

## ARTICLE OPEN



# Plasma cell-free RNA profiling distinguishes cancers from pre-malignant conditions in solid and hematologic malignancies

Breeshey Roskams-Hieter<sup>1,9</sup>, Hyun Ji Kim<sup>1,2,9</sup>, Pavana Anur<sup>1</sup>, Josiah T. Wagner<sup>1</sup>, Rowan Callahan<sup>1,2</sup>, Elias Spiliotopoulos<sup>1</sup>, Charles Ward Kirschbaum<sup>1</sup>, Fehmi Civitci<sup>1</sup>, Paul T. Spellman<sup>1,3,4</sup>, Reid F. Thompson<sup>1,2,4,5,6</sup>, Khashayar Farsad<sup>7</sup>, Willscott E. Naugler<sup>8</sup> and Thuy T. M. Ngo<sup>1,2,3</sup>✉

Cell-free RNA (cfRNA) in plasma reflects phenotypic alterations of both localized sites of cancer and the systemic host response. Here we report that cfRNA sequencing enables the discovery of messenger RNA (mRNA) biomarkers in plasma with the tissue of origin-specific to cancer types and precancerous conditions in both solid and hematologic malignancies. To explore the diagnostic potential of total cfRNA from blood, we sequenced plasma samples of eight hepatocellular carcinoma (HCC) and ten multiple myeloma (MM) patients, 12 patients of their respective precancerous conditions, and 20 non-cancer (NC) donors. We identified distinct gene sets and built classification models using Random Forest and linear discriminant analysis algorithms that could distinguish cancer patients from premalignant conditions and NC individuals with high accuracy. Plasma cfRNA biomarkers of HCC are liver-specific genes and biomarkers of MM are highly expressed in the bone marrow compared to other tissues and are related to cell cycle processes. The cfRNA level of these biomarkers displayed a gradual transition from noncancerous states through precancerous conditions and cancer. Sequencing data were cross-validated by quantitative reverse transcription PCR and cfRNA biomarkers were validated in an independent sample set (20 HCC, 9 MM, and 10 NC) with AUC greater than 0.86. cfRNA results observed in precancerous conditions require further validation. This work demonstrates a proof of principle for using mRNA transcripts in plasma with a small panel of genes to distinguish between cancers, noncancerous states, and precancerous conditions.

npj Precision Oncology (2022)6:28; <https://doi.org/10.1038/s41698-022-00270-y>

## INTRODUCTION

Although recent advances in cancer research offer new methods to treat cancer, early detection of malignancy still confers the highest chance of improving long-term patient survival. Currently, only 2.4% of metastatic liver cancer patients survive for more than 5 years<sup>1</sup>. Early detection of liver cancer, which has the most rapidly increasing incidence in the United States, has the potential to extend 5-year survival rates to 33% with current treatment options. Even with hematologic malignancies like multiple myeloma (MM), 95% of patients are diagnosed when cancer has already spread systemically, resulting in at least a 20% decrease in 5-year survival rates compared to detection at earlier stages<sup>2</sup>. Noninvasive, low cost and reliable cancer diagnostic assays could greatly benefit patients by facilitating accessibility to early cancer screening.

In many cancers, there are disease states known to be precursors of malignant disease. For example, MM, a cancer of antibody-producing plasma cells, is often preceded by monoclonal gammopathy of undetermined significance (MGUS), which is characterized by lower levels of abnormal antibodies. The prevalence of MGUS is about 3% in the Caucasian population, and the conversion rate from MGUS to multiple myeloma is ~1% per year<sup>3,4</sup>. Hepatocellular carcinoma (HCC), the most common form of liver cancer, is often preceded by liver cirrhosis (Cirr) characterized by irreversible fibrosis of the liver. The prevalence

of cirrhosis is between 4.5–9.5% of the global population<sup>5–7</sup>. The risk of developing *de novo* HCC in patients with liver cirrhosis ranges between 1 and 5% per year, depending on the etiology of the cirrhosis<sup>5–11</sup>. Most early cancer detection studies to date have focused on distinguishing cancer from healthy controls, rather than discriminating between cancer and common premalignant conditions. Therefore, there is an unmet clinical need for a simple blood test that can identify patients with premalignant conditions who require further intervention due to a higher likelihood of cancer incidence.

With current clinical practices, a cancer diagnosis is primarily initiated based upon costly imaging studies or invasive screening procedures. Alternatively, some cancers may only come to attention with clinical symptoms that present at more advanced stages. Liquid biopsy, a minimally invasive method for sampling and analyzing biomarkers in various body fluids, has the potential to improve cancer diagnosis and prognosis<sup>12–15</sup>. Several blood-based analytes have been explored for use in liquid biopsies for cancer detection such as circulating cells (circulating tumor cells (CTCs), circulating hybrid cells (CHCs), tumor-associated macrophages (TAMs))<sup>16–21</sup>, circulating tumor DNA (ctDNA)<sup>22–24</sup>, platelets<sup>25–27</sup>, and protein panels<sup>28</sup>. However, ctDNA and circulating cells are present at low levels, have varied characteristics between patients, and only weakly correlate with phenotypic changes in cancer<sup>17,29,30</sup>. Epigenetic features of ctDNA such as methylation

<sup>1</sup>Cancer Early Detection Advanced Research Center (CEDAR), Knight Cancer Institute, Oregon Health and Science University, Portland, OR, US. <sup>2</sup>Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR, US. <sup>3</sup>Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR, US. <sup>4</sup>Computational Biology Program, Oregon Health and Science University, Portland, OR, US. <sup>5</sup>Department of Radiation Medicine, Oregon Health and Science University, Portland, OR, US. <sup>6</sup>Division of Hospital and Specialty Medicine, VA Portland Healthcare System, Portland, OR, US. <sup>7</sup>Dotter Department of Interventional Radiology, Oregon Health and Science University, Portland, OR, US. <sup>8</sup>Division of Gastroenterology and Hepatology, School of Medicine, Oregon Health and Science University, Portland, OR, US. <sup>9</sup>These authors contributed equally: Breeshey Roskams-Hieter, Hyun Ji Kim. ✉email: [ngth@ohsu.edu](mailto:ngth@ohsu.edu)

and 5-hydroxymethylcytosine signatures, or ctDNA protected patterns may provide information about the tissue of origin for pan-cancer detection<sup>31–38</sup>. However, these methods may require large deep sequencing coverage to be effective and may have inadequate sensitivity and specificity. Recent transcriptome analysis of tumor-educated platelets has shown promise for pan-cancer detection<sup>25–27</sup>, but platelets are fragile, can be easily activated *in vitro*, and have highly variable characteristics depending on their preparation which make them challenging to utilize with existing clinical blood tests<sup>39</sup>. There is thus a need for robust liquid biopsy technology that can overcome these challenges in a safe, reliable, and cost-effective manner.

Circulating cell-free RNA (cfRNA) in the blood is released from cells by active secretion or through apoptosis and necrosis<sup>40,41</sup>. Plasma cfRNA has the potential to reflect the systemic response to growing tumors and provide information about the tissue of tumor origin specifically by cancer type. Previous work has demonstrated that global cfRNA profiles indicate temporal changes in organ-specific transcripts. Further analysis of these transcripts facilitated the prediction of pregnancy delivery, preterm birth, and distinction of cancer from healthy controls<sup>42–46</sup>. Here, we explore the potential of cfRNA profiles to distinguish cancers from their premalignant conditions. We sequenced total plasma cfRNA from plasma samples of patients with HCC and MM and their precancerous conditions including liver cirrhosis (Cirr) and MGUS, and non-cancer (NC) donors. We identified potential cfRNA biomarkers using plasma cfRNA-sequencing of a pilot sample set and validated potential cfRNA biomarkers in an independent sample set. We further validated the sequencing data using orthogonal measurement by quantitative reverse transcription PCR. Feature selection and classification models were built to explore the potential of cfRNA profiles in differentiating malignant from premalignant conditions.

## RESULTS

### Identification of plasma cfRNA biomarkers by sequencing

To identify cfRNA transcripts which potentially distinguish cancer patients from NC individuals, we prospectively collected blood samples from the following individuals: a pilot set of ten MM and eight HCC patients; 13 patients with premalignant conditions including eight MGUS and four Cirr; and 20 NC donors. Detailed clinical information of the samples is listed in the supplementary information (Supplementary Table 1). Plasma cfRNA samples were sequenced to saturation with a mean of 33.8 M raw reads with a range of 27.7 to 52.3 M (Supplementary Table 2 and Supplementary Figs. 2, 3a). After selecting for reads that mapped uniquely to the human genome, the cfRNA libraries had an average read depth of 14 M with a range from 2.3 to 43 M. On average, 80% of reads mapped to exons (Supplementary Table 2 and Supplementary Fig. 3b). A total of 39,374 annotated features were detected with at least one mapped read across all samples. The majority of detected RNAs were protein-coding with a mean fraction of 82% and a range from 65 to 89% (Supplementary Table 2 and Supplementary Fig. 3c). The fraction of reads mapping to exons and the distribution of read depths were uniform across all sample groups (Fig. S3b, c).

We then determined if cfRNA profiles can distinguish HCC and MM from NC donors. Principal component analysis (PCA) using the top 500 genes with the largest variance across all samples through pairwise comparison showed separation of HCC and MM cfRNA profiles from that of NC donors (Fig. 1b, c). Differential expression (DE) analysis of the pairwise comparison between individual cancer types with respect to NC donors using DESeq2 yielded 110, and 12 differentiating genes (adjusted  $p$  value  $<0.01$ ) for MM and HCC, respectively (Supplementary Table 3 and Supplementary Fig. 4a, b). Permutations of random sample shuffling in each pair with

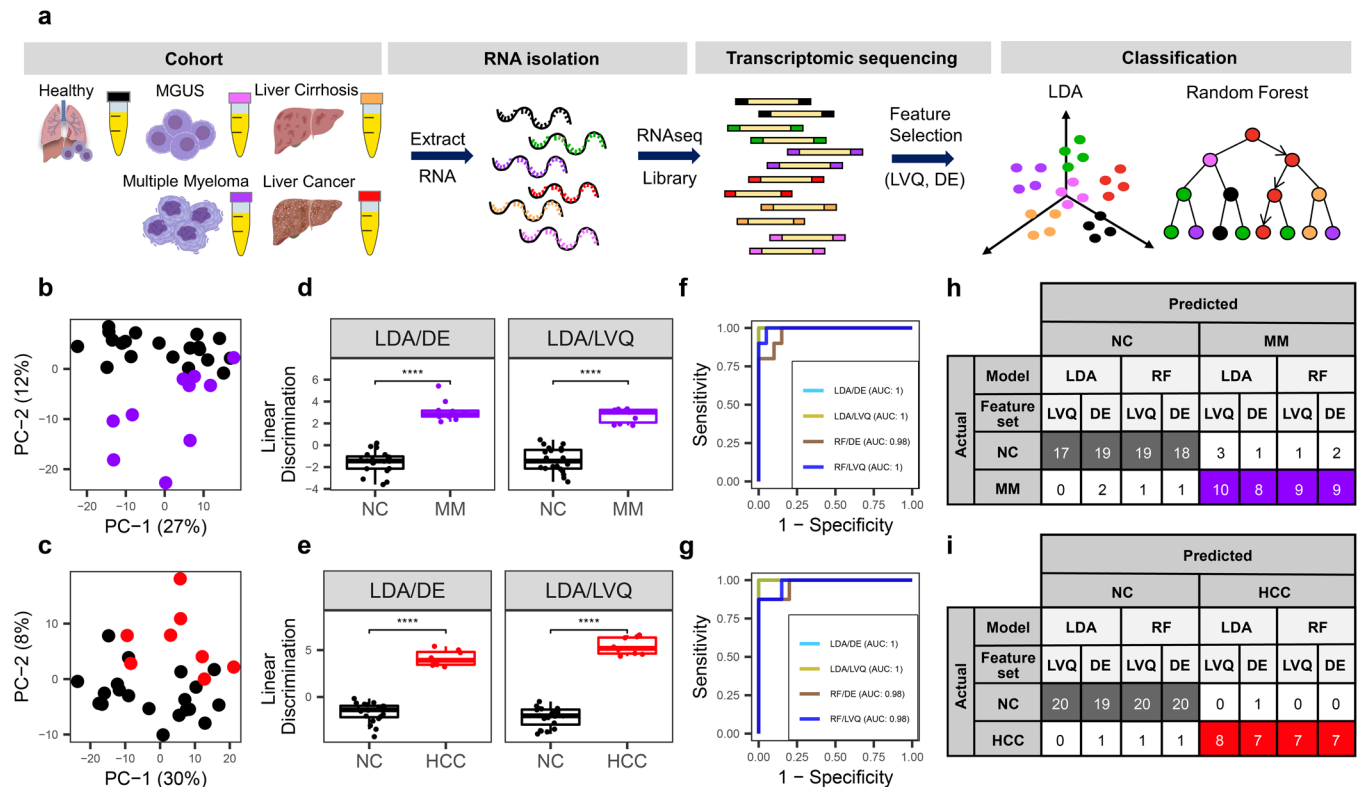
500 rounds resulted in zero significant differentiating genes ( $p_{\text{adj}} < 0.01$ ) in more than 95 and 94% of permutations for each pair comparing MM, and HCC to non-cancer donors, respectively (Supplementary Table 4 and Supplementary Fig. 4c, d). Gene ontology analysis revealed that MM upregulated genes were enriched for oxygen transport and gas transport (Supplementary Figure 5a). In HCC, the upregulated gene set was enriched for plasminogen activation (Supplementary Fig. 5b). This data collectively indicates the separation of cfRNA profiles in HCC and MM compared to NC donors.

To further explore the potential of cell-free RNA for cancer detection, we applied linear discriminant analysis (LDA) and a Random Forest (RF) algorithm to find combinations of discriminating genes to separate cancer from non-cancer individuals. Two independent methods were used to identify specific input gene lists for the classifying algorithms. First, discriminating genes using DESeq2 analysis with adjusted  $p$  value  $<0.01$  (Supplementary Table 3) were used as one feature set (DE gene set). Second, we implemented the learning vector quantization (LVQ) method to find the most important features that distinguished the two groups and selected the top ten as another feature set (LVQ gene set) (Supplementary Table 5). The linear combination for each gene set by LDA showed significant separation between HCC and MM from NC donors with  $p$  value of  $6.7 \times 10^{-8}$ ,  $6.7 \times 10^{-10}$ , and  $6.4 \times 10^{-7}$ ,  $6.4 \times 10^{-7}$  using the DE and top ten LVQ gene sets, respectively (Fig. 1d, e). We further employed the Random Forest (RF) method to develop orthogonal classification models. The area under the receiver operating characteristic (ROC) curve (AUC) was higher than 0.92 in both LDA and RF models with both DE and LVQ feature sets for the two cancer types (Fig. 1f, g).

To evaluate the significance and accuracy of our classification models, we employed the leave-one-out cross-validation (LOOCV) method. Both LDA and RF algorithms were trained on the described DE and LVQ gene sets, resulting in four classification models (Fig. 1f–i). Classifying MM from non-cancer donors yielded 90% accuracy (27/30) for all four models tested. HCC was correctly differentiated from NC donors with accuracies of 100% (28/28) and 93% (26/28) using the LDA method or 96% (27/28) and 96% (27/28) using the RF method with LVQ and DE feature sets, respectively. Overall, the LOOCV test confirmed that the biomarker sets determined by DESeq2 and LVQ methods, combined with our classification models using LDA and RF algorithms are statistically significant. LVQ gene sets yielded higher accuracy for both cancer types and were used as the feature sets for further validation.

### cfRNA profiles distinguish multiple myeloma from its premalignant condition, MGUS, and MGUS from non-cancer

We next examined if cfRNA profiles were able to recapitulate the transition from a precancerous condition to a cancerous one and distinguish between them. We chose to test our hypothesis on MM as it has the well-defined precancerous condition of MGUS. The top ten most significant genes that discriminate MM from NC donors as identified by LVQ displayed a gradual transition in cfRNA level from the non-cancer donors through MGUS to MM (Fig. 2a). Among these ten most significant genes, seven genes (CA1, EPB42, HBG1, HBG2, CENPE, CPOX, and NUSAP1) have higher expression in bone marrow, where cancerous plasma cells accumulate, compared to other tissue and cell types in publicly available data from the Human Protein Atlas<sup>47,48</sup> (Fig. 2b). Three genes resulting from the LVQ analysis are related to cell cycle processes: Centromere protein E (CENPE), a kinesin-like motor protein that accumulates in the G2 phase of the cell cycle and is highly expressed in the bone marrow<sup>49,50</sup>; Serine/threonine-protein kinase (NEK2), which is involved in mitotic regulation<sup>50,51</sup>; and Nucleolar and spindle associated protein 1 (NUSAP1), a nucleolar-spindle-associated protein that plays a role in spindle microtubule organization<sup>52</sup>.



**Fig. 1** cRNA profiles distinguish between cancer vs. healthy donors. **a** Schematic overview of the cRNA profiling workflow starting from plasma collected from the patients and NC donors in EDTA-coated tubes, cRNA extraction, sequencing, feature selection, and classification. **b, c** PCA analysis using the top 500 genes with the largest variance across NC and MM (**b**) or HCC samples (**c**). **d, e** Linear discriminant analysis (LDA) using DE genes with  $p_{adj} < 0.01$  and top ten most important genes identified by LVQ analysis.  $P$  value is derived from the Wilcoxon test. Center-line indicates the median value across all patients in that group, and the hinges represent the lower (Q1) and upper (Q3) quartile, with whiskers extending to the minimum and maximum of the resulting distribution. **f, g** ROC curves of the two classification models LDA and Random Forest (RF) model with two feature sets DE and LVQ. **h, i** LOOCV with the two models LDA and RF with two feature set DE and LVQ. DE genes are listed in Supplementary Table 3 and LVQ genes are listed in Supplementary Table 5.

An LDA plot using a combination of the top ten LVQ genes from pairwise comparisons MM—NC, and MGUS—NC displayed the separation of all three groups (Fig. 2c). An RF model using the top ten most important LVQ genes from MGUS—NC pairwise comparison yielded an accuracy of 89.3% (19/20 NC donors and 6/8 MGUS patients) (Fig. 2d). Classification of MM from MGUS yielded an accuracy of 100% (8/8 MGUS and 10/10 MM) using LOOCV with the RF classification method using the top ten most important genes from LVQ analysis of MM versus NC comparison as a feature set (Fig. 2e). The three-group classification resulted in an accuracy of 86.8% (18/20 NC, 6/8 MGUS, and 9/10 MM) defined by LOOCV using the RF method with the feature set composed of the combination of the top 10 LVQ genes from the comparison MM versus NC and MGUS versus NC donors (Fig. 2f).

### cRNA profiles distinguish liver cancer from its premalignant condition, cirrhosis, and cirrhosis from non-cancer

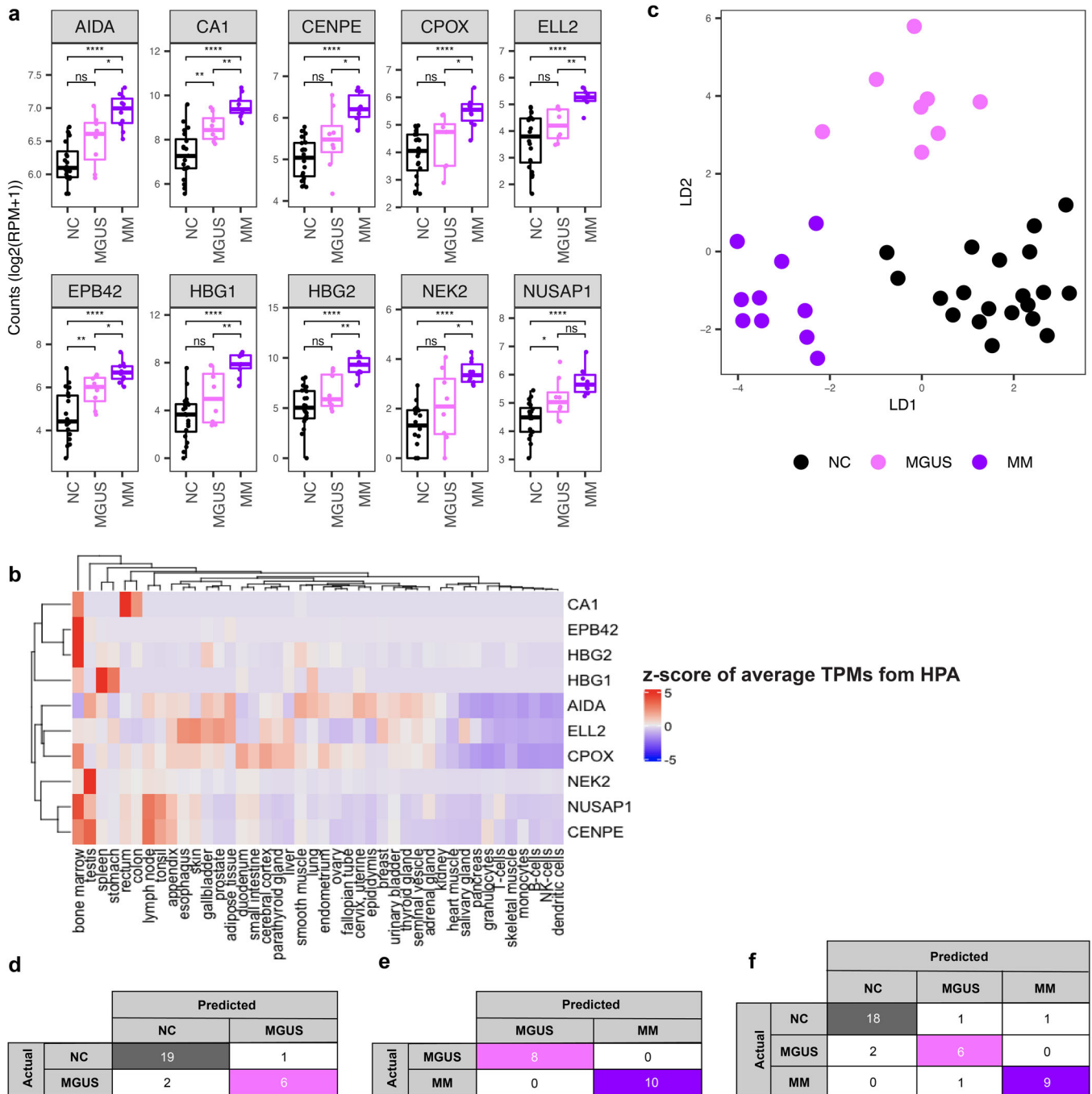
Next, we asked if we could distinguish between a solid tumor such as HCC and its precancerous condition, Cirr. Among the top ten most important genes that discriminate HCC from NC identified by the LVQ analysis, five genes also significantly differentiate HCC from Cirr (Fig. 3a). Interestingly, eight out of the top ten genes are expressed specifically in the liver and the corresponding proteins are secreted into the blood<sup>47,48</sup> (Fig. 3b). Apolipoprotein E (APOE) binds to the specific liver and peripheral cell receptors and is essential for the normal catabolism of triglyceride-rich lipoprotein constituents<sup>53</sup>. Complement C3 (C3) is synthesized in the liver and secreted to the plasma and is involved in both innate and adaptive immune responses<sup>54</sup>. Ceruloplasmin (CP) is a secreted

plasma metalloprotein from the liver that binds copper in the plasma and is involved in the peroxidation of Fe(II) transferrin to Fe(III) transferrin<sup>55</sup>. 24-dehydrocholesterol reductase DHCR24 catalyzes the reduction of sterol intermediates<sup>56</sup>. Fibrinogen alpha chain (FGA), fibrinogen beta chain (FGB), and fibrinogen gamma chain (FGG) encodes the coagulation factor fibrinogen, which is a component of blood clotting<sup>57</sup>. Histidine-rich glycoprotein (HRG) is a plasma glycoprotein that binds heparin sulfate on the surface of the liver, lung, kidney, and heart endothelial cells<sup>58</sup>.

We explored the potential of cRNA to distinguish HCC from Cirr and Cirr from NC individuals. An LDA plot using the feature set comprised of a combination of the top 10 LVQ genes identified for the pairwise comparisons of HCC—NC and Cirr—NC, shows a distinct separation between these groups (Fig. 3c). RF methods using the top ten important genes from Cirr—NC pairwise comparisons yielded 100% accuracy in classifying Cirr from NC samples using LOOCV (Fig. 3d). Classification of HCC from Cirr also yielded 100% accuracy using LOOCV with RF (Fig. 3e). We further attempted to classify three classes including NC, Cirr, and HCC in one model. The three-group classification resulted in 90.6% accuracy using LOOCV with RF (Fig. 3f).

### Validation of cRNA biomarkers

We designed a primer panel for the LVQ gene set to validate the sequencing data by quantitative reverse transcription PCR (RT-qPCR). RT-qPCR results from the pilot sample set were consistent with the sequencing data with a Pearson correlation coefficient  $> 0.77$  and a  $p$  value of  $2.2 \times 10^{-16}$  (Fig. 4a). We confirmed that the differential levels of cRNA transcripts of genes identified by the



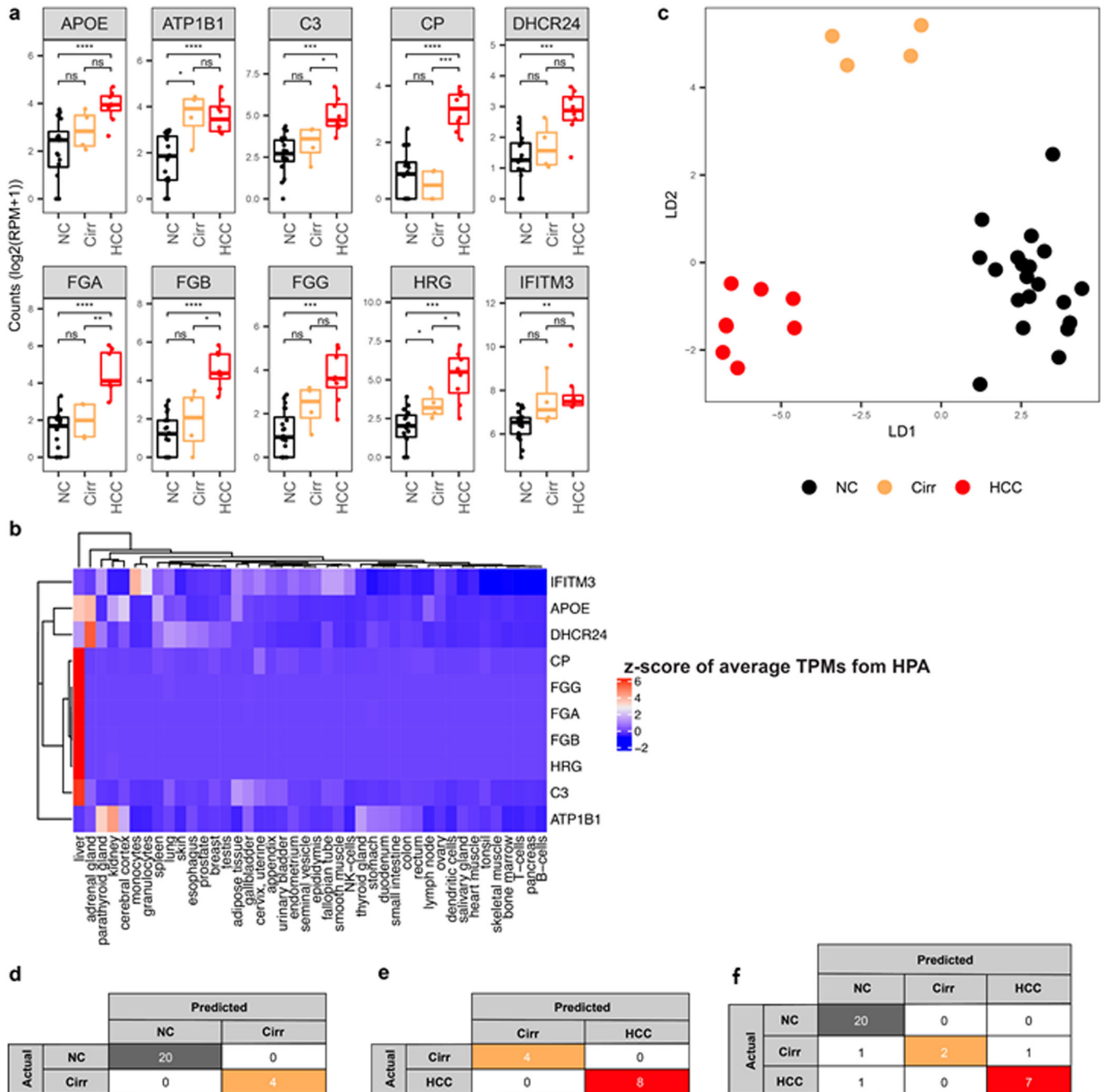
**Fig. 2** cRNA profiles distinguish between non-cancer, MGUS, and multiple myeloma donors. **a** Boxplots of representative top ten most significant genes resulted from the LVQ analysis for MM versus NC. *P* value was calculated for each pair by the *t*-test. Center-line indicates the median value across all patients in that group, and the hinges represent the lower (Q1) and upper (Q3) quartile, with whiskers extending to the minimum and maximum of the resulting distribution. **b** Heatmap of z-score across publicly available tissue-level expression data from the Human Protein Atlas (HPA) for the top ten LVQ genes identified in MM vs. NC. **c** LDA plot using ten genes from pairwise analysis across NC - MGUS and NC - MM pairs using the LVQ method. **d-f** LOOCV using the Random Forest (RF) model with top ten LVQ genes to discriminate NC and NC (d), MM vs MGUS (e), and three groups NC, MGUS, and MM (f). Genes included in the RF analysis are listed in Supplementary Table 5.

LVQ algorithm (HBG1, HBG2, NUSAP1, for MM and C3, CP, FGA, FGB for HCC) from RNA-sequencing were also observed with RT-qPCR (Fig. 4b).

To confirm that the feature sets and classification models defined in our pilot cohort were robust and generalizable, we collected a set of independent validation samples from ten NC controls, nine MM patients, and 20 HCC patients (Supplementary

Table 1 and Supplementary Figs. 6, 7). We validated the cRNA biomarkers identified from the pilot set *in silico* by measuring the classification accuracy on this independent sample set using the models trained with the pilot dataset using the LVQ gene sets. The linear combination identified by LDA in the pilot cohort of the LVQ feature set showed significant separation in the validation sample set between MM and HCC from NC donors, consistent with our



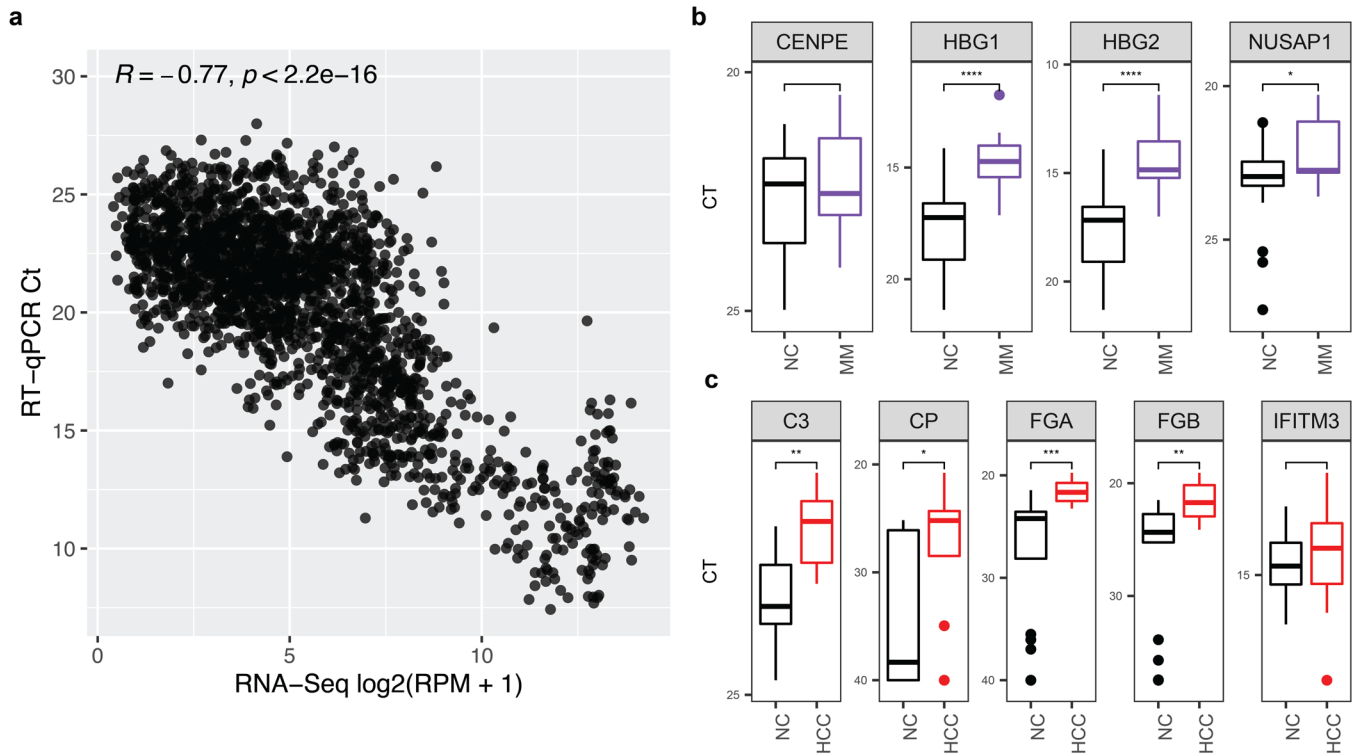


**Fig. 3** **c**RNA profiles distinguish between non-cancer, liver cirrhosis, and liver cancer donors. **a** Boxplots of representative top ten most significant genes resulted from the LVQ analysis for HCC vs. NC. *P* value was calculated for each pair by the *t*-test. Center-line indicates the median value across all patients in that group, and the hinges represent the lower (Q1) and upper (Q3) quartile, with whiskers extending to the minimum and maximum of the resulting distribution. **b** Heatmap of z-score across publicly available tissue-level expression data from the Human Protein Atlas (HPA) for the top ten LVQ genes identified in HCC vs. NC **c** LDA plot using top ten genes identified from each pairwise analysis between NC - Cirr and NC - HCC samples using the LVQ method. **d-f** LOOCV using the RF model with top ten LVQ genes to discriminate Cirr and NC (**d**), HCC vs Cirr (**e**), and three groups NC, Cirr, and HCC (**f**). Genes included in the RF analysis are listed in Supplementary Table 5.

previous results (Fig. 5a, c). Furthermore, both LDA and RF models trained on the pilot cohort with this same feature set were able to classify cancer from NC controls in our validation cohort, with an AUC >0.86 and 0.9 when classifying NC donors from MM and HCC, respectively (Fig. 5b, d).

Our cRNA classification model performed well for early and late clinical stages in the pilot set (Fig. 6a-d). In the validation sample set, the model displayed stage-dependent discrimination. It was

validated with an AUC of 0.74 for Barcelona Clinic Liver Cancer (BCLC) stage A in HCC (Fig. 6e, f) and an AUC of 0.64 for stage I in MM (Fig. 6g, h). For later stages, the model achieved a higher AUC of 0.91 for BCLC stages B and C in HCC (Fig. 6e, f) and 0.83 for stages II and III in MM (Fig. 6g, h) in the validation sample set. This stepwise increase in discrimination suggests that these biomarkers become more prevalent with cancer progression. HCC classification also showed significant discrimination compared to NC for



**Fig. 4** qRT-PCR of cRNA biomarkers is concordant with RNA-sequencing data. **a** Correlation plot of qRT-PCR data compared to RNA-sequencing data.  $P$  value was calculated by t-test. **b, c** qRT-PCR Ct values of top four LVQ genes identified from MM versus NC (**b**) and top 5 LVQ genes identified from HCC versus NC (**c**). Center-line for boxplots in both **b** and **c** indicates the median value across all patients in that group, and the hinges represent the lower (Q1) and upper (Q3) quartile, with whiskers extending to the minimum and maximum of the resulting distribution.

different etiologies (Fig. 7), and both HCC and MM showed discrimination for males and females (Supplementary Fig. 9) and are not age-dependent (Supplementary Fig. 9) in our pilot and validation sample sets.

## DISCUSSION

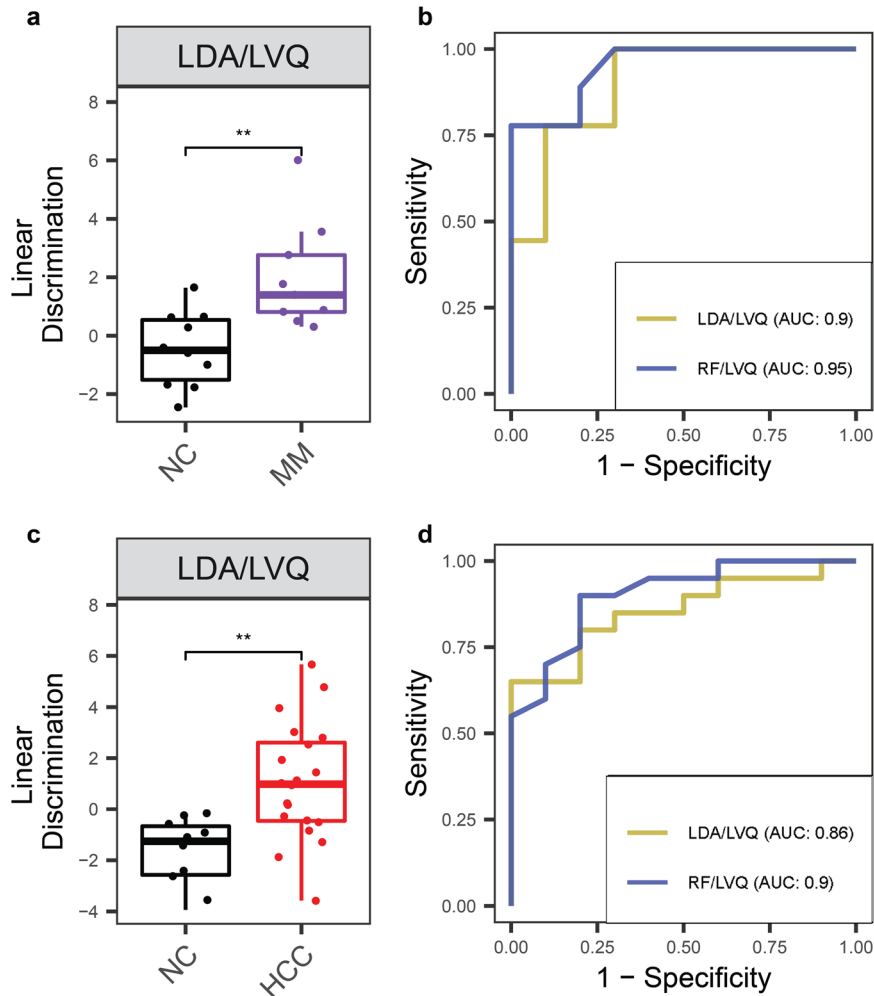
We sequenced cRNA from patients with two cancer types, one solid (HCC) and the other hematologic (MM), and their precancerous conditions (Cirr and MGUS, respectively), and from NC donors. Both cancer types can be distinguished from non-cancer controls and precancerous conditions using their cRNA profiles. To differentiate each cancer type from individuals without cancer, the combination of ten genes identified by learning vector quantization (LVQ) analysis in each pairwise comparison yielded higher accuracy compared to the use of a larger set of differentiating genes as evaluated by leaving one out cross-validation (LOOCV). RT-qPCR confirmation for a panel of selected biomarkers was consistent with the sequencing data. Plasma cRNA biomarkers identified from the sequencing data were further validated in an independent sample cohort. The use of a small gene panel potentially enables a cost-effective office-based assay for pan-cancer detection that can be highly useful in broad clinical applications.

To date, most investigations into the potential of blood-based methods for cancer detection have only focused on distinguishing cancers from healthy controls<sup>15,22,25,26,28,36</sup>. However, many cancer types have etiologies associated with precursor states such as MGUS for MM and Cirr for HCC. Here, we report that cRNA profiles can recapitulate the transition from a precancerous condition to cancer, and can effectively do so for both solid and hematologic cancers. We, therefore, propose that cRNA panels containing a small number of genes may distinguish cancers from

pre-malignant conditions and precursors from healthy individuals. This development might potentially enable a cost-effective screening strategy for early cancer detection during routine exams in high-risk patients.

Liver and bone marrow have been reported to contribute heavily to the abundance of cell-free nucleic acids in plasma<sup>42,45,46</sup>. This may explain the source of cRNA biomarkers found in these cancer types. In HCC, eight out of the top ten genes used in the classification model are specifically synthesized in the liver and encode secreted proteins found in the blood that mediate plasminogen activation and fibrinolysis processes. In MM, seven out of ten genes among the most important cRNA biomarkers have relatively high expression in bone marrow compared to other tissue and cell types and are related to cell cycle processes. These findings indicate that the identified cRNA biomarkers potentially originate from the tissue of origin of the tumor. Further investigation is needed to better define the tissue and cell-type origin of the biomarkers, and how they may associate with disease initiation and progression.

Our study has important limitations. This is cross-sectional single sampling with a small sample size for both the discovery and validation sets. Furthermore, the sample sets do not represent the wider distribution of cancer subtypes and precursor lesions in the overall population with different underlying etiologies. Another limitation is that the majority of patients and controls are white, so further studies are needed to examine if these results can be extrapolated to a more diverse population of other races. However, accurate classification is not sex- or age-dependent. Although the control population has a higher female/male ratio than the cases, our classification model showed significant discrimination for both males and females in the pilot and validation sets. Despite the median age of controls being 9 and 6 years younger than cases for the pilot and validation sets,



**Fig. 5** cfrNA biomarkers and classification models validated in independent sample set. **a, c** Linear discriminant analysis in the validation cohort using top ten LVQ genes identified and classification models trained on the pilot cohort for MM versus NC, and HCC versus NC. *P* value was calculated for each pair by the Wilcoxon rank-sum test. Center-line indicates the median value across all patients in that group, and the hinges represent the lower (Q1) and upper (Q3) quartile, with whiskers extending to the minimum and maximum of the resulting distribution. **b, d** ROC curves of these same classification models, trained on the pilot sample set and tested with the validation sample set, using the top ten LVQ genes identified from the pilot sample set.

respectively, discrimination does not depend on age. Our results for precancerous conditions are promising but require further validation. In addition, we have not fully characterized the stability of cell-free RNA and the biological origin of the identified cfrNA biomarkers. Before the tests developed from this work can be clinically applied, large-scale clinical studies will be required to validate the potential of cfrNA as a cancer biomarker and to build robust classification models. Such large-scale clinical studies will also help to determine if the test can be applied to a broader risk population without specific predispositions.

In summary, we report a proof of principle that global profiling of cell-free mRNA has the potential to establish a platform for longitudinal monitoring of disease progression across both solid and hematologic cancers. This work lays the foundation for developing inexpensive assays that measure transcript levels of mRNA in plasma for a small panel of genes that can differentiate pan-cancer from premalignant conditions and otherwise healthy donors. Intriguingly, organ-specific enriched mRNA transcripts were identified as biomarkers that might indicate the tissue of origin for the tumor. These cell-free plasma RNA biomarkers could be readily combined with other nucleic acid-based and protein-based approaches for potentially increased diagnostic sensitivity and specificity.

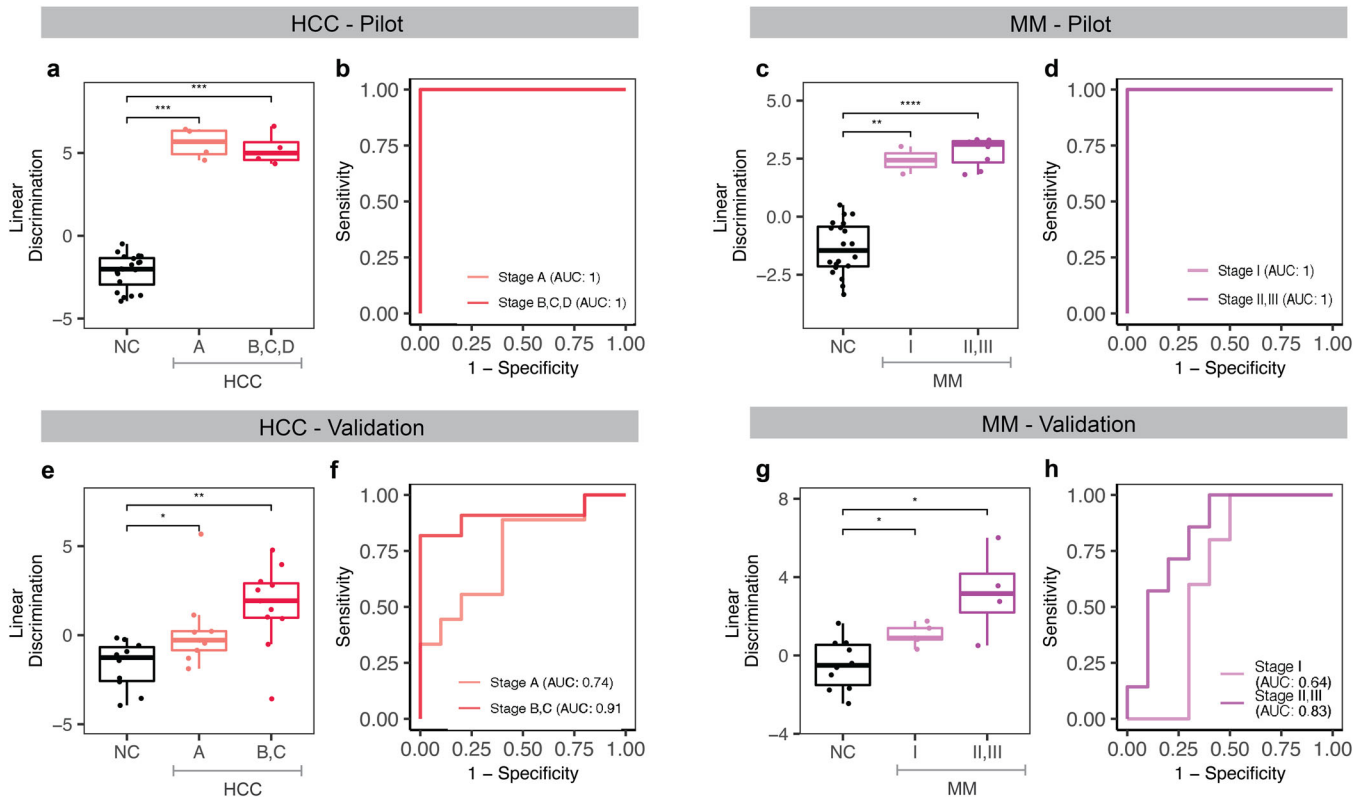
## METHODS

For a detailed summary of the methods used and the general workflow of our study please see Supplementary Fig. 1.

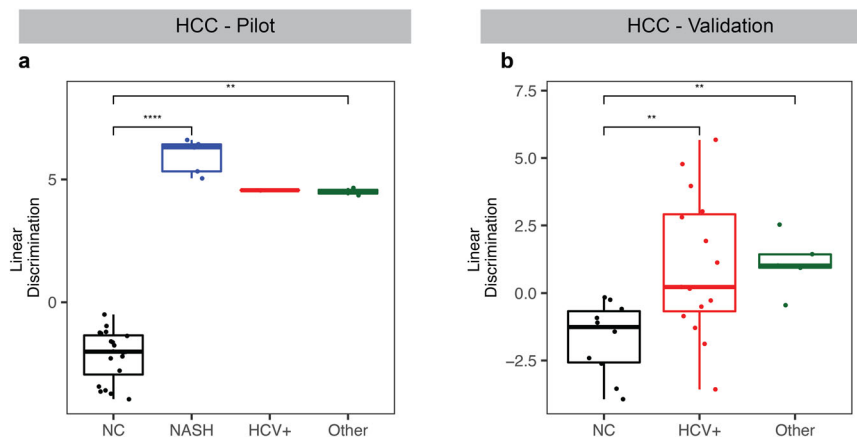
### Patient samples

Blood samples from non-cancer donors and patients with monoclonal gammopathy of undetermined significance (MGUS), multiple myeloma, liver cirrhosis, and liver cancer were obtained from Oregon Health and Science University (OHSU) by Knight Cancer Institute Biobank and Oregon Clinical and Translational Research Institute (OCTRI). All samples were collected under institutional review board (IRB) approved protocols by Oregon Health and Science University. Participants provided written informed consent to take part in the study. Individuals who had no recorded previous history of cancer were considered to be non-cancer donors.

All samples with various diagnoses within the same sample set were collected and processed using a uniform protocol by the same staff at Oregon Health and Science University. The validation set and the pilot set were collected and processed independently by two groups of staff. The clinical information regarding study participants are given in the Supplementary Table S1. The pilot set includes 10 MM and 8 HCC patients; 13 patients with premalignant conditions including eight MGUS and four Cirr; and 20 NC donors. The validation set includes ten NC controls, nine MM patients, and 20 HCC patients. All Cirr patients underwent abdominal



**Fig. 6** **cfRNA biomarkers show clinical stage-dependent discrimination in pilot and validation sample sets.** **a–d** Linear discriminant analysis using the top ten LVQ genes and model trained in the pilot cohort shows significant discrimination and classification by clinical stage in both HCC (**a, b**) and MM (**c, d**). **e–h** When classifying the independent validation cohort with these same models, we see stage-dependent classification for both HCC (**e, f**) and MM (**g, h**). *P* value for each pair in (**a, c, e, g**) was calculated by the Wilcoxon rank-sum test, and elements of the boxplots include the median value across all patients in that group shown by the center-line and the hinges which represent the lower (Q1) and upper (Q3) quartile, with whiskers extending to the minimum and maximum of the resulting distribution.



**Fig. 7** **cfRNA biomarkers for HCC show discrimination between various etiologies.** **a** Linear discriminant analysis trained on the pilot cohort with the top ten LVQ genes show significant discrimination between NC and HCC on the background of NASH, HCV+, and other etiologies in the pilot cohort and the validation cohort (**b**). *P* value for each pair was calculated by the Wilcoxon rank-sum test. Center-line in each boxplot indicates the median value across all patients in that group, and the hinges represent the lower (Q1) and upper (Q3) quartile, with whiskers extending to the minimum and maximum of the resulting distribution.

US or MRI and all MGUS patients had an evaluation of bone marrow or assessment of serum-free light chain ratio within 6 months of blood collection for this study.

### Processing of whole blood

For all cohorts, whole blood samples were collected in EDTA-anticoagulated vacutainers. Within 2 h of collection, blood samples were

first centrifuged at 1000×g for 10 min at 4°C followed by 15,000×g for 10 min at 4°C. Plasma was then stored at –80°C until RNA isolation.

### cfRNA isolation

Samples were randomly shuffled for RNA extraction, library preparation, and sequencing in Illumina flow cells (Fig. 1a). Total RNA purification was performed by using a plasma/serum circulating and exosomal RNA purification kit (Norgen Biotek) from 3 ml of human plasma according to



the manufacturer's protocol. To digest trace amounts of contaminating DNA, RNA was treated with 10X Baseline-ZERO DNase. DNase I treated RNA samples were purified and further concentrated using RNA clean and concentrator-5 (Zymo Research) according to the manufacturer's manuals. Final eluted RNA was stored immediately at  $-80^{\circ}\text{C}$ .

### Library preparation

We prepared stranded RNA-Seq libraries using Clontech SMARTer stranded total RNA-seq kit v2- pico input mammalian (Takara Bio) according to the manufacturer's instructions. For cDNA synthesis, we used option 2 (without fragmentation), starting from highly degraded RNA. The input of 7  $\mu\text{l}$  of RNA samples were used to generate cDNA libraries suitable for next-generation sequencing. For the addition of adapters and indexes, we employed SMARTer RNA unique dual index kit  $-96$  U. SMARTer RNA unique dual index of each 5' and 3' PCR primer were added to each sample to distinguish pooled libraries from each other. The amplified RNA-seq library was purified by immobilization onto the AMPure XP PCR purification system (Beckman Coulter). The library fragments originated from rRNA and mitochondrial rRNA were treated with ZapR v2 and R-Probes according to the manufacturer's protocols. For final RNA-seq library amplification, 16 cycles of PCR were performed and the final 20  $\mu\text{l}$  was eluted in Tris buffer following amplified RNA-seq library purification. The amplified RNA-seq library was stored at  $-20^{\circ}\text{C}$  prior to sequencing.

### Sequencing data processing and quality control

Each sample was sequenced to more than 20 million paired-end reads using an Illumina Nextseq or HiSeq sequencer. Adapter sequences were trimmed using sickle tool<sup>59</sup>. After trimming, the quality of the reads were checked using FastQC (v0.11.7)<sup>60,61</sup> and RSeQC (v2.6.4)<sup>62</sup>. Reads were aligned to the hg38 human genome using the STAR aligner (v2.5.3a)<sup>63</sup> with two pass mode flag. Duplicated reads were removed using the Picard tool (v1.119)<sup>64</sup>. Read counts for each gene were calculated using the htseq-count tool (v0.11.2)<sup>65</sup> in intersection-strict mode. The number of mapped reads to each gene were normalized to the total number of reads in the whole transcriptome (Reads Per Million - RPM). For each sample, we calculated exon, intron, intergenic fractions, and protein-coding fractions (CDS exons) using RSeQC<sup>62</sup> and the read\_distribution.py script. Samples with an exon fraction larger than 0.35 were kept for further analysis.

### Identification of cfRNA biomarkers (DESeq and LVQ and GO analysis)

Two independent methods were applied to select cfRNA features for building classification models. Differentiating genes between all pairwise comparisons were identified with the R package DESeq2 (v1.24.0) using the Wald test<sup>66</sup> with adjusted  $p$  value ( $p_{\text{adj}} < 0.01$ ) (Supplemental Table S3) were used as one feature set (DE gene set). The second method for feature selection uses the LVQ algorithm built-in an R package caret (v6.0-84)—with tenfold cross-validation repeated three times<sup>67</sup>. The top ten most important features were selected by ranking the varImp parameter (LVQ gene set) (Supplemental Table S5). Gene Ontology (GO) analysis was implemented on the top differentiating genes from the DESeq2 analysis with  $p_{\text{adj}} < 0.01$  using the package topGO (v2.37.0) and a Fischer statistical test to measure significant enrichment of each Gene Ontology term<sup>68</sup>.

### Cancer type classification (LDA and RF)

Two methods were used to build models for classifying cancer types using feature sets identified from pairwise comparison using DESeq2 and LVQ methods. LDA models were built using the R package MASS (v7.3-51.4)<sup>69</sup>. Random Forest models were built using the R package randomForest (v4.6-14)<sup>70</sup>.

### Statistical consideration (permutation test and leave-one-out cross-validation)

To test the significance of the differential expression results for each pairwise comparison of cancer to NC donors, we performed a permutation test in which differential expression analysis between two groups of randomized samples was compared using the DESeq2 package. For each pair, 500 permutations of random shuffling were performed and the number of differentiating genes with  $p_{\text{adj}} < 0.01$  were documented for building a histogram and compared to the number of significant genes

( $p_{\text{adj}} < 0.01$ ) for the group with correct labeling. To determine the significance and accuracy of our classification models, we employed the LOOCV method. Briefly, in LOOCV, LDA, or RF algorithms classified each sample based on the training models obtained from all other samples (total number of samples in each pair minus the testing sample). The test was repeated until all individual samples were classified and cross-validated.

### Tissue specificity of LVQ feature sets using publicly available databases

To evaluate whether our LVQ gene sets were tissue-specific to the tissue of origin (TOO), we downloaded publicly available average tissue-level expression values (transcripts per million; TPMs) from the Human Protein Atlas (ref: <https://www.proteinatlas.org/about/download>). The methodology used to normalize and calculate average expression values can be found here: [https://www.proteinatlas.org/about/assays+annotation#hpa\\_rna](https://www.proteinatlas.org/about/assays+annotation#hpa_rna). We then subsetted this matrix of counts values for our two gene sets (top ten LVQ for MM versus non-cancer, and top ten LVQ for HCC versus non-cancer), and calculated a z-score across tissue types to evaluate which tissue types the genes were enriched in. Next, we generated a heatmap of this transformed matrix using ComplexHeatmap (v2.4.3).

### Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

cfRNA sequencing data have been deposited in the Gene Expression Omnibus Repository (GSE182824).

### CODE AVAILABILITY

In-house scripts used in this manuscript, which includes data processing, downstream analysis, and the scripts used to generate figures is publicly available on [GitHub](https://github.com)

Received: 1 September 2021; Accepted: 16 March 2022;

Published online: 25 April 2022

### REFERENCES

- SEER. *Cancer Stat Facts: Liver and Intrahepatic Bile Duct Cancer* (National Cancer Institute, 2018)
- Howlander N. et al. *SEER Cancer Statistics Review, 1975–2018* (National Cancer Institute, 2021)
- Kyle, R. A. & Rajkumar, S. V. Management of monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). *Oncol.* **25**, 578–586 (2011).
- Dhodapkar, M. V. MGUS to myeloma: a mysterious gammopathy of under-explored significance. *Blood* **128**, 2599 (2016).
- Llovet, J. M. et al. Hepatocellular carcinoma. *Nat. Rev. Dis. Prim.* **2**, 16018 (2016).
- Fateen, W. & Ryder, S. D. Screening for hepatocellular carcinoma: patient selection and perspectives. *J. Hepatocell. Carcinoma* **4**, 71–79 (2017).
- Starr, S. P. & Raines, D. Cirrhosis: diagnosis, management, and prevention. *Am. Fam. Physician* **84**, 1353–1359 (2011).
- Laursen, L. A preventable cancer. *Nature* **516**, S2 (2014).
- Goh, G. B., Chang, P. E. & Tan, C. K. Changing epidemiology of hepatocellular carcinoma in Asia. *Best. Pr. Res. Clin. Gastroenterol.* **29**, 919–928 (2015).
- Wong, V. W. et al. Clinical scoring system to predict hepatocellular carcinoma in chronic hepatitis B carriers. *J. Clin. Oncol.* **28**, 1660–1665 (2010).
- Yang, H. I. et al. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. *Lancet Oncol.* **12**, 568–574 (2011).
- Bai, Y. & Zhao, H. Liquid biopsy in tumors: opportunities and challenges. *Ann. Transl. Med* **6**, S89 (2018).
- Palmirotta, R. et al. Liquid biopsy of cancer: a multimodal diagnostic tool in clinical oncology. *Ther. Adv. Med Oncol.* **10**, 1758835918794630 (2018).
- Marrugo-Ramírez, J., Mir, M. & Samitier, J. Blood-based cancer biomarkers in liquid biopsy: a promising non-invasive alternative to tissue biopsy. *Int. J. Mol. Sci.* **19**, 2877 (2018)

15. Esposito, A. et al. Liquid biopsies for solid tumors: understanding tumor heterogeneity and real time monitoring of early resistance to targeted therapies. *Pharm. Ther.* **157**, 120–124 (2016).
16. Sundling, K. E. & Lowe, A. C. Circulating tumor cells: overview and opportunities in cytology. *Adv. Anat. Pathol.* **26**, 56–63 (2019).
17. Millner, L. M., Linder, M. W. & Valdes, R. Jr. Circulating tumor cells: a review of present methods and the need to identify heterogeneous phenotypes. *Ann. Clin. Lab Sci.* **43**, 295–304 (2013).
18. Thiele, J. A. et al. Circulating tumor cells: fluid surrogates of solid tumors. *Annu. Rev. Pathol. Mechanisms Dis.* **12**, 419–447 (2017).
19. Liu, Y. & Cao, X. The origin and function of tumor-associated macrophages. *Cell. Mol. Immunol.* **12**, 1 (2014).
20. Adams, D. L. et al. Circulating giant macrophages as a potential biomarker of solid tumors. *Proc. Natl Acad. Sci. USA* **111**, 3514 (2014).
21. Gast, C. E. et al. Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival. *Sci. Adv.* **4**, eaat7828 (2018).
22. Newman, A. M. et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).
23. Corcoran, R. B. & Chabner, B. A. Application of cell-free DNA analysis to cancer treatment. *N. Engl. J. Med.* **379**, 1754–1765 (2018).
24. Abbosh, C. et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
25. Best, M. G. et al. RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell* **28**, 666–676 (2015).
26. Best, M. G., Wesseling, P. & Wurdinger, T. Tumor-educated platelets as a non-invasive biomarker source for cancer detection and progression monitoring. *Cancer Res.* **78**, 3407–3412 (2018).
27. In't Veld, S. G. J. G. & Wurdinger, T. Tumor-educated platelets. *Blood* **133**, 2359–2364 (2019).
28. Cohen, J. D. et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926 (2018).
29. Abbosh, C., Birkbak, N. J. & Swanton, C. Early stage NSCLC - challenges to implementing ctDNA-based screening and MRD detection. *Nat. Rev. Clin. Oncol.* **15**, 577–586 (2018).
30. Haque, I. S. & Elemento, O. Challenges in using ctDNA to achieve early detection of cancer. Preprint at *bioRxiv* 237578 (2017).
31. Salta, S. et al. A DNA methylation-based test for breast cancer detection in circulating cell-free DNA. *J. Clin. Med.* **7**, 420 (2018).
32. Xu, R.-h et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* **16**, 1155 (2017).
33. Song, C.-X. et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res.* **27**, 1231 (2017).
34. Shen, S. Y. et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579–583 (2018).
35. Moss, J. et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **9**, 5068 (2018).
36. Cristiano, S. et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389 (2019).
37. Liu, M. C. et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* **31**, 745–759 (2020).
38. Chen, X. et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat. Commun.* **11**, 3475 (2020).
39. Gemmell, C. H. Activation of platelets by in vitro whole blood contact with materials: increases in microparticle, procoagulant activity, and soluble P-selectin blood levels. *J. Biomater. Sci. Polym. Ed.* **12**, 933–943 (2001).
40. Heitzer, E. et al. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat. Rev. Genet.* **20**, 71–88 (2019).
41. Wan, J. C. M. et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223 (2017).
42. Koh, W. et al. Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc. Natl Acad. Sci. USA* **111**, 7361–7366 (2014).
43. Pan, W. et al. Simultaneously monitoring immune response and microbial infections during pregnancy through plasma cfRNA sequencing. *Clin. Chem.* **63**, 1695–1704 (2017).
44. Ngo, T. T. M. et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science* **360**, 1133 (2018).
45. Larson, M. H. et al. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat. Commun.* **12**, 2357 (2021).
46. Ibarra, A. et al. Non-invasive characterization of human bone marrow stimulation and reconstitution by cell-free messenger RNA sequencing. *Nat. Commun.* **11**, 400 (2020).
47. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
48. Uhlen, M. et al. Tissue-based map of the human proteome. *Science* **347**, 394 (2015).
49. Sardar, H. S. & Gilbert, S. P. Microtubule capture by mitotic kinesin centromere protein E (CENP-E). *J. Biol. Chem.* **287**, 24894–24904 (2012).
50. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
51. Fry, A. M. The Nek2 protein kinase: a novel regulator of centrosome structure. *Oncogene* **21**, 6184–6194 (2002).
52. Mills, C. A. et al. Nucleolar and spindle-associated protein 1 (NUSAP1) interacts with a SUMO E3 ligase complex during chromosome segregation. *J. Biol. Chem.* **292**, 17178–17189 (2017).
53. Srivastava, R. A., Bhasin, N. & Srivastava, N. Apolipoprotein E gene expression in various tissues of mouse and regulation by estrogen. *Biochem. Mol. Biol. Int.* **38**, 91–101 (1996).
54. Jia, Q. et al. Association between complement C3 and prevalence of fatty liver disease in an adult population: a cross-sectional study from the Tianjin Chronic Low-Grade Systemic Inflammation and Health (TCLSIHealth) cohort study. *PLoS ONE* **10**, e0122026 (2015).
55. Zeng, D. W. et al. Serum ceruloplasmin levels correlate negatively with liver fibrosis in males with chronic hepatitis B: a novel noninvasive model for predicting liver fibrosis in HBV-related liver disease. *PLoS ONE* **8**, e77942 (2013).
56. Waterham, H. R. et al. Mutations in the 3beta-hydroxysterol Delta24-reductase gene cause desmosterolosis, an autosomal recessive disorder of cholesterol biosynthesis. *Am. J. Hum. Genet.* **69**, 685–694 (2001).
57. Fort, A. et al. A liver enhancer in the fibrinogen gene cluster. *Blood* **117**, 276–282 (2011).
58. Gram, J. et al. Plasma histidine-rich glycoprotein and plasminogen in patients with liver disease. *Thromb. Res.* **39**, 411–417 (1985).
59. Joshi, N. A. & Sickle, F. J. A sliding-window, adaptive, quality-based trimming tool for FastQ files. <https://github.com/najoshi/sickle> (2011).
60. Leggett, R. M. et al. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Front. Genet.* **4**, 288–288 (2013).
61. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010)
62. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
63. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1–33 (2013).
65. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
66. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
67. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
68. Alexa A. & Rahnenfuhrer, J. topGO: enrichment analysis for gene ontology. R package version 2.36.0. (2019)
69. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* (Springer, 2002)
70. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R. N.* **2**, 18–22 (2002).

## ACKNOWLEDGEMENTS

This work was supported by the Cancer Early Detection Advanced Research Center (CEDAR), Knight Cancer Institute, Oregon Health and Science University (CEDAR3250918), Cancer Research UK/OHSU Project Award (C63763/A27122 to T.T.M.N.), the OCTRI CTSa grant (UL1TR000128), the Kuni Foundation, the Department of Defense (W81XWH2110853 to T.T.M.N.) and the Susan G. Komen Foundation (CCR21663959 to T.T.M.N.). We would like to acknowledge the CEDAR repository and the Biolibrary for helping with sample collection and processing and Gordon Mills, Joseph Estabrook, and Ashley Woodfin for helpful discussion.

## AUTHOR CONTRIBUTIONS

T.T.M.N. designed and supervised all aspects of the study and wrote the manuscript; All authors contributed to and edited the manuscript; H.J.K., J.T.W., C.W.K., E.S., and T.T.M.N. designed experiments and carried out RNA extraction, library preparation, and RT-qPCR measurement; T.T.M.N., P.A., B.R.-H., and R.C. processed the sequencing data; T.T.M.N., P.A., and B.R.-H. developed statistical tools and analyzed the data; F.C.,

P.S., R.F.T., K.F., and W.E.N. contributed to the study design and the manuscript. B.R-H and H.J.K. are co-first authors.

### COMPETING INTERESTS

Oregon Health and Science University has filed patent applications based on this work. The authors declare no competing interests.

### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-022-00270-y>.

**Correspondence** and requests for materials should be addressed to Thuy T. M. Ngo.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022