



Investigating the impact of RNA integrity variation on the transcriptome of human leukemic cells

Suliman A. Alsagaby¹

Received: 31 December 2021 / Accepted: 17 June 2022 / Published online: 7 July 2022
© King Abdulaziz City for Science and Technology 2022

Abstract

High RNA integrity is essential for good quality of transcriptomics profiling. Nevertheless, in some cases samples with low RNA integrity is the only available material to study. This work was set to investigate the impact of thermal-induced RNA degradation on the transcriptomic profiles of human leukemic cells. DNA microarray-based transcriptomics was conducted on two groups of samples; high RNA integrity samples ($n=4$) and low RNA integrity samples ($n=5$). RNA degradation caused limited but noticeable changes in the transcriptomes. Only 1945 (6.7%) of 29,230 genes showed altered quantitation (fold change \geq two-fold, p value ≤ 0.03 , corrected p value ≤ 0.05). RNA degradation had the most impact on short transcripts and those with short distance between their 5' end and the probe binding position. Overall, the present work identified the genes whose relative quantification is sensitive to RNA degradation. Therefore, altered expression of these genes should be interpreted with caution when studied in low integrity RNA samples.

Keywords Transcriptomics · RNA Degradation · Misquantitation · Leukemia

Introduction

In contrast to conventional methods of molecular biology where the expression of a single or few genes can be evaluated in a single assay, transcriptomics approach enables gene expression profiling of all human coding-genes in a single experiment (Lowe et al. 2017). Therefore, transcriptomic approach saves time and effort; and most importantly provides an almost comprehensive picture of the whole transcriptome (the sum of all RNA transcripts) of a sample. This has made transcriptomics to be commonly utilized for studying the molecular bases of human diseases, especially tumors (Iacobas 2020; Uhlen et al. 2017). Transcriptomics-based findings have heavily contributed to a better understanding of the molecular pathology of human cancer; and facilitated the discovery of therapeutic targets and biomarkers (Alsagaby 2019a; Rodon et al. 2019; Supplitt et al. 2021; Wang et al. 2017).

There are two main approaches of transcriptomics; DNA-based hybridization, where a large number of oligonucleotides (probes) that correspond to all human coding-genes are attached on a surface of a microscopic slide (Barbulovic-Nad et al. 2006). The location of each probe on the microarray slide is annotated; hence each probe location corresponds to a specific gene. When fluorescent-labeled cDNA converted from total RNA of a sample is added to the microarray slide, the fluorescent-labeled cDNAs hybridize specifically with probes. Next, the intensity of the fluorescent signal at a probe location on the microarray slide is used to indicate the relative quantity of a transcript (Barbulovic-Nad et al. 2006). The other approach of transcriptomics is RNA sequencing (RNASeq), in which the total RNA expressed in a sample is sequenced and the number of sequenced RNA copies from a gene is used as quantitative data of the gene expression (Ozsolak et al. 2011). Each approach has its strengths and weaknesses, but they are both well established and widely used (Tachibana 2015).

Good quality of high-throughput technology-based findings, such as those produced by “omics” approaches, is dependent on high quality samples (Alsagaby 2019b; Horgan et al. 2011; Lowe et al. 2017). Preserved integrity of RNA sample is greatly needed to truly reflect the status of in vivo transcriptome (Husseini et al. 2021). RNA degradation is a

✉ Suliman A. Alsagaby
s.alsagaby@mu.edu.sa

¹ Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Majmaah University, AL-Majmaah 11952, Saudi Arabia

physiological process that is tightly regulated to determine the fate of RNA (Houseley et al. 2009). Sample storage and in vitro handling negatively impact RNA integrity (Hatzis et al. 2011; Koh et al. 2021a). Therefore, the assessment of RNA integrity is an essential quality control check prior to performing a transcriptomics analysis. The RNA integrity number (RIN) that is calculated on the basis of an entire electrophoretic trace of RNA using Agilent's Bioanalyzer provides a simple and reproducible method for determining RNA quality (Mueller et al. 2004; Schroeder et al. 2006). As a result, RIN is widely accepted as quality control checker of RNA quality in transcriptomics studies (Dong et al. 2013). Although there has been no consensus on a RIN value to be used as determinant for sample inclusion, RIN value equal or above 6 appeared to be most commonly reported in transcriptomics studies (Alsagaby et al. 2020; Hoffman 2004; Letzkus et al. 2014; Vartanian et al. 2009).

In clinical settings, where sample storage or RNA isolation and processing cannot be made immediately, RNA decaying is unavoidable (Botling et al. 2009; Chung et al. 2008b; Hentze et al. 2019; Koh et al. 2021b; Ohashi et al. 2004). Moreover, in cancer patients, chemotherapy disrupts RNA of malignant cells, reducing their RIN value

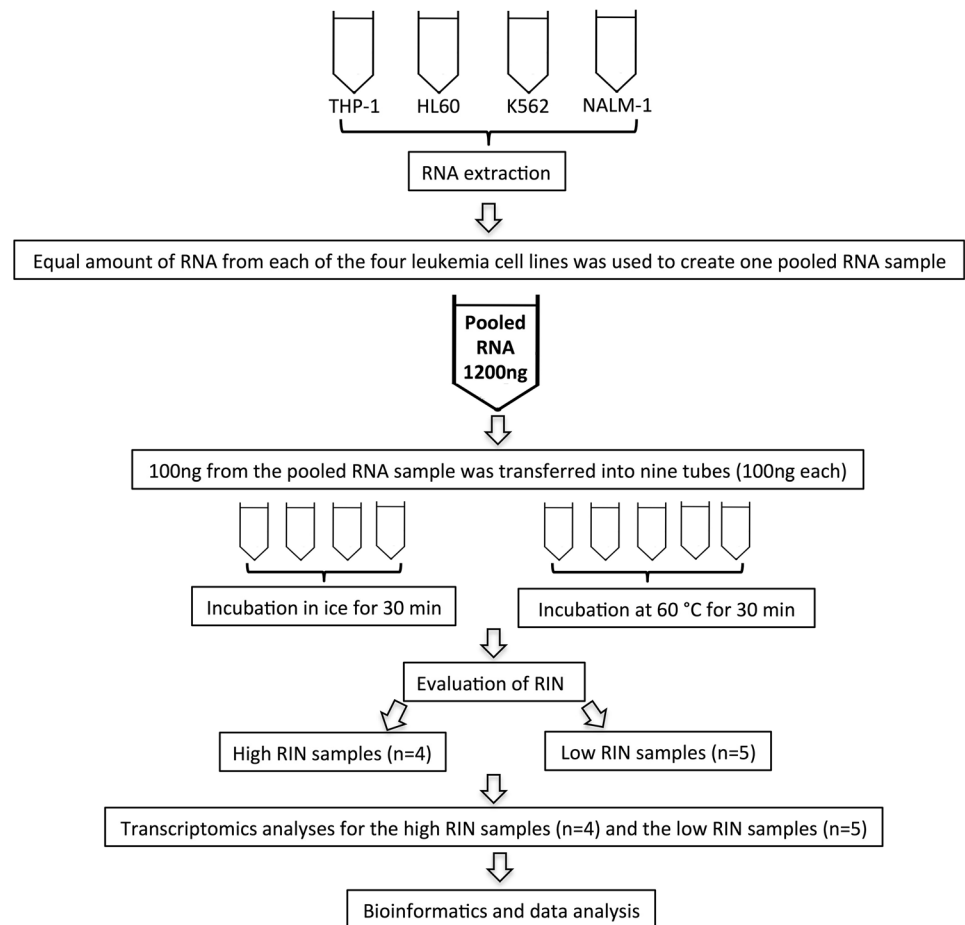
(Narendrula et al. 2016). In fact, low RIN value of cancer cells was proposed as a biomarker of good response to chemotherapy (Parissenti et al. 2015). Therefore, in these situations investigators are left with no other option but to use samples with compromised RNA integrity. In the present study, the goal was to investigate the impact of RNA degradation on the complete transcriptome of leukemia cells.

Materials and methods

Cell culture

The methodology workflow of the current study is represented in Fig. 1. To make the present study relevant to different types of leukemia, four human leukemia cell lines (representing different type of leukemia) were obtained from the American Type Culture Collection (ATCC) and used in this study; THP-1 (cell line of acute monocytic leukemia), HL60 (cell line of acute promyelocytic leukemia), K562 (cell line of chronic myelogenous leukemia) and NALM-1 (cell line of acute lymphoblastic leukemia). RPMI 1640 medium supplemented with 10% fetal calf serum (FCS), 100 U/mL

Fig. 1 Workflow of the major steps of the study's methodology. Four leukemia cell lines were cultured. Next, RNA was extracted from the leukemia cell lines and 300 ng RNA from each leukemia cell line was transferred into one tube to make a pooled RNA sample (1200 ng). Then, 100 ng from the pooled RNA sample was transferred into nine tubes (100 ng each). Four tubes were kept in ice for 30 min and the other five tubes were incubated at 60 °C for 30 min. Post incubation, RIN was measured in the nine tubes. Next, transcriptomics analyses were conducted on the high RIN samples ($n=4$) and the low RIN samples ($n=5$). Finally, bioinformatics and data analysis were conducted on the transcriptomics profiles. *RIN* RNA integrity number



penicillin, 40 µg/mL gentamycin, 100 µg/mL streptomycin sulphate, 4.5 mg/mL glucose and 2 mg/mL sodium bicarbonate was used to cultivate the cancer cells at 37 °C under an atmosphere of humidified air containing 5% CO₂.

Extraction of RNA and measurement of RNA integrity number

RNeasy Mini Kits from QIAGEN were used to extract total RNA from the leukemia cells. Next, NanoDrop ND-1000 spectro-photometer (Thermo Fisher Scientific, Inc.) was employed to evaluate the purity and measure the quantity of the isolated RNA. Samples that were admitted for further analysis were those that possessed a good purity of the extracted RNA (260 nm/280 nm > 2).

Induction of RNA degradation and evaluation of RNA integrity number

Equal amount (300 ng) of the extracted RNA from each one of the leukemia cell lines (THP-1, HL60, K562 and NALM-1) was transferred into a single tube and mixed to create a pool sample (RNA = 1200 ng). Next, 100 ng of the pooled RNA was transferred into nine tubes (each tube contained 100 ng). Of these tubes, four were kept in ice for 30 min, whereas the other five were incubated in water bath at 60 °C for 30 min. The RNA integrity number (RIN) in each of the nine tubes was then assessed using the materials and instructions of the Agilent Nano 6000 chip kit (capillary electrophoresis assay) and the Agilent 2100 Bioanalyzer. Post running the RNA samples in the Agilent Nano 6000 chips, the Agilent 2100 Bioanalyzer produces a gel image with an electropherogram of the analyzed RNA and calculate the RIN.

Transcriptomics profiling

The Agilent procedure of “one-color microarray-based gene expression analysis” was used to conduct transcriptomic profiling of the leukemia cells (nine samples: four samples with RIN ≥ 7.9 and five samples with RIN ≤ 3.8). Briefly, the low input quick amp labeling (one-color) kit (Agilent) was applied to 80 ng of the extracted RNA and the RNA spikes from RNA Spike-In Kit, One-Color (Agilent) in order to convert them to cDNA. Next, using the same kit, the cDNA was transcribed and amplified to fluorescent-labeled cRNA. The materials and protocol of RNeasy Mini Kits (QIAGEN) was used to purify the labeled cRNA. The quality and quantity of the labeled cRNA was assessed using NanoDrop ND-1000 spectro-photometer (Thermo Fisher Scientific, Inc.). Samples that were further processed were those that exhibited the following characteristics: with 260 nm/280 nm > 2, cRNA yield ≥ 825 ng and

specific activity (pmol Cy3/µg cRNA) ≥ 6. To prepare the hybridization samples, 600 ng of the labeled cRNA with 5 µL of 10× Gene Expression Blocking Agent (Agilent) and 1 µL of 25× Fragmentation Buffer (Agilent) were mixed (sample volume = 25 µL). Incubation for 30 min at 60 °C was conducted to induce fragmentation of the labeled cRNA in hybridization samples. Post cooling of the hybridization samples, the cleaving of the cRNA was stopped by adding 25 µL of 2× Hi-RPM hybridization Buffer (Agilent) (sample volume = 50 µL). Next, 40 µL of each hybridization sample was transferred onto a gasket that was assembled to SurePrint G3 Human Gene Expression 8×60 K Microarray slide (Agilent) and incubated at 65 °C for 17 h. Post incubation, the Gene Expression Washing Buffer 1 and 2 with 0.005% Triton X-102 (Agilent) were used to the microarray slides. The scanning of the microarray slide was performed using Agilent SureScan Microarray Scanner with the AgilentG3_GX_1color protocol. Agilent Feature Extraction Software (version 11.0.1.1; Agilent Technologies, Inc.) was employed to extract the quantitative data of probe features from the scanned microarray slide. The quantitative data were saved for further analysis.

Data analysis and filter criteria

The quantitative data extracted earlier from the microarray slides (see previous section) were loaded into the GeneSpring GX version 12.1 software package (Agilent Technologies, Inc.). Next, the raw signal data of the probes were log₂ transformed and normalized using the percentile shift algorithm with percentile target = 75. The analysis was restricted to probes with a flag call of either “detected” or “not detected”. The study included two of conditions: RIN ≥ 7.90 (four samples) and RIN ≤ 3.80 (five samples). The flag “detected” in all samples in one of the two conditions was used to filter the probes included in the analysis. Differentially expressed genes (DEGs) in the two conditions were identified on according to fold change (FC) that set at twofold or greater with corrected *p* value ≤ 0.05. Cluster analyses and heatmap presentations of DEGs were conducted using hierarchical algorithm and were based on probes (normalized signal value of probes) and conditions with Euclidean method for similarity measure and Wards for linkage rule. gProfiler (Raudvere et al. 2019) was employed for the gene ontology (GO) enrichment analysis; the statistical domain scope was set to be “only annotated genes” and corrected *p* value cut-off was set at ≤ 0.05.

Statistical analysis

The *p* values were calculated on the basis of Student’s *t*-test. For the corrected *p* values, Benjamini–Hochberg false discovery rate (FDR) was used. GeneSpring GX

version 12.1 software package (Agilent Technologies, Inc.) was employed for calculation of p values and corrected p values of the DEGs. The p value of the other analyses was calculated using excel software. Column graphs were also constructed using Excel software.

Results

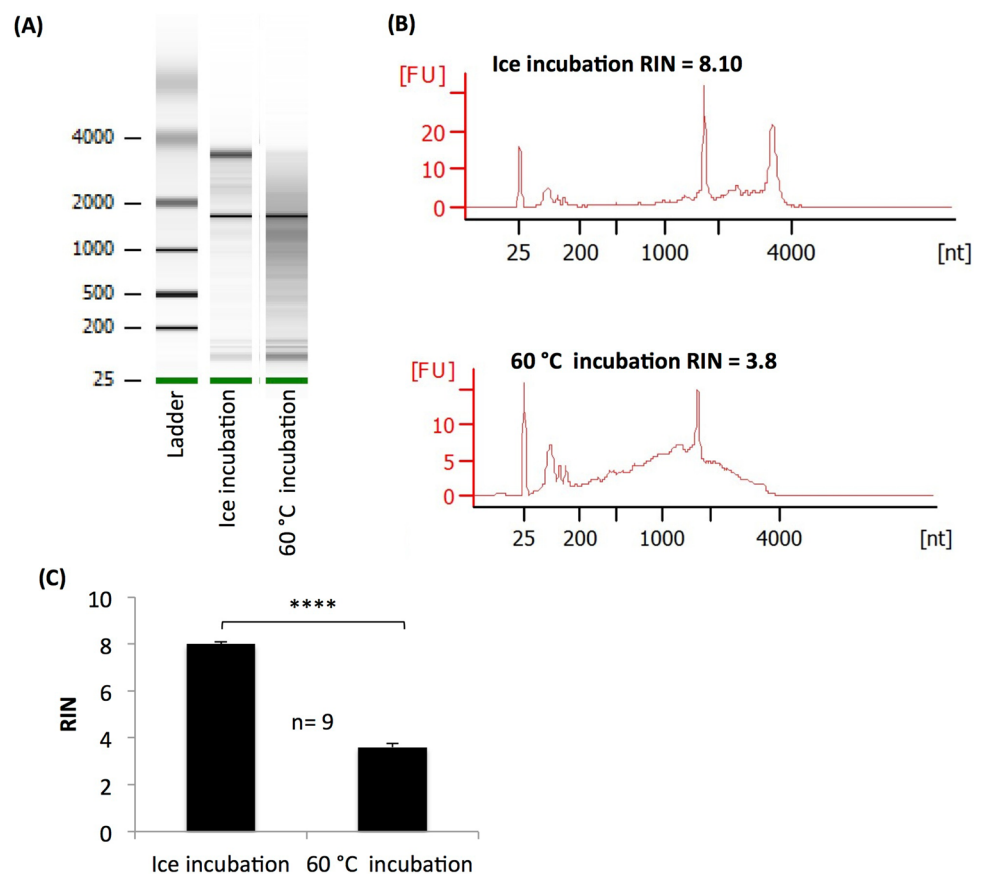
Evaluation of RNA integrity number

The RNA integrity number (RIN) of RNA samples ($n=4$) that were incubated in ice for 30 min and the RNA samples ($n=5$) that were incubated at 60 °C for 30 min was measured (Fig. 2). RIN ranged from 7.90 to 8.10 (average = 8) for the samples that were kept in ice for 30 min (these samples termed high RIN) and from 3.40 to 3.80 (average = 3.60) for the samples that were incubated at 60 °C for 30 min (these samples denoted low RIN). As expected, heat induced a significant drop in the RIN (Fig. 2; $p = 3 \times 10^{-10}$). The high RIN samples ($n=4$) and low RIN samples ($n=5$) were used for the subsequent transcriptomics analysis.

Impact of RNA degradation on transcriptome profile

In attempt to investigate the impact of RNA degradation on the global gene expression profiling, DNA-microarray-based transcriptomics analysis was conducted for each one of the high RIN samples ($n=4$) and the low RIN samples ($n=5$). Then, the transcriptomic profiles of the high RIN samples were compared with that of the low RIN samples. The comparison was limited to genes (probes) that were detected in all samples of either condition; high RIN or low RIN (number of probe IDs = 29,230). Principle component analysis (PCA) showed that compromised RNA integrity induced noticeable changes in the transcriptomes of the samples; the low RIN samples were clustered together and distinctly separated from the high RIN samples (Fig. 3A). Next, fold change (\geq two-fold) with p value ≤ 0.03 and corrected p value ≤ 0.05 were used to identify alteration in the relative quantitation of gene expression in the low RIN samples versus the high RIN samples. A total of 1945 genes (probe IDs), representing 6.7% of the studied transcriptomes, showed significant changes in their expression in the low RIN samples compared with the high RIN samples (Online Resource 1). In contrast, 27,285 genes (probe IDs), which represent 93.3% of the examined transcriptomes, showed stable expression in two groups of samples, as they did not meet the criteria made

Fig. 2 Heat-induced degradation of RNA. The RNA integrity number (RIN) was determined using Nano 6000 chips and the Agilent 2100 Bioanalyzer in two group of samples: RNA samples incubated in ice for 30 min ($n=4$) and RNA samples incubated at 60 °C for 30 min ($n=5$). Representative examples of electrophoretic trace of RNA samples are shown (A, B). Heat induced a significant decrease in the RIN values (C). The p values were calculated on this basis Student's t -test using Excel software. Asterisk symbol: **** is p value ≤ 0.0001



to defined altered relative gene expression; these genes were termed “stable genes”. Figure 3B shows a volcano plot of the evaluated transcriptomes of the two groups of samples. Of the 1945 genes, 1146 (59%) showed reduced quantitative data in the low RIN samples compared with high RIN samples (denoted down-represented genes). These genes were down-represented probably because they were more degradable than the rest of the transcripts. In contrast, 781 genes (41%) of the 1945 exhibited increased quantitative data in the low RIN samples as opposed to the high RIN samples (termed over-represented genes). These genes were over-represented perhaps because they were more stable (less degradable) compared with rest of the transcripts. Similar to PCA, cluster analysis with heatmap presentation grouped the low RIN samples closely together and high RIN samples closely together; and identified genes that were impacted by the heat-induced RNA decoy (Fig. 3C, $FC \geq$ two-fold, p value ≤ 0.03 , corrected p value ≤ 0.05).

Type of RNA enrichment

Next, investigation on the type of RNA among the three groups of genes (the stable genes, over-represented genes and down-represented genes) was conducted to determine whether a type of RNA was enriched in either one of the three groups (Fig. 4). Protein coding RNA appeared to dominate the three groups of genes; the stable genes showed the largest proportion (58%) followed by the down-represented genes (53%) and the over-represented genes (46%). The recorded percentage of other types of RNA was below 20% in the three groups of genes. Interestingly, long non-coding RNA (lncRNA) showed the least proportion in the down-represented genes (3%) versus 10% and 12% for the stable genes and the over-represented genes, respectively. In contrast, retained introns showed bigger percentage (17%) in the down-represented genes compared with the stable genes (11%) and the over-represented genes (11%).

Sequence characteristics

Next, analyses were conducted on the three groups of genes with respect to their sequence features. The analysis reported no significant differences between the three groups of genes in the context of 5' UTR length (Fig. 5A). In contrast, the over-represented genes ($p \leq 0.01$) and the down-represented genes ($p \leq 0.0001$) possessed significantly shorter 3' UTR compared with the stable genes (Fig. 5B). Transcript length and gene length appeared to be the longest in the over-represented genes ($p \leq 0.0001$) and the shortest in the down-represented genes ($p \leq 0.0001$, Fig. 5C, D). The percentage of GC was found to be higher in the stable genes ($p \leq 0.0001$) and down-represented genes ($p \leq 0.0001$) compared with over-represented genes (Fig. 5E). The distance between a

transcript complementary sequence to its corresponding probe and the 5' end of the same transcript was the longest in the over-represented genes ($p \leq 0.0001$) and the shortest in the down-represented genes ($p \leq 0.0001$, Fig. 5F).

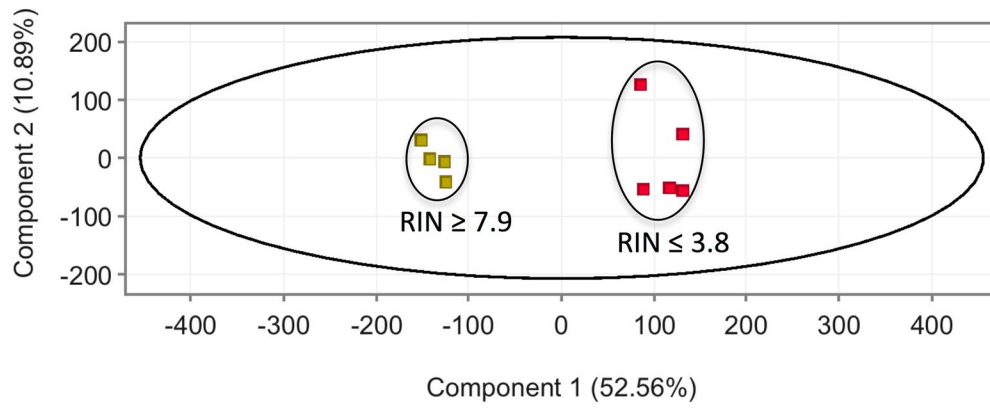
Gene ontology enrichment

Gene ontology (GO) analysis using gProfiler was performed on the over-represented and the down-represented genes in order to determine whether significant enrichments of GO terms can be found in the two groups of genes (Tables 1 and 2). Protein binding appeared to be a significantly enriched molecular function term in both groups of genes. Nevertheless, catalytic activity ($p = 0.007$) and enzyme binding ($p = 0.007$) were specifically enriched in the down-represented genes. On the other hand, nucleic acid binding ($p = 0.04$), DNA binding ($p = 0.05$) and chromatin binding ($p = 0.05$) were enriched only in the over-represented genes. While metabolic process was commonly enriched as a biological process term in the two groups of genes, metabolic process of RNA ($p = 3.97 \times 10^{-5}$) and nucleic acid ($p = 8.93 \times 10^{-5}$) appeared to be specifically enriched in the over-represented genes. Furthermore, the over-represented genes uniquely enriched for biological processes, such as regulation of gene expression ($p = 3.97 \times 10^{-5}$), positive regulation of gene expression ($p = 0.0002$), positive regulation of RNA biosynthetic process ($p = 0.0002$) and mRNA transport ($p = 0.0003$). Nucleus, nuclear lumen and nucleoplasm were commonly enriched in both groups of genes. However, cytoplasm ($p = 2.69 \times 10^{-19}$), vesicle ($p = 1.47 \times 10^{-18}$), Golgi-associated vesicle ($p = 0.001$) and endoplasmic reticulum ($p = 0.001$) were enriched specifically in the down-represented genes.

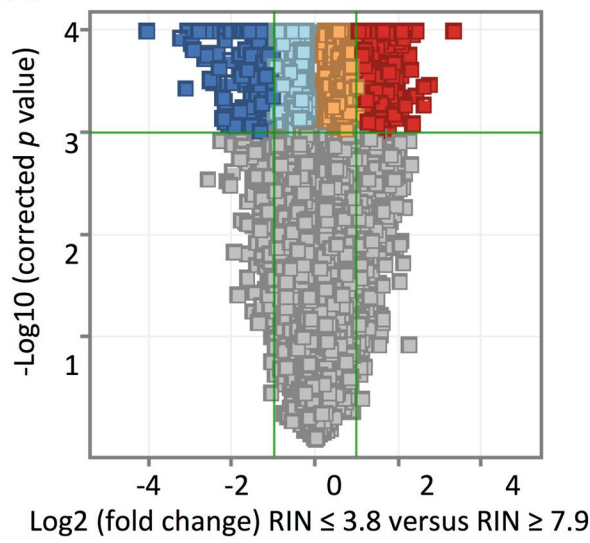
Discussion

In global gene expression studies, preserving RNA integrity is a very essential step to insure true reflection of cellular transcriptome (Lightfoot et al. 2005; Lowe et al. 2017). However, saving RNA from decoying in clinical samples is not always possible despite the big effort made in the optimization and standardization of protocols that deal with the extraction, handling and processing of RNA (Chung et al. 2008a; Debey et al. 2006; Farina et al. 2014; Hatzis et al. 2011; Hewitt et al. 2008). As a result, working on clinical samples with compromised RNA integrity is inevitable in some cases. In such cases, knowledge of: (i) to what extent low integrity of RNA influences overall transcriptomic profile; (ii) what are the name of transcripts or what are the characteristics of transcripts whose relative quantification does not show true measurement of gene expression due to RNA degradation, may provide a mean through which

(A)



(B)



(C)

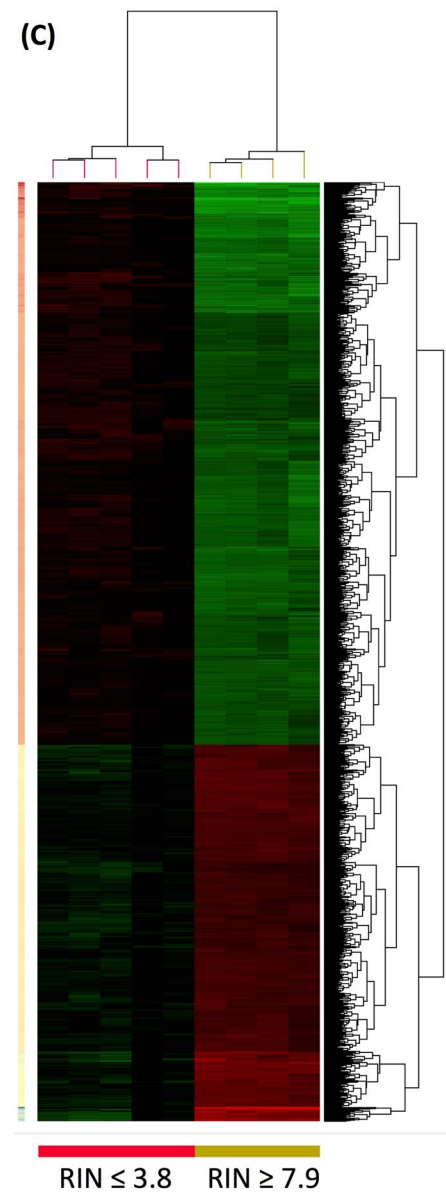


Fig. 3 Heat-induced degradation of RNA alters the transcriptome of human cancer cells. Principle component analysis (A) shows that low RIN samples ($RIN \leq 3.8$) were grouped together and distinctly separated from the high RIN samples ($RIN \geq 7.9$). Volcano plot (B) compares the gene expression in the two groups of samples. Cluster analysis with heatmap presentation (C) also grouped the low RIN samples together and the high RIN samples together; and identifies genes with altered quantification ($FC \geq$ two-fold, p value ≤ 0.03 , corrected p value ≤ 0.05)

transcripts with RNA decoy-dependent altered expression can be excluded from the transcriptomic profile. This may give tolerance for working on low quality RNA samples without compromising the credibility of the results. Following the same concept and attempting to contribute to such knowledge, the present work studied the impact of heat-induced RNA degradation on the transcriptome of human leukemic cells.

The results of the PCA, the cluster analysis and the volcano plot showed RNA degradation-dependent changes in the transcriptomes of the human leukemic cells. However, these alterations appeared to have happened to a small proportion of the studied transcriptomes. In fact, only 6.7% (1945 transcripts; $FC \geq 2$, p value ≤ 0.03 , corrected p value ≤ 0.05) of the leukemic cells transcriptome was affected by RNA degradation. In contrast, 93.3% (27,285 transcripts) of the transcriptomes showed stable expression. This finding may give rationale that working on precious samples from leukemia patients with low RIN values might be justifiable, as the vast majority of the transcripts studied here appeared to be stable. This idea becomes more sensible when pre-existing knowledge of the transcripts that are likely to be affected by RNA degradation is available for researchers. This way, investigators can rule out these transcripts (whose relative quantification is prone to alteration

by RNA degradation) from the rest of the transcriptomic profile.

Normalization of the relative quantitative data of DNA microarray experiments is a main step that has to be performed in order to reduce the impact of technical (handling) variations and make the transcriptomic profiles comparable (Do et al. 2006). The normalization done on this study may explain the over-representation of the 781 transcripts ($FC \geq 2$, p value ≤ 0.03 , corrected p value ≤ 0.05). These genes are the least transcripts affected by the thermal degradation of RNA compared with the rest of the transcripts. In contrast, the down-representation of the 1146 transcripts ($FC \geq 2$, p value ≤ 0.03 , corrected p value ≤ 0.05) indicates that these transcripts are the most affected by the thermal decoy of RNA compared with the rest of the transcripts. Taken together, the over and down-represented transcripts are of interest for researchers working on the transcriptome of clinical leukemic samples with low RIN value, as altered expression of these transcripts may highlight the poor integrity of RNA samples more than variation in the biology of the disease between samples.

Transcript length appeared in the present study to be a determinant factor of whether a transcript is sensitive to thermal degradation; long transcripts (over-represented genes) were the least affected by the thermal-induced degradation, whereas short transcripts (down-represented genes) were the most sensitive to heat-induced decoy. Similar findings were reported in colorectal cancer cells, where lost RNA integrity had the least impact on long transcripts and the most effect on short RNA (Opitz et al. 2010). Another stability marker of transcript against thermal-induced degradation was the distance between a probe complementary sequence and 5' end of a transcript. This distance was found to be the longest in the over-represented transcripts and shortest in the down-represented transcripts. Different mechanisms were

Fig. 4 Comparison of the types of RNA in the three groups of genes. Ensemble database was interrogated to identify the type of RNA transcribed from the stable genes, the over-represented genes and the down-represented genes. *lncRNA* long noncoding RNA

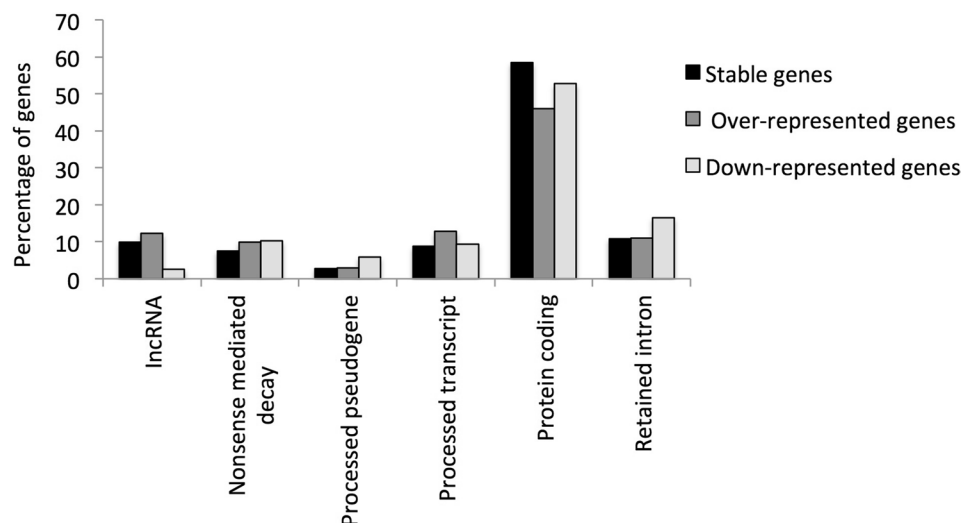
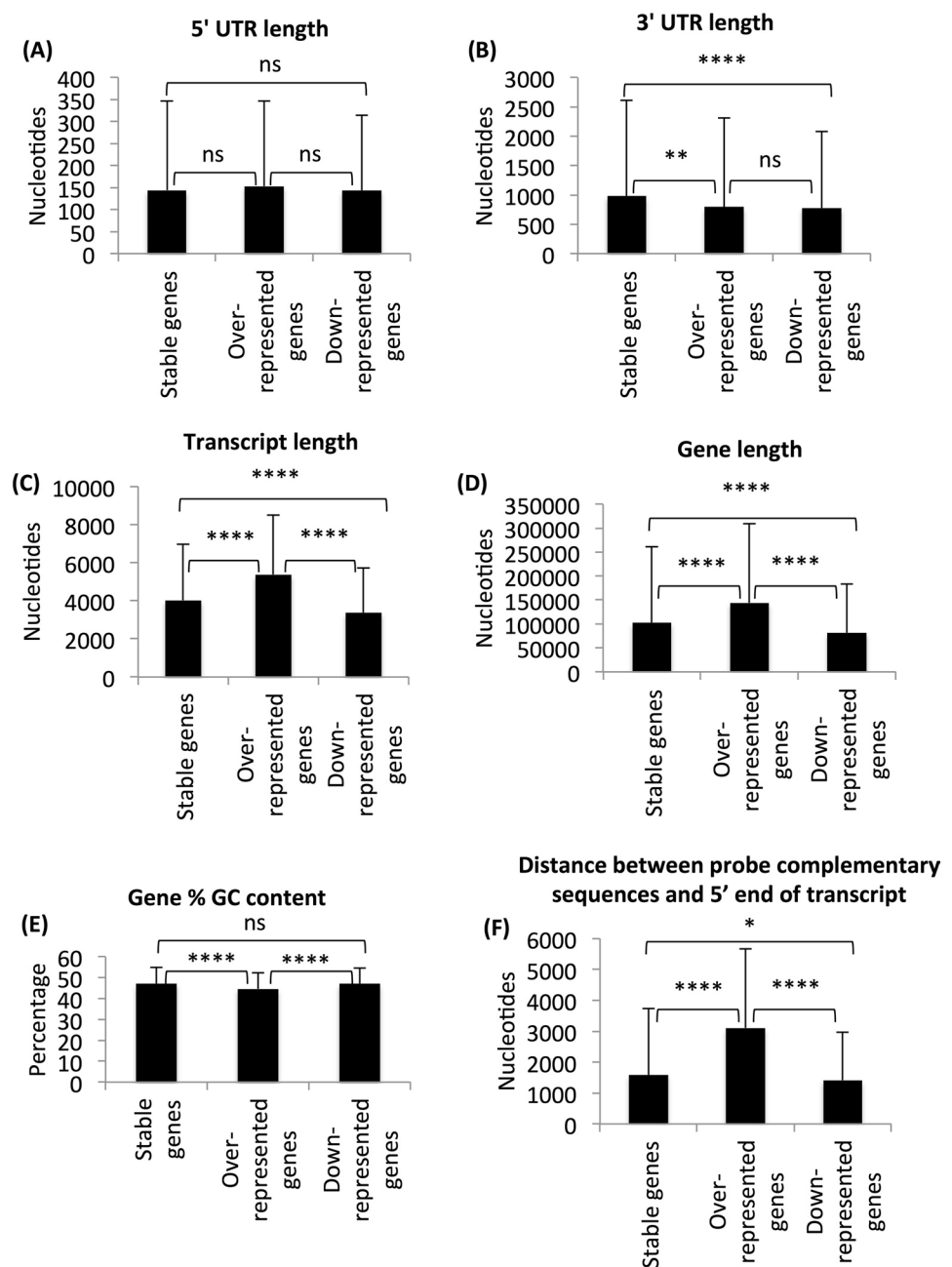


Fig. 5 Sequence characteristics of the three groups of genes. Sequence features were extracted from Ensembl database and NCBI reference RNA sequence database. 5' UTR length of transcripts (A); 3' UTR length of transcripts (B); transcript length (C); gene length (D); percentage of GC content in transcripts (E); distance between probe positions and the 5' end of transcripts (F). *UTR* untranslated region. The *p* values were calculated on this basis Student's *t*-test using Excel software. Asterisk symbols: * is *p* value ≤ 0.05 ; ** is *p* value ≤ 0.01 ; *** is *p* value ≤ 0.001 ; **** is *p* value ≤ 0.0001



described for RNA degradation in eukaryotic cells (Parker et al. 2004; Tatosyan et al. 2020), one of which was shown to take place in the direction from 5' to 3' (Cannistraro et al. 1985). The fact that the conversion of RNA to cDNA in Agilent DNA microarray was conducted using a universal primer that base pairs with the poly A tail (3' end); and the direction of RNA digestion from 5' to 3' made the majority of Agilent probes to be complementary to sequences close to the 3' end of transcripts. This may explain why long transcripts and those with long distance between their 5' end the sequence complementary to their corresponding probes were the least affected by thermal-induced RNA degradation.

The over and down-represented genes significantly enriched for GO terms. Transcripts involved in catalytic activity and enzymatic binding appeared to be most affected by RNA degradation. On the other hand, transcripts that play roles in nucleic acid binding, regulation of RNA biosynthesis, and mRNA transport were the least sensitive to RNA decoy. The enrichment of GO terms here was related to the sequence features of the down-represented and over-represented genes. For example, in contrast to the down-represented genes that enriched for catalytic activity and enzymatic binding, the over-represented genes that enriched for nucleic acid binding, regulation of RNA biosynthesis,

Table 1 Gene ontology enrichment analysis of the down-represented genes

	Term name	Corrected <i>p</i> value (FDR)	Term size	Intersection size
Molecular function	Protein binding	2.66×10^{-8}	14,752	749
	Binding	0.0007	17,052	816
	Catalytic activity	0.0007	5879	334
	Enzyme binding	0.0007	1866	128
Biological process	Cellular metabolic process	0.00007	10,855	574
	Metabolic process	0.01	11,998	606
	Response to nutrient	0.01	183	23
	Cellular component organization or biogenesis	0.01	6767	367
	Protein metabolic process	0.01	5806	321
	Phosphorus metabolic process	0.01	3125	188
	Tricarboxylic acid cycle	0.01	36	9
	Vacuole organization	0.01	182	22
	Protein modification process	0.02	4027	232
	Protein catabolic process	0.02	983	72
	Autophagy	0.03	545	45
	Macromolecule modification	0.03	4245	240
	Macroautophagy	0.03	312	30
	Organic substance catabolic process	0.03	2271	140
	Organic substance metabolic process	0.04	11,403	571
	Endoplasmic reticulum to Golgi vesicle-mediated transport	0.04	207	22
	Cell component	Autophagosome organization	0.04	102
Cytoplasm		2.69×10^{-19}	11,888	659
Membrane-bounded organelle		1.47×10^{-18}	12,836	693
Vesicle		5.61×10^{-10}	4071	265
Nuclear lumen		6.05×10^{-8}	4343	269
Endomembrane system		1.76×10^{-7}	4633	281
Endomembrane system		1.76×10^{-7}	4633	281
Nucleus		0.0003	7565	399
Nucleolus		0.0008	938	69
Golgi-associated vesicle		0.001	178	21
Endoplasmic reticulum		0.002	1975	123

FDR False discovery rate, *term size* the number of genes that were annotated to a term, *intersection size* the number of genes found in the present study that were annotated to a term

and mRNA transport had longer gene and transcript length and longer distance between the probe position and their 5' end. As expected from the experimental design of this study, where the thermal treatment was applied directly on isolated RNA rather on living cells, GO terms (or pathways) like heat shock response, apoptosis, necrosis, DNA fragmentation, or RNA disruption, which would have been activated if the thermal treatment was applied directly to the cell (Dinh et al. 2001; Furusawa et al. 2011; Ren et al. 2019), were not enriched.

Overall, the present work provided valuable insights into the impact of RNA degradation on the transcriptomes of leukemia cells, and highlighted the genes and the characteristics of genes whose relative quantification

is influenced by RNA decoy. Therefore, this study has the potential to help investigators to distinguish between RNA decoy-dependent changes in genes expression from pathology-dependent alteration of gene expression when low quality RNA samples are the only available material to use. However, cautions should be taken when relating the current findings to clinical settings because of two reasons: (i) the present work was conducted on human leukemic cell lines rather than clinical samples from leukemic patients, (ii) the RNA degradation induced by thermal treatment was a simulation (not necessarily identical) to RNA decoy of patient samples in clinical settings. Further evaluation of the relative expression of some of the over-represented and under-represented using quantitative

Table 2 Gene ontology enrichment analysis of the over-represented genes

	Term name	Corrected <i>p</i> value (FDR)	Term size	Intersection size
Molecular Function	Protein binding	4.92×10^{-5}	14,752	330
	Histone binding	0.03	241	15
	Nucleic acid binding	0.04	4353	115
	DNA binding	0.04	2519	74
	Chromatin binding	0.05	604	25
Biological Process	Regulation of RNA metabolic process	3.97×10^{-5}	3850	118
	Regulation of nucleobase-containing compound metabolic process	3.97×10^{-5}	4123	125
	Regulation of gene expression	3.97×10^{-5}	5162	147
	Regulation of macromolecule metabolic process	6.34×10^{-5}	6626	176
	Nucleic acid metabolic process	8.93×10^{-5}	5247	146
	Positive regulation of gene expression	0.0002	2396	79
	Positive regulation of nucleic acid-templated transcription	0.0002	1679	61
	Positive regulation of RNA biosynthetic process	0.0002	1680	61
	Nucleobase-containing compound metabolic process	0.0002	5766	154
	Positive regulation of transcription by RNA polymerase II	0.0002	1276	50
	mRNA transport	0.0003	154	14
	Positive regulation of transcription, DNA-templated	0.0003	1592	58
	RNA metabolic process	0.0003	4732	131
	Gene expression	0.0006	6270	161
	Nucleobase-containing compound biosynthetic process	0.0007	4211	117
	Cellular macromolecule metabolic process	0.0007	8357	202
	Regulation of transcription, DNA-templated	0.0008	3494	101
	Nitrogen compound metabolic process	0.001	10,057	233
	Nucleic acid-templated transcription	0.001	3734	105
	RNA biosynthetic process	0.002	3749	105
	Transcription by RNA polymerase II	0.002	2735	82
	Organic substance metabolic process	0.007	11,403	253
	Protein metabolic process	0.02	5806	142
Cell cycle	0.03	1914	56	
Cell Component	Nucleoplasm	5.21×10^{-6}	3991	122
	Nuclear lumen	4.1×10^{-6}	4343	124
	Nucleus	0.001	7565	182
	Plasma membrane region	0.04	1255	38

FDR False discovery rate, *term size* the number of genes that were annotated to a term, *intersection size* the number of genes found in the present study that were annotated to a term

real-time polymerase chain reaction (qRT-PCR) may add validity to the transcriptomics data. However, limitations in the project budget and in the logistic support, which has been caused by the COVID19 pandemic, did not allow performing additional experimental work for the assessment of gene expressing using qRT-PCR. Our planned future work will apply the same concept of the present work on leukemia samples isolated from leukemia patients to identify genes whose transcripts' relative quantitation is sensitive to RNA degradation in clinical settings.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13205-022-03223-1>.

Acknowledgements The author would like to thank the Deanship of Scientific Research at Majmaah University in Saudi Arabia for supporting this study for supporting this study under project number [R-2022-204].

Author contribution The entire study was conducted by SAA.

Declarations

Conflict of interest The author has no competing interests to declare that are relevant to the content of this article.

Research involving human participants and/or animals Not applicable because the research was not conducted on human participants and/or animals.

Informed consent Not applicable because the research was not conducted on human participants.

References

- Alsagaby SA (2019a) Transcriptomics-based validation of the relatedness of heterogeneous nuclear ribonucleoproteins to chronic lymphocytic leukemia as potential biomarkers of the disease aggressiveness. *Saudi Med J* 40:328–338
- Alsagaby SA (2019b) Understanding the fundamentals of proteomics. *Curr Top Pept Protein Res* 20:25–39
- Alsagaby SA, Vijayakumar R, Premanathan M, Mickymaray S, Alturaiki W, Al-Baradie RS, AlGhamdi S, Aziz MA, Alhumaydhi FA, Alzahrani FA (2020) Transcriptomics-based characterization of the toxicity of ZnO nanoparticles against chronic myeloid leukemia cells. *Int J Nanomed* 15:7901–7921
- Barbulovic-Nad I, Lucente M, Sun Y, Zhang M, Wheeler AR, Bussmann M (2006) Bio-microarray fabrication techniques—a review. *Crit Rev Biotechnol* 26:237–259
- Botling J, Edlund K, Segersten U, Tahmasebpoor S, Engström M, Sundström M, Malmström P-U, Micke P (2009) Impact of thawing on RNA integrity and gene expression analysis in fresh frozen tissue. *Diagn Mol Pathol* 18:44–52
- Cannistraro VJ, Kennell D (1985) Evidence that the 5' end of lac mRNA starts to decay as soon as it is synthesized. *J Bacteriol* 161:820–822
- Chung J-Y, Braunschweig T, Williams R, Guerrero N, Hoffmann KM, Kwon M, Song YK, Libutti SK, Hewitt SM (2008a) Factors in tissue handling and obtained from formalin-fixed, processing that impact RNA paraffin-embedded tissue. *J Histochem Cytochem* 56:1033–1042
- Chung J-Y, Braunschweig T, Williams R, Guerrero N, Hoffmann KM, Kwon M, Song YK, Libutti SK, Hewitt SM (2008b) Factors in tissue handling and processing that impact RNA obtained from formalin-fixed, paraffin-embedded tissue. *J Histochem Cytochem* 56:1033–1042
- Debey S, Zander T, Brors B, Popov A, Eils R, Schultze JL (2006) A highly standardized, robust, and cost-effective method for genome-wide transcriptome analysis of peripheral blood applicable to large-scale clinical trials. *Genomics* 87:653–664
- Dinh H-KB, Zhao B, Schuschereba ST, Merrill G, Bowman PD (2001) Gene expression profiling of the response to thermal injury in human cells. *Physiol Genom* 7:3–13
- Do JH, Choi D-K (2006) Normalization of microarray data: single-labeled and dual-labeled arrays. *Mol Cells* 22:2
- Dong Z, Chen Y (2013) Transcriptomics: advances and approaches. *Sci China Life Sci* 56:960–967
- Farina NH, Wood ME, Perrapato SD, Francklyn CS, Stein GS, Stein JL, Lian JB (2014) Standardizing analysis of circulating microRNA: clinical and biological relevance. *J Cell Biochem* 115:805–811
- Furusawa Y, Tabuchi Y, Wada S, Takasaki I, Ohtsuka K, Kondo T (2011) Identification of biological functions and gene networks regulated by heat stress in U937 human lymphoma cells. *Int J Mol Med* 28:143–151
- Hatzis C, Sun H, Yao H, Hubbard RE, Meric-Bernstam F, Babiera GV, Wu Y, Pusztai L, Symmans WF (2011) Effects of tissue handling on RNA integrity and microarray measurements from resected breast cancers. *J Natl Cancer Inst* 103:1871–1883
- Hentze JL, Kringelbach TM, Novotny GW, Hamid BH, Ravn V, Christensen IJ, Høgdall C, Høgdall E (2019) Optimized biobanking procedures for preservation of RNA in tissue: comparison of snap-freezing and RNA later-fixation methods. *Bio-preserv Biobank* 17:562–569
- Hewitt SM, Lewis FA, Cao Y, Conrad RC, Cronin M, Danenberg KD, Goralski TJ, Langmore JP, Raja RG, Williams PM (2008) Tissue handling and specimen preparation in surgical pathology: issues concerning the recovery of nucleic acids from formalin-fixed, paraffin-embedded tissue. *Arch Pathol Lab Med* 132:1929–1935
- Hoffman EP (2004) Guidelines: Expression profiling—best practices for data generation and interpretation in clinical trials. *Nat Rev Genet* 5:229–238
- Horgan RP, Kenny LC (2011) 'Omics' technologies: genomics, transcriptomics, proteomics and metabolomics. *Obstet Gynaecol* 13:189–195
- Houseley J, Tollervy D (2009) The many pathways of RNA degradation. *Cell* 136:763–776
- Hussein AA, Derakhshandeh M, Tatlisu NB (2021) Comprehensive review of transcriptomics (RNAs) workflows from blood specimens. *Sep Purif Rev* 2:1–21
- Iacobas DA (2020) Powerful quantifiers for cancer transcriptomics. *World J Clin Oncol* 11:679
- Koh EJ, Yu SY, Kim SH, Kim SJ, Lee E-I, Hwang SY (2021a) Understanding confounding effects of blood handling strategies on RNA quality and transcriptomic alteration using RNA sequencing. *Bio-Chip J* 2:1–8
- Koh EJ, Yu SY, Kim SH, Kim SJ, Lee E-I, Hwang SY (2021b) Understanding confounding effects of blood handling strategies on RNA quality and transcriptomic alteration using RNA sequencing. *Bio-Chip J* 15:1–8
- Letzkus M, Luesink E, Starck-Schwartz S, Bigaud M, Mirza F, Hartmann N, Gerstmayer B, Janssen U, Scherer A, Schumacher MM (2014) Gene expression profiling of immunomagnetically separated cells directly from stabilized whole blood for multicenter clinical trials. *Clin Transl Med* 3:1–13
- Lightfoot S, Salowsky R, Buhlmann C (2005) RNA integrity number: towards standardization of RNA quality assessment for better reproducibility and reliability of gene expression experiments. *Breast Cancer Res* 7:1
- Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T (2017) Transcriptomics technologies. *PLoS Comput Biol* 13:e1005457
- Mueller O, Lightfoot S, Schroeder A (2004) RNA integrity number (RIN)—standardization of RNA quality control. *Agilent Appl* 1:1–8
- Narendrula R, Mispel-Beyer K, Guo B, Parissenti AM, Pritzker LB, Pritzker K, Masilamani T, Wang X, Lannér C (2016) RNA disruption is associated with response to multiple classes of chemotherapy drugs in tumor cell lines. *BMC Cancer* 16:1–17
- Ohashi Y, Creek KE, Pirisi L, Kalus R, Young SR (2004) RNA degradation in human breast tissue after surgical removal: a time-course study. *Exp Mol Pathol* 77:98–103
- Opitz L, Salinas-Riester G, Grade M, Jung K, Jo P, Emons G, Ghadimi BM, Beißbarth T, Gaedcke J (2010) Impact of RNA degradation on gene expression profiling. *BMC Med Genomics* 3:1–14
- Ozsolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12:87–98
- Parissenti AM, Guo B, Pritzker LB, Pritzker KPH, Wang X, Zhu M, Shepherd LE, Trudeau ME (2015) Tumor RNA disruption

- predicts survival benefit from breast cancer chemotherapy. *Breast Cancer Res Treat* 153:135–144
- Parker R, Song H (2004) The enzymes and control of eukaryotic mRNA turnover. *Nat Struct Mol Biol* 11:121–127
- Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J (2019) g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47:W191–W198
- Ren MQ, Kazman JB, Abraham PA, Atias-Varon D, Heled Y, Deuster PA (2019) Gene expression profiling of humans under exertional heat stress: comparisons between persons with and without exertional heat stroke. *J Therm Biol* 85:102423
- Rodon J, Soria J-C, Berger R, Miller WH, Rubin E, Kugel A, Tsimberidou A, Saintigny P, Ackerstein A, Braña I (2019) Genomic and transcriptomic profiling expands precision cancer medicine: the WINETHER trial. *Nat Med* 25:751–758
- Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* 7:1–14
- Supplitt S, Karpinski P, Sasiadek M, Laczanska I (2021) Current achievements and applications of transcriptomics in personalized cancer medicine. *Int J Mol Sci* 22:1422
- Tachibana C (2015) Transcriptomics today: microarrays, RNA-seq, and more. *Science* 349:544–546
- Tatosyan KA, Ustyantsev IG, Kramerov DA (2020) RNA degradation in eukaryotic cells. *Mol Biol* 54:485–502
- Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F (2017) A pathology atlas of the human cancer transcriptome. *Science* 357:2
- Vartanian K, Slotke R, Johnstone T, Casale A, Planck SR, Choi D, Smith JR, Rosenbaum JT, Harrington CA (2009) Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis. *BMC Genom* 10:1–16
- Wang X, Kaczor-Urbanowicz KE, Wong DTW (2017) Salivary biomarkers in cancer detection. *Med Oncol* 34:1–8