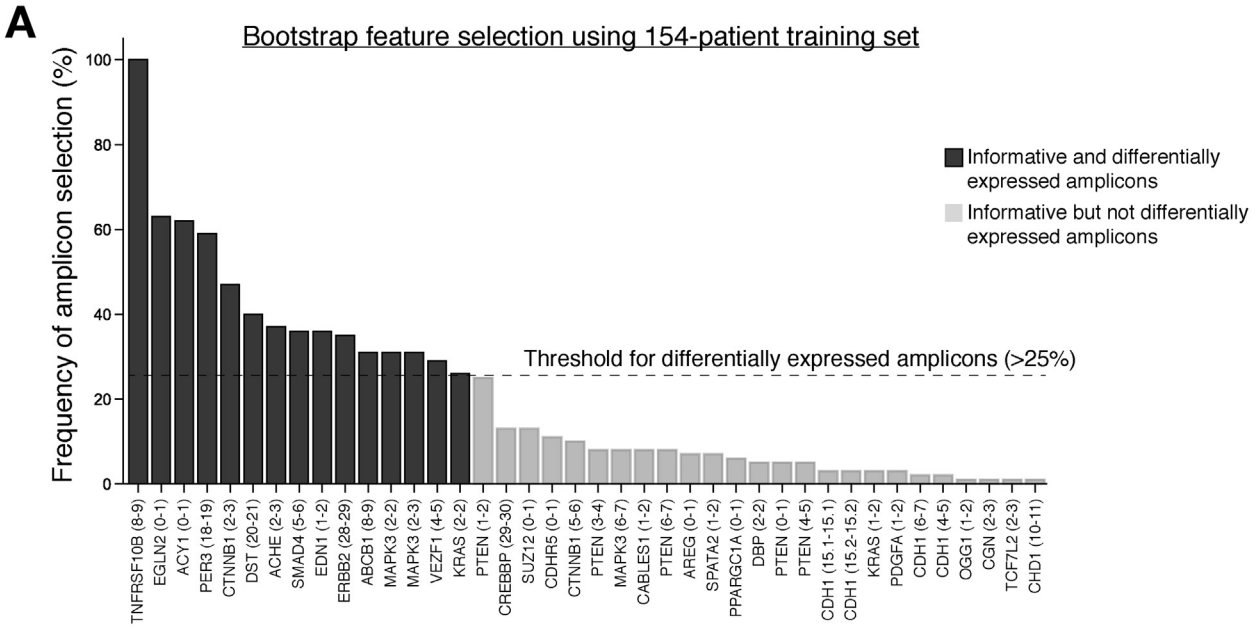
Model performance was subsequently tested by applying it to the prospective hold out test set (n = 110 samples, each from a unique donor). Model output correlated with disease

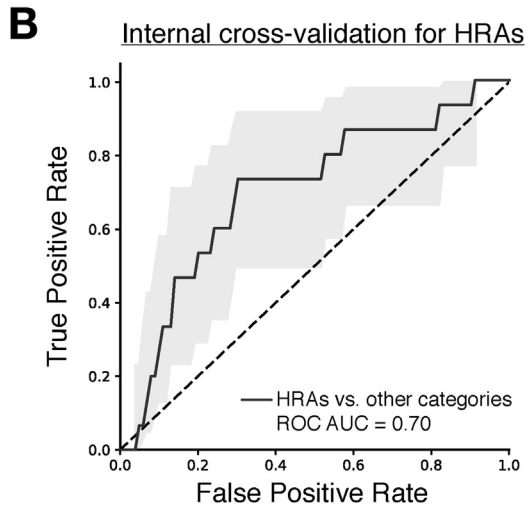
*Authors share co-first authorship.

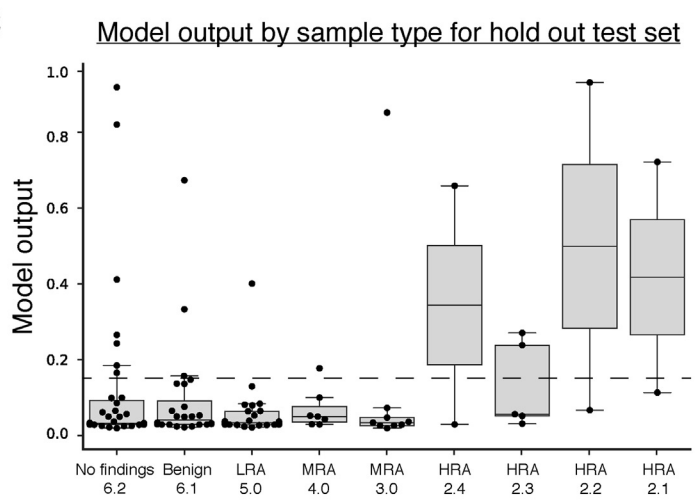
Abbreviations used in this paper: **AUC, area under the curve; CI, confidence interval; CRC, colorectal cancer; HRA, high-risk adenoma; ROC, receiver operating characteristic; seRNA, stool-derived eukaryotic RNA.**

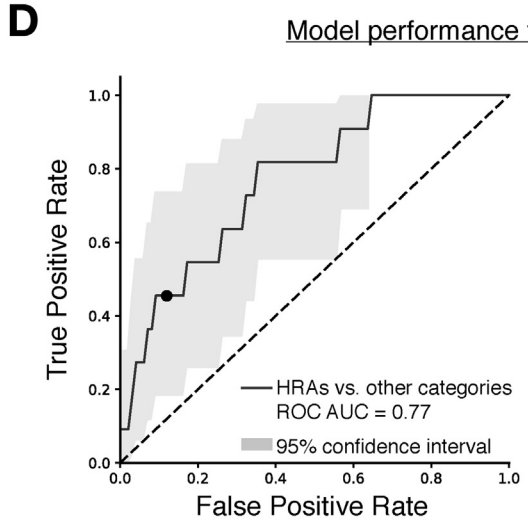
Most current article

0016-5085
https://doi.org/10.1053/j.gastro.2019.05.058









	Positive	Negative	Sen.	Spec.
High-risk adenomas	5	6	45%	--
Medium-risk adenomas	2	14	--	88%
Low-risk adenomas	1	23	--	96%
Benign polyps	3	21	--	88%
No findings	7	28	--	80%

severity (1-way analysis of variance; $P = .017$), which was not provided as a feature for model training (Figure 1*C*). 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 cutoff point of 0.1415 to the ROC curve, the model demonstrated a 45% sensitivity for HRAs (n = 11 samples; 95% confidence interval [CI] 18.4%–73.4%), 93% blended specificity for medium- and low-risk adenomas (n = 40; 95% CI 83.4%–98.6%), 88% specificity for benign polyps (n = 24; 95% CI 72.6%–97.5%), and an 80% specificity for no findings on a colonoscopy (n = 35; 95% CI 65.5%–91.3%) (Figure 1*D*).

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 CIs. 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 before the next screening recommendation is approximately 12%, given a 3-year screening interval, and approximately 40%, given a 10-year screening interval.

Limitations of this study include use of a single organization (3 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 detect HRAs noninvasively, with potential to improve screening accuracy and compliance for the millions of Americans who are currently noncompliant with existing screening guidelines.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at https://doi.org/10.1053/j.gastro.2019.05.058.

References

1. Siegel RL, et al. CA Cancer J Clin 2018;68:7–30.
2. Inadomi JM. N Engl J Med 2017;376:149–156.
3. Imperiale TF, et al. N Engl J Med 2014;370:1287–1297.
4. Dib J Jr. N Engl J Med 2017;376:1598–1600.
5. Xi X, et al. Noncoding RNA 2017;3(1).
6. Diehl F, et al. Gastroenterology 2008;135:489–498.
7. Dickinson BT, et al. Gut 2015;64:1485–1494.
8. Brenner H, et al. Gut 2007;56:1585–1589.

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*) Ten-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 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 (ie, no finding 6.2 = least severe, HRA 2.1 = most severe) (Supplementary Table 1). Each *dot* represents a single sample employed in the analysis. The *box* encases the first and third quartile of the data set, 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. 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 (Supplementary Table 2). LRA, low-risk adenoma; MRA, medium-risk adenoma; Sen., sensitivity; Spec., specificity.

Received February 5, 2019. Accepted May 28, 2019.

Reprint requests
Address requests for reprints to: Obi L. Griffith, PhD, McDonnell Genome Institute, Washington University School of Medicine, 4444 Forest Park Avenue, Campus Box 8501, St Louis, Missouri 63108. e-mail: obigriffith@wustl.edu; and Aadel A. Chaudhuri, MD, PhD, Department of Radiology Oncology, Washington University School of Medicine, 4511 Forest Park Avenue, St Louis, Missouri 63108. e-mail: aadel@wustl.edu.

Acknowledgments
This work would not have been possible without the dedication and commitment of the Geneoscopy Scientists: Andrew R. Barnell, Division of Gastroenterology and Hepatology, Geneoscopy LLC, St Louis, Missouri; Katie M. Campbell, McDonnell Genome Institute, Washington University School of Medicine, St Louis, Missouri and Division of Gastroenterology and Hepatology, Geneoscopy LLC, St Louis, Missouri; and Kimberly R. Kruse, Division of Gastroenterology and Hepatology, Geneoscopy LLC, St Louis, Missouri. We would like to thank our scientific advisors: Dave Messina, Ira Kodner, Mark Manary, and Phil Needleman. This research would not have been possible without the support and assistance of 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.

Author contributions: EKB and YK contributed to the study concept and design. EKB, YK, ARB, KMC, and EMW contributed to the acquisition, analysis, and interpretation of data. EKB, YK, ARB, KMC, KRK, and EMW contributed to drafting of the manuscript. MG, AAC, and OLG contributed to critical revision of the manuscript. EKB, YK, ARB, and KMC contributed to statistical analysis. EKB, YK, and ARB obtained funding. MG, AAC, and OLG contributed to study supervision. All authors have seen and approved the final draft.

Conflicts of interest
These authors disclose the following: Erica K. Barnell, Yiming Kang, Andrew R. Barnell, Katie M. Campbell, and Elizabeth M. Wurtzler are inventors of the intellectual property owned by Geneoscopy. Erica K. Barnell, Yiming Kang, and Andrew R. Barnell are owners of Geneoscopy. Aadel A. Chaudhuri is a scientific advisor for Geneoscopy. Andrew R. Barnell, Elizabeth M. Wurtzler, and Kimberly R. Kruse are employees of Geneoscopy. Aadel A. 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, and is an inventor of intellectual property licensed to Biocognitive Labs. The remaining authors disclose no conflicts.

Funding
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 National Human Genome Research Institute under award number K99HG007940. Aadel A. Chaudhuri is supported by the National Cancer Institute (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 L. 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 National Institutes of Health.

Supplementary Methods

Study Design

Stool samples were prospectively and consecutively collected from patients undergoing CRC screening prior to treatment or surgical resection. Patients with a family history of CRC were considered eligible for the study; however, patients who had a personal history of CRC (eg, previously diagnosed with CRC via colonoscopy) or any other inflammatory gastrointestinal disease (eg, inflammatory bowel disease) were considered ineligible. The Washington University School of Medicine (St Louis, MO) Institutional Review Board approved the protocol and research procedures (IRB #20111107). The primary outcome was the feasibility of using seRNA to assess risk for HRAs. These human RNA biomarkers are derived from the enterocytes that are sloughed from the lumen of the colon and excreted into the stool as whole cells. Isolated seRNA was subjected to targeted amplification (TruSeq Targeted RNA Custom Panel; Illumina) and next-generation sequencing (NextSeq 550; Illumina). Sequencing reads were used for transcript selection, model development, and model validation.

Eligible Patients and Sample Collection

Stool samples were obtained by the Digestive Diseases Research Core Center at Washington University School of Medicine. All patients were sent a stool sample collection kit by mail and returned the kit via courier to the Digestive Diseases Research Core Center. Clinical data (eg, demographic information, and colonoscopy results) were collected by the Digestive Diseases Research Core Center. Stool samples were evaluated for occult blood using a OC-Light S FIT (Polymedco, Cortlandt Manor, NY) before seRNA extraction. Each patient recruited for the study had a colonoscopy performed and those with positive findings underwent biopsy and subsequent histopathologic review to determine neoplastic classification. Adenoma classification was stratified based on histopathology (benign vs premalignant), number of polyps, size of polyps, and differentiation. If the patient had no findings during the colonoscopy, he or she was labeled as healthy. If a patient had multiple findings on their colonoscopy, the ultimate label represented the lesion with the highest severity of disease. No samples were excluded from the analysis based on colonoscopy result. The relative proportions of colonoscopy results were not significantly different from those observed in other prospective clinical trials.[1]

Development of a Training Set and a Testing Set

In total, 264 stool samples, each from a unique individual, were prospectively collected. One hundred and fifty-four stool samples were used as a training set and 110 stool samples were used as a hold-out test set. The training set and hold-out test set were evaluated for categorical, demographic, and handling differences using a *t* test (population means) or Fisher's exact test (population frequencies) and significance was indicated if the *P* value was $<.05$.

Stool Sample Enrichment for Human RNA From Exfoliated Enterocytes Using Differential Centrifugation

Stool samples were aliquoted into 50-mL conical tubes and filled to 45 mL using suspension buffer (10 mM Tris, 1 mM EDTA, 0.005% Tween-20, 80U RNase Inhibitor, pH 7.5). Samples were homogenized and subjected to differential centrifugation using a swing-bucket centrifuge for 10 minutes at 4°C to separate the homogenate into a human cell layer below an enriched bacterial supernatant. The supernatant was discarded to eliminate bacterial noise and the pellet was suspended into a guanidine thiocyanate buffer to lyse the enterocytes and expose the human biomarkers.[2] From the lysate layer containing enriched human nucleic acids, 2 mL of the solution was purified using a NucliSENS easyMag automated system (bioMérieux, Durham, NC).[3,4] The final solution was subjected to Baseline-ZERO DNase treatment (Epicentre Technologies, Madison, WI) and clean-up using the NucliSENS easyMag.[5] The RNA in the final solution was likely derived from the colorectum, however, the solution potentially had upper gastrointestinal and bacterial contamination.

Custom Amplicon Panel Development

The custom amplicon panel was developed using previously conducted research and literature review.[6] First, transcripts were selected based on a microarray experiment (Gene Expression Omnibus Accession #GSE99573).[6] For this experiment, total seRNA was extracted from stool samples and expression was assessed using the Affymetrix Human Transcriptome Array 2.0 (ThermoFisher Scientific, Waltham, MA). Microarray expression profiles derived from patients with CRC or premalignant adenomas (diseased cohort) were compared to expression profiles from patients with no findings on colonoscopy (healthy cohort). Transcripts with significant differential expression ($P < .03$) were selected for the amplicon panel. Additional transcripts were selected based on a NanoString experiment.[6] Again stool samples were obtained from a diseased cohort and a healthy cohort. Total seRNA was extracted from stool samples and expression was assessed using the nCounter PanCancer Pathways Panel (NanoString, Seattle, WA) and the nCounter PanCancer Progression Panel (NanoString). Differentially expressed transcripts were identified by comparing the diseased cohort to the healthy cohort using the nSolver differential expression analysis platform. Finally, the literature was evaluated for additional transcripts implicated in CRC. This included searching the ClinVar,[7] COSMIC,[8] and CIViC[9] databases, as well as other pertinent studies.[10,11] A custom amplicon panel was developed for targeted enrichment using the Illumina DesignStudio. Use of exons provided in Figure 1 can be used to guide amplicon development.

Stool-Derived Eukaryotic RNA Quality Check and Sequencing

The seRNA integrity and size was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and mass was determined using a Qubit Fluorometer (ThermoFisher Scientific). Samples required >200 ng of RNA to be eligible for library preparation. Libraries were prepared using a TruSeq Targeted RNA Custom Panel (Illumina). No complementary DNA fragmentation was required for sequencing. Sequencing was performed using the NextSeq 550 System (Illumina) with 2 × 150 bp reads. A PhiX spike-in was used for quality control. All samples in the training set were sequenced across 4 independent sequencing runs and all samples in the testing set were sequenced across 4 independent sequencing runs.

Quantification Stool-Derived Eukaryotic RNA Expression

Raw sequencing reads were aligned to the reference genome (GRCh38) via HISAT2 with default parameters, and counts were summarized within amplicon regions of interest.[12] Samples required >100,000 aligned reads for study eligibility. For each sample, raw amplicon expression was normalized to *GAPDH*, the internal housekeeping gene, such that reported expression equates to amplicon read count per million mapped *GAPDH* reads. Raw *GAPDH* values were used as a measure for total eukaryotic RNA in each sample. Raw *GAPDH* values were also eligible as a feature for model development. Normalized amplicon values and raw *GAPDH* values along with sample demographic information is available in Supplementary Table 2.

Replicate Analysis to Assess Transcript Expression Variability

Five patients were used to assess reproducibility of seRNA extraction and sequencing. Across these patients, 3 separate reproducibility experiments were performed: technical sequencing replicates, biological library preparation replicates, and biological sequencing replicates. Pearson *r* correlation was used to assess change in transcript expression for all replicates across all amplicons in the custom panel.

The technical sequencing replicate experiment assessed reproducibility of independent sequencing runs for transcript expression quantification. For this experiment, a single sample was subjected to extraction and library preparation. Subsequently, 2 separate aliquots of the library preparation were sequenced on different sequencing runs.

The biological library preparation replicate experiment assessed reproducibility of 2 different library preparations for transcript expression quantification. For this experiment, a stool sample from the same bowel movement was split into 2 aliquots that were homogenized and extracted separately. One sample used 400 ng of library input and was subjected to 28 cycles of polymerase chain reaction amplification. The second sample used 200 ng of library input and was subjected to 30 cycles of polymerase chain reaction amplification. Both replicates underwent parallelized sequencing on the same sequencing run.

The biological sequencing replicate experiment assessed reproducibility of independent sequencing runs for transcript expression quantification. For this experiment, a stool sample from the same bowel movement was split into 2 aliquots that were homogenized and extracted separately. Both samples used 400 ng of library input and were subjected to 30 cycles of polymerase chain reaction amplification. The samples were subsequently sequenced on independent sequencing runs.

Parameters for Ordinal Regression Model Development

Linear ordinal regression models were trained using logistic loss in combination with the All-Threshold loss function[13,14] to respect the ordinality of predicted labels (or the progression of disease). Progression of disease included negative findings (no findings on a colonoscopy, benign polyps, low-risk adenomas), medium-risk adenomas, and HRAs. Distance between the ordinal predicted labels were equally distant. The optimization methods employed were L-BFGS-B, an extension of the limited-memory BFGS optimization algorithm. A Python implementation of the algorithm mord (version 0.3) was employed in our analysis.[14] Other software used included Python (version 3.6.5), scipy (version 1.0.0), and sklearn (version 0.20.0).

Assessment of Internal Cross-Validation Performance

Ten-fold internal cross-validation was performed using bootstrapping of the training set (n = 154 samples). Specifically, the training set was segregated into 10 different 9:1 splits (ie, subtraining and subtesting sets), whereby feature selection was performed for each split. A feature was eligible for model development if the absolute $\log_2$ fold-change was >1 in both contrast groups (HRA vs low-risk adenomas, benign polyps, no findings on a colonoscopy; medium-risk adenomas vs low-risk adenomas, benign polyps, no findings on a colonoscopy) and the analysis of variance between the contrast groups had a *P* value <.05. Within each split, the subtraining set and selected features were used to build an ordinal regression model. Model performance was assessed by employing this model on the subtesting set. Output from the model provided a prediction between 0 and 1 for each category (normal, medium-risk adenoma, and HRA), whereby a larger number reflected increased confidence in a positive finding. Predictions on samples from each fold of the internal cross-validation for all splits were compared to results from colonoscopy and ROC curves were developed. A Bayesian approach was used to compute the 95% CI for the ROC curve.[15] Specifically, for each sensitivity point estimate on the ROC curve, a posterior distribution was calculated by employing the β function (prior β used = β [0,0]). The 95% CI was bound between 2.5 and 97.5 percentiles of the posterior distribution. Finally, an optimal threshold was developed to attain at least 85% specificity for all negative findings (MRAs, LRAs, benign polyps, no findings on a colonoscopy). This cutoff point was used to assess model performance for HRA sensitivity.

Assessment of Hold-Out Test Set Performance

To further assess performance, a new model was built using all samples in the training set and employed on the hold-out test. Feature selection for model development was performed using bootstrapping of the training set (n = 154 samples). Specifically, the training set was segregated into 100 different 9:1 splits, whereby each split was assessed for informative amplicons. An amplicon was considered informative if the absolute $\log_2$ fold-change was >1 in both contrast groups (HRA vs LRAs, benign polyps, no findings on a colonoscopy; MRAs vs LRAs, benign polyps, no findings on a colonoscopy) and the analysis of variance between the contrast groups had a P value < .05. If an amplicon was deemed informative in at least 25% of all bootstrapped splits, it was considered differentially expressed and eligible as a feature for model development. This threshold was set by evaluating the performance of models with inclusion of the most informative transcripts and maximizing ROC AUC. Raw *GAPDH* values were also evaluated for significance between contrast groups and were considered eligible for model development. CI development was performed according to the methods described here.

An ordinal regression model was built using all 154 samples in the training set and all eligible features.[13] Specifically, this model penalizes the false prediction proportionally according to the distance between prediction and true label (ie, the further away a false prediction is from truth, the more penalty). The goal of training is to minimize the sum of such penalties. Output from the model provided a prediction between 0 and 1 for each category (normal = 0, medium-risk adenoma = 0.5, and HRA = 1), whereby a larger number reflected increased confidence in a positive finding. Predictions on samples in the hold-out test set were then compared to colonoscopy results and ROC curves were developed. The predetermined threshold from internal cross-validation performance was used to assess model performance on the hold-out test set.

References

1. Imperiale TF, et al. Ann Intern Med 2019 Feb 26. https://doi.org/10.7326/M18-2390 [Epub ahead of print].
2. Paulo JA, et al. Biochim Biophys Acta 2013;1834:791–797.
3. Powell EA, Mortensen JE. Ann Clin Microbiol Antimicrob 2016;15:54.
4. Loens K, et al. J Clin Microbiol 2007;45:421–425.
5. Deschaght P, et al. BMC Microbiol 2009;9:244.
6. Barnell E, et al. Gastroenterology 2018;154:S–578.
7. Landrum MJ, et al. Nucleic Acids Res 2014;42:D980–D985.
8. Forbes SA, et al. Nucleic Acids Res 2017;45:D777–D783.
9. Griffith M, et al. Nat Genet 2017;49:170–174.
10. Stelzer G, et al. Curr Protoc Bioinform 2016;54:1.30.1–1.30.33.
11. Dienstmann R, et al. J Clin Orthod 2014;32:3511–3511.
12. Kim D, et al. Nat Methods 2015;12:357–360.
13. Rennie JDM, Srebro N. Proceedings of the IJCAI Multidisciplinary Workshop on Advances in Preference Handling. Norwell, MA: Kluwer, 2005:180–186.
14. Pedregosa F, et al. J Mach Learn Res 2017;18:1–35.
15. Edwards W, et al. Psychol Rev 1963;70:193–242.