

integrate into hepatocyte genome in the liver and cause chronic hepatocyte cell death and inflammation (chronic hepatitis). Repeated reparative response to the hepatitis replaces hepatocytes with scar-forming cells (fibroblasts), thus liver cirrhosis. The accumulation of genetic mutations in the hepatocyte genome during prolonged cell death and regeneration results in malignant transformation of hepatocytes, i.e. hepatocellular carcinoma (HCC). HBV-related HCC accounts for ~80% of the liver cancer in some localities, e.g. Hong Kong.

[0005] Detection of cellular abnormalities and the presence of disease in an organ system commonly requires direct tissue sampling (biopsy) of the organ of interest, which can carry infection and bleeding risk of invasive procedures. Non-invasive assessment by imaging, such as ultrasound scan, provides morphological and specific functional information of organ, such as blood flow. Liver ultrasonography has been employed in the screening of liver cancer in chronic HBV hepatitis patients and uterine artery Doppler analysis is used in preeclampsia prediction in early pregnancy. These however requires well-trained operators for assessment and does not assess the cellular aberrations directly.

[0006] Non-invasive methods of detecting cellular abnormalities and the presence of a disease in an organ system are desired. These and other improvements are addressed.

SUMMARY

[0006a] In one aspect, the present disclosure refers to a method of identifying an expressed marker to differentiate between different levels of a condition, the method comprising: for each cell of a plurality of cells obtained from one or more first subjects: analyzing RNA molecules from the cell to obtain a set of reads, thereby obtaining a plurality of sets of reads; for each read of the set of reads: identifying, by a computer system, an expressed region in a reference sequence corresponding to the read; for each of a plurality of expressed regions: determining an amount of reads corresponding to the expressed region; determining an expression score for the expressed region using the amount of reads corresponding to the region, thereby determining a multidimensional expression point comprised of the expression scores for the plurality of expressed regions; grouping, by the computer system, the plurality of cells into a plurality of clusters using the multidimensional expression points corresponding to the plurality of cells, the plurality of clusters being less than the plurality of cells; for each cluster of the plurality of clusters, determining a set of one or more preferentially expressed regions that are expressed in

cells of the cluster at a specified rate more than cells of other clusters; for each of a plurality of cell-free RNA samples: analyzing a plurality of cell-free RNA molecules to obtain a plurality of cell-free reads, wherein the plurality of cell-free RNA samples are from a plurality of cohorts of second subjects, wherein each cohort of the plurality of cohorts has a different level of the condition; and for each set of one or more preferentially expressed regions of the plurality of sets of one or more preferentially expressed regions: measuring a signature score for the corresponding cluster using cell-free reads corresponding to the set of one or more preferentially expressed regions; identifying, based on the signature scores, one or more of the sets of one or more preferentially expressed regions as one or more expressed markers for use in classifying future samples to differentiate between different levels of the condition.

[0006b] In another aspect, the present disclosure refers to a method of determining a level of a condition in a subject, the method comprising: receiving a plurality of cell-free reads from analysis of cell-free RNA molecules from a biological sample obtained from the subject; for each preferentially expressed region of one or more expressed markers, the one or more expressed markers determined by the method as disclosed herein: determining an amount of reads for the preferentially expressed region, and comparing the amount of reads to a reference value for one or more preferentially expressed regions to one or more reference values; and determining, based on the comparisons of the amount of reads for each preferentially expressed regions to one or more reference values, the level of the condition for the subject.

[0006c] In another aspect, the present disclosure refers to a method of determining a level of a condition in a subject, the method comprising: receiving a plurality of cell-free reads from analysis of cell-free RNA molecules from a biological sample obtained from the subject; determining a value of a temporal parameter related to the condition; determining, using the value of the temporal parameter, an expressed markers for the condition at a time of the value of the temporal parameter, the expressed marker comprising one or more sets of preferentially expressed regions; for each preferentially expressed region of the expressed marker: determining an amount of reads corresponding to the preferentially expressed region; comparing the amount of reads for one or more preferentially expressed regions to one or more reference values; and determining, based on the comparison of the amount of reads for one or more preferentially expressed regions to one or more reference values, the level of the condition for the subject.

[0006d] In another aspect, the present disclosure refers to a computer product comprising a computer readable medium storing a plurality of instructions for controlling a computer system to perform the method as disclosed herein.

5 **[0006e]** In another aspect, the present disclosure refers to a system comprising one or more processors configured to perform the method as disclosed herein.

[0007] Examples of the present technology involve integrative single-cell and cell-free plasma RNA transcriptomics. Examples allow for the determination of expressed regions that can be used to identify, determine, or diagnosis a condition or disorder in a subject. Methods described herein analyze cell-free RNA molecules for certain expressed regions. The specific expressed
10 regions analyzed were previously determined to be indicative for a certain type of cell or grouping of cells. As a result, the amounts of cell-free reads at the specific expressed regions may be related to the number of cells in a tissue or organ. The number of cells in the tissue or organ may change as a result of cell death, metastasis, or other dynamics. A change in the number of cells in the tissue or organ may then be reflected in certain expressed regions in cell-
15 free RNA.

[0008] Example methods in the present technology include analyzing reads from cellular RNA molecules obtained from a plurality of first subjects. The RNA molecules are grouped into