An inexpensive semi-automated sample processing pipeline for cell-free RNA

Mira N. Moufarrej[1] and Stephen R. Quake[1,†]

[1] Departments of Bioengineering and Applied Physics, Stanford University and Chan Zuckerberg Biohub, Stanford, CA USA 94305

[†] Corresponding author: steve@quake-lab.org

Abstract

Despite advances in automated liquid handling and microfluidics, preparing samples for RNA sequencing at scale generally requires expensive equipment, which is beyond the reach of many academic labs. Manual sample preparation remains a slow, expensive, and error-prone process. Here, we describe a low-cost, semi-automated pipeline to extract cell-free RNA (cfRNA) that increases sample throughput by 12-fold while reducing time spent and cost by nearly 11-fold and 3-fold respectively. This pipeline is generalizable for many nucleic acid extraction applications, thereby increasing the scale of studies, which can be performed in small research labs.

Introduction

Liquid biopsies offer an unprecedented opportunity to study human disease on a molecular scale and sample from the entire human body at once (Fan et al., 2012; Koh et al., 2014; Kowarsky et al., 2017; Lo et al., 2010; Pan et al., 2017; Song, Diao, Brunger, & Quake, 2016; Tsang et al., 2017). Applications of this technique, which rely on about a milliliter of blood plasma, span vast spaces like oncology (Crowley, Di Nicolantonio, Loupakis, & Bardelli, 2013; Liu et al., 2020; Song et al., 2017), transplant care (Sharon et al., 2017; Snyder, Khush, Valantine, & Quake, 2011), and prenatal diagnostics (Fan, Blumenfeld, Chitkara, Hudgins, & Quake, 2008; Ngo et al., 2018). Numerous groups and companies have now demonstrated the utility of "omic"-based blood analysis in proof-of-concept work (Aghaeepour et al., 2017, 2018; Ghaemi et al., 2019; Sharon et al., 2017; Song et al., 2017; Tsang et al., 2017). However, scaling such discovery research remains obstructed by the lack of robust sample processing. Often overlooked in manuscripts detailing novel results, sample processing – especially for human samples that may come from sources with disparate collection and storage guidelines – remains paramount to yield trustworthy results.

Most discovery-focused work has four main stages – sample collection, sample processing, data generation, and data analysis. Sample collection, data generation, and data analysis have progressed tremendously due to the advent of mature biobanks, massively parallel sequencing (e.g. NovaSeq), and cloud-based computing (e.g. Amazon Web Services, Google Cloud) respectively. Sample processing however has remained in the dark ages in most labs – still done by hand by technicians, graduate students, and post-doctoral scholars alike despite increasing sample numbers. As a result, this laborious step constrains the number of samples processed at once and can lead to batch effects.

Although the biotechnology industry has long moved to robotic liquid handling solutions (Tan & Yiap, 2009), many labs have not. Most automated solutions for lab work available today focus on

production quality pipelines and do not accommodate research-based lab work in at least one of these ways – (1) No high-volume pipetting (e.g. Bravo) (2) No easy user-level customization (e.g. Hamilton, Eppendorf) (3) High initial expense ($20K-100K) and custom consumables (e.g. Bravo, Hamilton, Eppendorf).

Here, we report on the design and validation of a semi-automated pipeline to extract cell-free RNA (cfRNA) by outfitting an existing open-source robotic platform, the Opentrons 1.0 (OT1). This platform has two key advantages – its low price point and open-source API. To use the OT1, we designed new 3D printed parts to mount a manual eight-channel large volume (1 milliliter per channel) pipette onto the robot securely, wrote new liquid handling protocols to accommodate our high-volume sample processing, and validated the platform end-to-end. Both RT-qPCR and RNA-sequencing data highlight comparable results to the gold-standard - manual sample processing. Importantly, any lab already has most of the parts required (manual pipette, corresponding tips, corresponding kits) to build a semi-automated sample processing pipeline of their own. They would only need to purchase or 3D-print a few extra parts (~$5.5K total) to start processing samples in a semi-automated manner. Such robust and scalable solutions will allow us to test hypotheses faster and spend less time on repetitive but important tasks like sample processing.

Results

System overview

Like many RNA isolation protocols, processing cfRNA samples can be decomposed into three subparts – RNA extraction, DNA digestion, and RNA cleaning and concentration (Fig 1A). Our process requires a single robotic system than can easily fit within a chemical hood – necessary to process bodily fluids – and can be configured to use any individual eight-channel pipette. Nearly all steps are performed on the robot with minimal user input. At the system's prompting, the user fills reagent boats, transfers plates for centrifugation, or swaps out tip boxes. Sample addition is performed by hand to easily accommodate various sample storage strategies (e.g. 1.5 mL microcentrifuge tubes, matrix tubes). Similarly, DNAse master mix and eluant addition are also performed by hand to maintain the system's simplicity since these steps stand out as the only small volume (<15 μL) pipetting in the process. By using standard lab equipment and interfacing with the user at key steps, our semi-automated pipeline streamlines sample processing while maintaining flexibility key to discovery stage research.

System setup

The OT1 system is composed of motors that drive motion in the x, y, and z directions across 10 deck positions where the user can place reagent boats, plates, tip boxes, and trash (Fig 1B). To use it, the user must mount their pipette of choice so that it is rigidly fixed in all directions and only moves with the motors themselves. We achieved this by designing and 3D-printing several parts to work with our pipette of choice. To fix motion in the x-direction, we mounted two U-shaped clasps (Fig 1C, dark blue arrows). These clasps are then fastened to a hard back (Fig 1C, light blue box). This back board coupled with a piece mounted on top of the pipette (Fig 1C, light blue arrow) prevents unexpected z-axis motion during tip retrieval and disposal. Finally, two more L-shaped holds (Fig 1C, green arrows) prevent y-axis motion in the pipette body, which can otherwise rotate freely. Together, these 3D-printed fixtures allow the system to remain properly calibrated across

multiple runs while mounting and disposing of tips to aspirate, mix, transfer, and dispense reagents (Fig 1D, Supp Movie 1).

Semi-automated protocol overview and validation

Like many RNA extraction protocols, the reagent volumes required to isolate cfRNA depend on initial sample volume (typically 1 mL for our process). To extract cfRNA from 1 mL of plasma requires 5 mL of added reagents yielding a maximum total volume of 6 mL. Since this is far beyond what a typical 96-well deep well plate can accommodate (~1-2 mL), we opted to use 48 deep-well plates to maintain system scalability (Fig 1A). Using a 48 deep-well plate not only interfaces well with the robotic system, but it also allows us to move away from the classic 50 mL conical tubes used for this step, since these can be difficult to process quickly in an automated fashion (Fig 1A). To process 96 samples using 48-well plates requires that the samples are split into two batches briefly and then recombined into 1 plate for subsequent steps (Fig 2A). Specifically, we thaw and extract cfRNA from 48 samples (Fig 2A left side) and then refrigerate the isolated cfRNA while thawing and processing cfRNA from another 48 samples (Fig 2A right side).

We investigated whether this refrigeration step had adverse effects on cfRNA concentration by spiking in a known concentration of a single RNA oligonucleotide at the start of extraction. We then measured its concentration using RT-qPCR pre- and post- refrigeration as compared to that for control samples processed without any refrigeration (Fig 2B left). Across 2 independent trials, we find that refrigeration of isolated cfRNA for 1.5 hours has no adverse effects on final yield (Bonferroni adjusted p-value = 0.43 and 1.0 per trial, Wilcoxon signed-rank test) (Fig 2B right). Further, we find no loss in yield for samples post-refrigeration as compared to control unrefrigerated samples (Bonferroni adjusted p-value = 1.0 and 1.0 per trial, one-sided Mann-Whitney rank test). We also separately confirmed that the custom RNA oligonucleotide and its corresponding Taqman probes work as expected (Supp Fig 1).

Subsequently once cfRNA from all 96 samples has been extracted, all samples are treated with DNAse to remove any lingering DNA, cleaned, and concentrated to a final volume of 12 µL (Fig 2A). In total, this process takes 4.5 hours with each sample at room temperature for 3 hours at a given time (Fig 2A). Finally, we quantified the likelihood of cross-contamination between samples by spiking in RNA only into every other well so that each 48-well plate then contained 24 sample wells and 24 blank wells.

We then checked for cross-contamination by measuring if RNA could be detected in blank wells after sample processing. Across 2 independent trials with 48 blank wells each (24 per 48-well plate), we find no evidence of cross-contamination (Fig 2C). Specifically, we find that blank samples do not yield a distribution of RT-qPCR cycle thresholds (C_t) significantly lower than that of true non-template controls (NTC, n = 17) (Bonferroni adjusted p-value = 0.16 and 0.10 per trial, one-sided Mann-Whitney rank test). We also note that all blank samples contained far less RNA than positive controls (PC) as indicated by higher C_t (Fig 2C).

RT-qPCR validation

To compare the semi-automated pipeline to the gold standard of traditional manual pipetting, we compared the concentration of RNA rescued for a known control relative to the concentration of

RNA expected with total rescue (ΔC_t). Over 3 independent trials per protocol, we find that the manual and semi-automated methods perform comparably (Fig 2D).

Globally, when we combine ΔC_t values from all 3 trials per protocol, we find that extracting RNA using the semi-automated method (n = 144, 48 per trial) does not significantly decrease RNA rescue when compared to the manual protocol (n = 24, 8 per trial) (Bonferroni adjusted p-value = 0.11, one-sided Mann-Whitney rank test). However, when comparing individual trials in a pairwise fashion, we do observe some significant shifts to more positive ΔC_t values specifically between all 3 manual (M) trial pairs, 1 semi-automated (SA) trial pair, and 4 of 9 manual and semi-automated trial pairs (Bonferroni adjusted p-value = 0.031, 0.0075, 0.016, 1.83×10^{-14}, 0.0001, 5.86×10^{-5}, 0.002, 0.00013 for pairs M1:M2, M1:M3, M2:M3, SA2:SA3, M1:SA1, M1:SA3, M2:SA1, M2:SA3 respectively, one-sided Mann-Whitney rank test). We note that because the manual trials (n=8) require smaller batch sizes than their semi-automated counterparts (n=48), pairwise comparison differences may be the result of sub-sampling rather than differences in protocol yield. The lack of a significant difference under global testing conditions supports this theory.

RNA-sequencing validation

Finally, we processed and sequenced 646 plasma samples collected at 2 separate locations denoted as sets 1 (n = 485) and 2 (n = 161). We then asked whether the semi-automated strategy led to significant differences in RNA quality upon sequencing compared to samples previously processed manually (n = 107). All other processing steps such as library preparation and sequencing were held constant.

First, we compared the number of hours required to isolate RNA to the number of samples processed (Fig 3A). We find that time required to process one sample sharply decreases as expected for both sets processed in a semi-automated fashion (0.056 hours) as compared to manually (0.59 hours) – nearly an 11-fold decrease in time for a 12-fold increase in samples processed. Specifically, we processed 485 samples from set 1 in 27 hours and 161 samples from set 2 in 9 hours – significantly less than the 63 hours required to process 107 samples manually.

We then asked if both methods recovered comparable amounts of cfRNA (Fig 3B). We find no decrease in RNA yield for samples prepared semi-automatedly for set 1 but find a significant drop for set 2 as compared to manually processed samples (Bonferroni adjusted p-value = 1.0, 5.7×10^{-10} for sets 1 and 2 respectively). We also find a significant decrease in RNA yield when comparing set 2 to set 1 (Bonferroni adjusted p-value = 1.8×10^{-4}, one-sided t-test).

To reconcile the discrepancy in RNA yield between sets 1 and 2, we measured other metrics of RNA quality from sequencing data. We find that across three metrics of RNA quality – degradation, mapping, and ribosomal fraction of total reads – samples from set 1 yield comparable results to manual processing (Bonferroni adjusted p-value $\geq$ 0.85 for all metrics) (Fig 3C left to right).

Samples from set 2 however contain RNA that is significantly more degraded than both manually and semi-automatedly processed set 1 samples (Bonferroni adjusted p-value = 9.18×10^{-14} and 1.71×10^{-14} respectively, one-sided t-test) (Fig3C left). To quantify sample degradation, we

leveraged that exonucleases present in plasma digest RNA from the 5' to 3' end (Sorrentino, 2010) and counted the fraction of genes in a sample for which all reads map exclusively to the 3' most exon. For manually processed samples and samples from set 1, we find that an average of 28.7% of genes detected have reads that exclusively map to the 3' exon as compared to an average of 36.2% for set 2.

Consistent with the observation that samples from set 2 contain degraded RNA, a significantly smaller fraction of reads from set 2 uniquely map to the human genome as compared to both manual samples and semi-automated set 1 (Bonferroni adjusted p-value = 7.28×10^{-30} and 2.61×10^{-53} respectively). Samples from set 1 and manual processing however yielded comparable mapped fractions (Bonferroni adjusted p-value = 0.85, one-sided t-test) with an average of 69.2% and 70.2% uniquely mapping reads as compared to 43.1% from set 2 (Fig3C middle).

Finally, we measured what fraction of reads mapped to the ribosome – an indication of how informative the RNA isolated within a given sample is likely to be (Fig 3C right). We find that no increase in ribosomal fraction in both set 1 (Bonferroni adjusted p-value = 1.0) and 2 (Bonferroni adjusted p-value = 1.0) as compared to manual samples with an average of 16.3%, 7.6%, and 12.2% for manually processed and semi-automated sets 1 and 2 respectively. However, we do identify a significant increase in ribosomal fraction for set 2 as compared to semi-automated set 1 (Bonferroni adjusted p-value = 3.45×10^{-14}).

All together, we conclude that manual and semi-automated sample processing does not yield significant changes in RNA concentration post isolation and quality upon sequencing. Differences may still arise however due to differences in sample collection strategies at distinct sites as indicated by the higher level of RNA degradation and lower quality statistics for semi-automated set 2 as compared to both manually processed samples and samples from semi-automated set 1.

Cost and time comparison

When compared to manual methods assuming a batch size of 8 samples, semi-automated parallel processing results in a 12-fold increase in sample batch size and nearly an 11- and 3-fold reduction in time and cost respectively (Table 1). We note that this RNA isolation protocol like many of its kind lends itself well to parallelization because incubation and centrifugation compose nearly half of total processing time. As sample number increases, such steps remain near constant in time. Further, reduced plastic usage, specifically by shifting to plate-based (vs individual tube) consumables and reduced tip usage, majorly contributed to decreased cost per sample processed.

Discussion

Here, we demonstrated a low-cost, semi-automated pipeline to extract cfRNA in an academic lab setting. We highlighted the benefits of such a system including an increase in sample throughput by 12-fold and reduction in time spent and cost by nearly 11-fold and 3-fold respectively. Such quantitative benefits are coupled with several qualitative improvements distinct from what one can expect from a manual protocol. Specifically, a semi-automated protocol allows for consistent sample preparation quality across users, reduced personnel time, and reduced user concentration during the protocol. When contrasted with fully automated systems, we find that a semi-automated system is well-suited to the needs of a typical academic lab performing discovery research.

Although a semi-automated system requires the user to remain nearby, its initial cost (~$5.5K) remains significantly cheaper than its automated counterparts (~$20-100K).

Importantly, the system's low cost and easy user-level customization can be applied to automate any RNA isolation required for discovery research in an academic lab. Because the system is specifically built to leverage existing 96-well kits for RNA isolation, cleaning, and concentration, such a protocol can easily be transferred to isolate RNA from other sources like tissues, urine, or other bodily fluids with relatively few changes. The user would only have to program the new liquid handling steps or modify existing ones using a common high-level programming language (Python). This easy customization stands in stark contrast to the proprietary software or lower-level programming languages like C required to modify protocols on most automated systems.

The semi-automated protocol we describe is not without its limitations – namely that the OT1 system selected here does not permit for easy pipette changes restricting the user to a single multi-channel pipette's volume range. In an academic lab setting, we do not see volume restrictions as a major issue since most RNA protocols require reagent volumes within the range of a single pipette. The user can perform any steps that require reagent volumes outside the range of the mounted pipette by repeatedly transferring a smaller volume within the upper limit for larger volumes or transferring by hand for smaller volumes. We find performing the three steps that require smaller, variable volumes by hand provides additional flexibility in the protocol. For instance, adding samples by hand easily accommodates various sample storage strategies like microtubes and matrix tubes. Additionally, adding the eluant by hand permits the user to easily modify the volume used for each RNA isolation to match their chosen downstream application.

We also found that setting up and calibrating manual pipettes to work with this platform required significant effort. These pain-points have since been addressed in the second generation of this robotic system, the OT2 (~$6.5K including pipettes). Building a system that integrates pipettes permits academic labs to skip system setup and start processing samples even faster within a reasonable cost. Although the OT1 since its use in this work has been phased out, all necessary hardware and software required to independently build the system have been made available on GitHub ("GitHub - Opentrons/otone_hardware: 3D models and Bill of Materials for the opensource OT.One lab automation platform.," n.d.). Further, both systems are open-source permitting users to easily and inexpensively tinker to exactly suit their needs.

Overall, we have shown that semi-automated sample processing can be performed affordably and quickly within individual academic labs. Shifting to semi-automated sample processing alleviates a significant bottleneck in discovery-focused work and allows for seamless transitions from sample collection to data generation and analysis. With this robust and scalable solution, we can test hypotheses faster and spend less time on repetitive but important tasks like sample processing.

Methods

Sample collection

All sample collection protocols were reviewed and approved by the appropriate site-specific Institutional Review Board. Written consent was obtained for all participants. Only participants above 18 years of age were eligible. Blood samples were collected into EDTA-coated Vacutainer

tubes and processed within 8 hours of sample collection. Until processing the samples were stored at 4 degrees Celsius. Plasma was separated from blood using standard clinical blood centrifugation protocol.

Synthesized RNA controls

We synthesized a custom RNA oligonucleotide consistent with the 3' end of External RNA Controls Consortium 54 (ERCC54) using Integrated DNA Technologies (IDT). The RNA sequence for the synthesized ERCC54 oligonucleotide is listed below.

rUrUrU rArGrA rArUrG rCrUrU rArArA rGrArU rGrGrC rArGrA rGrUrU rGrGrA rGrGrA rGrArG rArUrU rUrGrC rCrArA rUrCrA rCrArA rArCrC rArArA rUrCrA rGrUrU rGrArG rUrGrG rArGrG rGrCrA rArCrA rArCrA rGrArG rArGrU rUrGrC rUrArU rArGrC rGrArG rGrGrC rUrUrU rGrGrC rArArA rCrArA rCrCrC rArCrC

System setup
We set up our OT1 robot per vendor instructions and subsequently mounted an eight-channel 1.2 mL Rainin pipette (Cat No 17014496). To attach the pipette, we designed and 3D printed several supplementary holds using a uPrint 3D printer. In total, we mounted four clasps – two U-shaped along the neck and two L-shaped along the base. We also manufactured and mounted two parts for the OT1 system itself as well as a tip box holder for the non-standard Rainin tip footprint. All mounts are pictured in Fig 1 and specific 3D printing STL files can be found on Github at the following address.
https://github.com/miramou/cfRNA_pipeline_automation/tree/master/Opentrons/custom_solidworks_parts

Manual sample processing
We extracted RNA from either 1 mL of plasma or 1 mL of nuclease-free H2O spiked with a known concentration of ERCC54. RNA was isolated using using Norgen's Plasma/Serum Circulating and Exosomal RNA Purification Kit (Slurry Format) (Cat No 42800) and subsequently treated with Lucigen Baseline-ZERO DNAse (Cat No DB0715K). It was then cleaned and concentrated into 12 µL using Zymo's RNA Clean and Concentrator-5 kit (Cat No R1013).

Following cfRNA extraction from plasma samples, isolated RNA concentrations were estimated using Agilent's Bioanalyzer RNA 6000 Pico Kit (Cat No 5067-1513) per manufacturer instructions.

Semi-automated sample processing
We extracted RNA from either 1 mL of plasma or 1 mL of nuclease-free H2O spiked with a known concentration of ERCC54. RNA was isolated using using Norgen's Plasma/Serum Circulating and Exosomal RNA Purification 96-Well Kit (Slurry Format) (Cat No 29500) and subsequently treated with Lucigen Baseline-ZERO DNAse (Cat No DB0715K). It was then cleaned and concentrated into 12 µL using Zymo's RNA Clean and Concentrator-96 kit (Cat No R1080).

Following cfRNA extraction from plasma samples, isolated RNA concentrations were estimated using Agilent's Bioanalyzer RNA 6000 Pico Kit (Cat No 5067-1513) per manufacturer instructions.

All code used for system operation can be found on Github at the following address.
https://github.com/miramou/cfRNA_pipeline_automation/tree/master/Opentrons

RT-qPCR validation

Samples that contained a known concentration of ERCC54 were prepared for RT-qPCR using Biorad's iTaq Universal Probes One-Step Kit (Cat No 1725141). Specifically, we combined 1 µL of isolated RNA with 0.5 µL of the corresponding Taqman probes (Thermo Fisher Cat No 4448490, Assay ID Ac03459999_a1) and 8.5 µL of master mix following manufacturer instructions. We then performed real time RT-qPCR on the Biorad CFX system (either 96 or 384) using the following program:

(1) 50°C for 10 minutes
(2) 95°C for 1 minute
(3) 95°C for 10 seconds
(4) 60°C for 20 seconds
(5) Return to step 3 for a total of 40 cycles

Finally, we obtained cycle thresholds from the corresponding CFX software.

Library preparation and sequencing

cfRNA sequencing libraries were prepared with Takara's SMARTer Stranded Total RNAseq Kit v2 - Pico Input Mammalian Components (Cat No 634419) from 2-8 µL of eluted cfRNA according to the manufacturer's instructions. Samples were barcoded using Takara's SMARTer RNA Unique Dual Index Kit – 96U Set A (Cat No 634452). Samples were then pooled in an equimolar fashion and sequenced on Illumina's NovaSeq platform (2x75 bp) to an average depth of 50 million reads per sample.

Read mapping and quality metrics

Raw sequencing reads were trimmed with trimmomatic (Bolger, Lohse, & Usadel, 2014) and then aligned to the human reference genome (hg38) with STAR (Dobin et al., 2013). Mapping quality statistics were aggregated using MultiQC (Ewels, Magnusson, Lundin, & Käller, 2016).

To estimate degradation, we first counted the number of reads per exon and annotated each exon with its corresponding gene and exon number using htseq-count (Anders, Pyl, & Huber, 2015). We then counted the number of genes for which all reads mapped exclusively to the 3' most exon per sample and divided by the total number of genes detected to obtain the fraction of genes where all reads mapped to the 3' most exon.

Finally, we estimated ribosomal read fraction by counting the number of reads that mapped to the ribosomal region (GL000220.1:105424-118780) using samtools view (Li et al., 2009).

Further computational analyses

All further computational analysis was performed using Python 3.7. The specific environment requirements and notebook used can be found on Github at the following address. https://github.com/miramou/cfRNA_pipeline_automation/tree/master/Manuscript_Figs

Acknowledgements
We are deeply grateful to both Norma Neff and Rene Sit for their sequencing expertise. We also thank Brian Yu for helping brainstorm and providing indispensable advice around how to best use the Opentrons system. Figures 1A and 2A were created with BioRender.com.

Fig. 1 **Overview of semi-automated RNA isolation system.** (**A**) Graphical overview of sample collection, processing, data generation and analysis of cfRNA data. Step 2 compares manual and semi-automated sample processing. (**B**) Representative photograph of full system. (**C**) Close-up of mounted eight-channel high volume pipette. Dark blue, green, and light blue indicators highlight 3D printed fixtures to prevent motion in the x, y, and z directions respectively. Black arrow indicates moving fixture used to eject tips. (**D**) Representative freeze-frame photographs of various tasks performed on the robotic system.

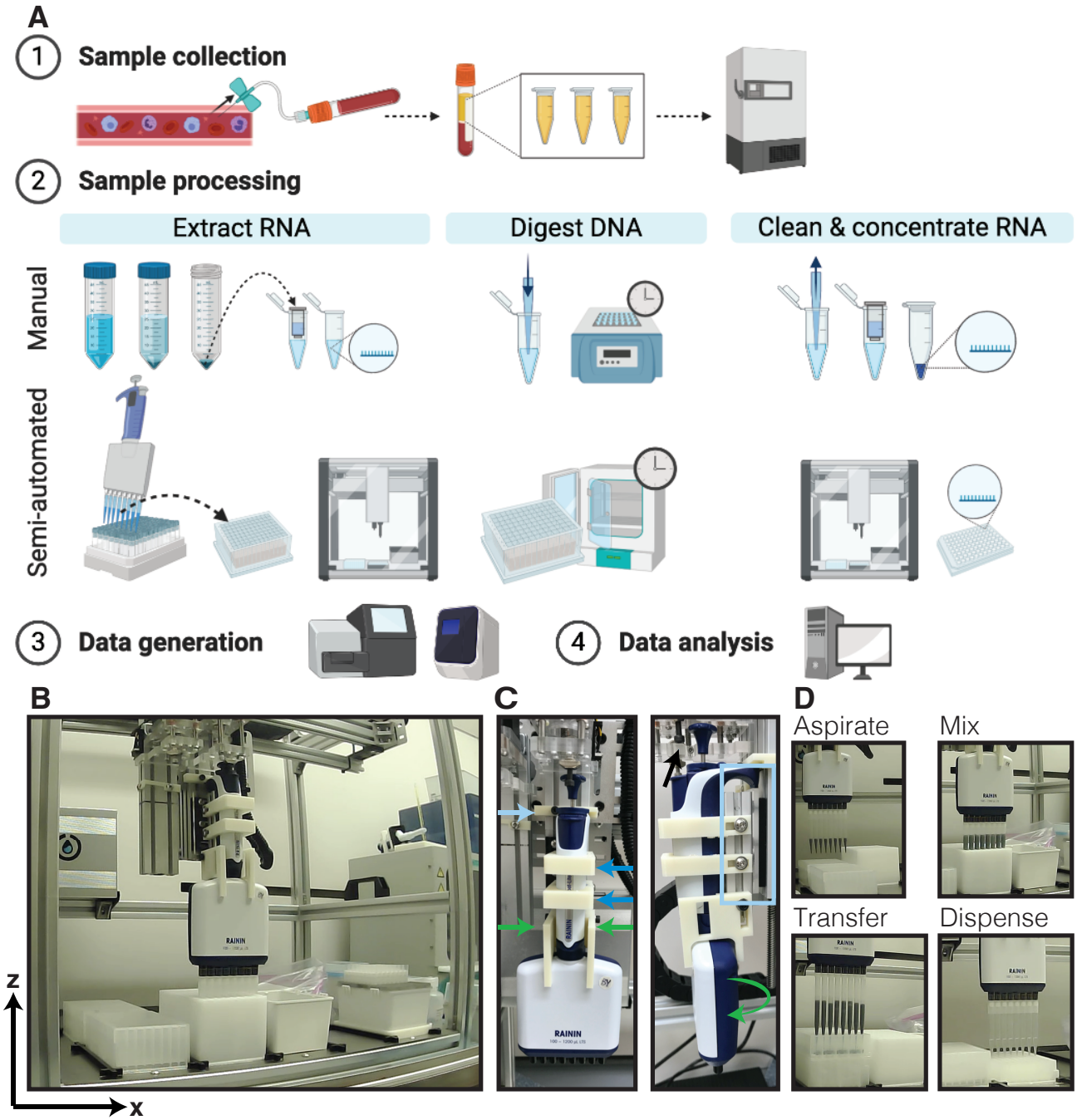


Fig 2 **Semi-automated RNA isolation system validation using RNA control.** (**A**) Graphical overview of semi-automated sample processing. (**B**) Comparison of RT-qPCR cycle thresholds (C_t) for paired samples pre- and post-refrigeration (light and dark blue) as compared to unrefrigerated samples (green) for two independent trials (rows). Histograms (right column) further highlight the distribution of difference in C_t between paired samples post- and pre-refrigeration. (**C**) Boxplot to identify any cross-contamination during processing. Yellow and orange boxplots show the distribution of C_t for processed water samples from two independent trials as compared to those for known non-template controls (NTC) in green. Red points indicate positive controls. (**D**) Comparison of manual and semi-automated protocols with three trials per method. Top and bottom violin plot compare all trials in either an independent fashion or grouped together respectively. Symbols * and ** indicate statistically significant differences between the two indicated groups at thresholds less than of 0.05 and 10^{-10} respectively.

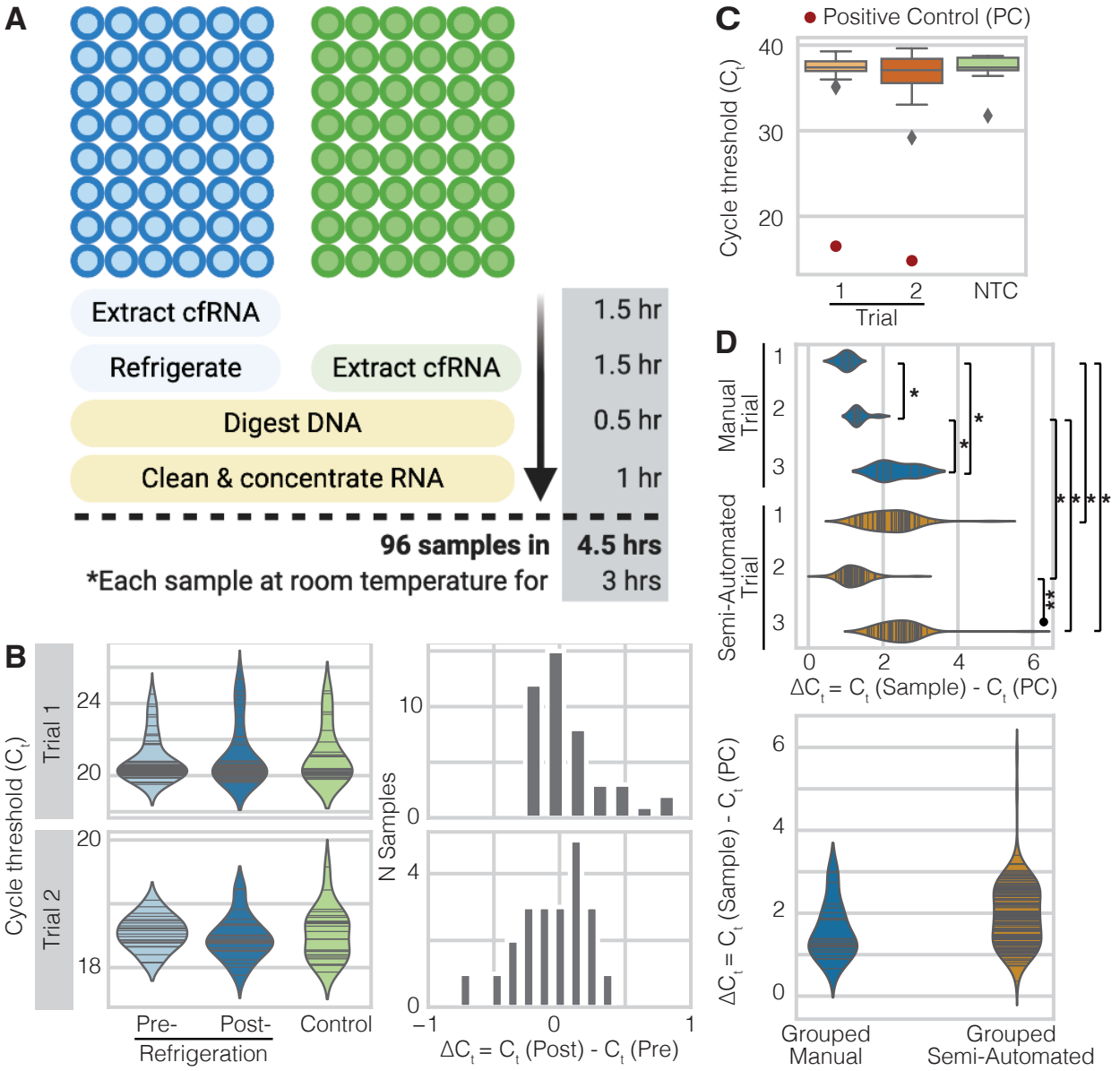


Fig 3 **Semi-automated RNA isolation system validation using cfRNA samples.** (**A**) Comparison of time spent (hours) vs samples processed for manual vs semi-automated processes. Total number of hours and samples shown for each trial (left) vs number of hours spent per sample (right). (**B**) Violin plot comparing total cfRNA plasma concentration measured post-extraction for samples processed manually vs semi-automatedly. (**C**) Left to right – Boxplots comparing three metrics of RNA quality: Fraction of genes where all reads mapped to the 3' exon (Left), total reads that uniquely mapped (Middle), and total reads that mapped to the ribosome (Right). Symbol * indicates statistically significant differences between the two indicated groups at thresholds less than of 2 x 10^{-4}.

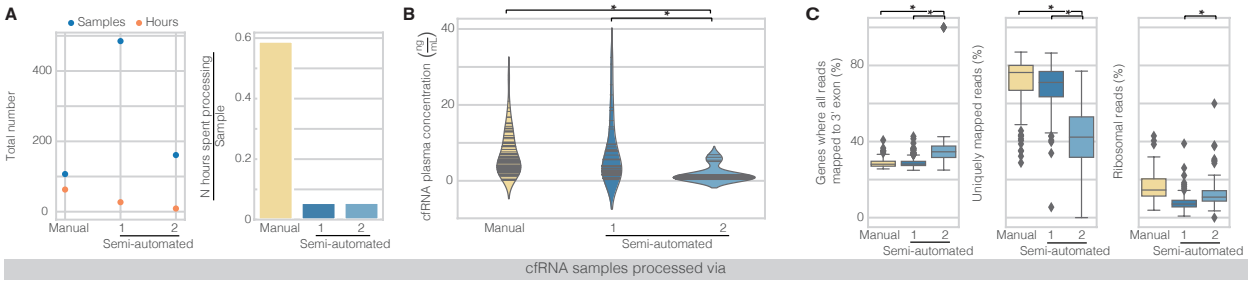


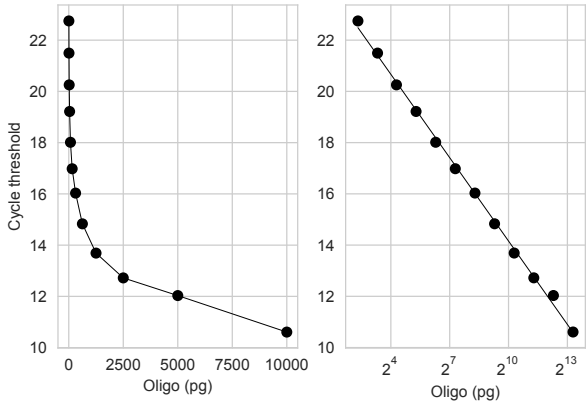
Table 1 **Time and cost comparison for manual and semi-automated RNA isolation protocols.**

Sample processing method	*Batch size*	*Time per batch (hours)*	*Cost per sample ($)*
Manual	8	4.5	41.25
Semi-automated	96	4.5	15.24

Supp Movie 1 **Full-protocol at 20x speed. Slides in white indicate user steps like centrifugation.**

See attached file or https://youtu.be/g6RsSaNvSNA.

Supp Figure 1 **Validation of custom RNA control, ERCC54, and its corresponding probes across a range of oligonucleotide input values.** Left and right plots show linear and log x-axis values respectively.



Bibliography

Aghaeepour, N., Ganio, E. A., Mcilwain, D., Tsai, A. S., Tingle, M., Van Gassen, S., … Gaudilliere, B. (2017). An immune clock of human pregnancy. *Science Immunology*, *2*(15). doi:10.1126/sciimmunol.aan2946

Aghaeepour, N., Lehallier, B., Baca, Q., Ganio, E. A., Wong, R. J., Ghaemi, M. S., … Angst, M. S. (2018). A proteomic clock of human pregnancy. *American Journal of Obstetrics and Gynecology*, *218*(3), 347.e1-347.e14. doi:10.1016/j.ajog.2017.12.208

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. doi:10.1093/bioinformatics/btu638

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. doi:10.1093/bioinformatics/btu170

Crowley, E., Di Nicolantonio, F., Loupakis, F., & Bardelli, A. (2013). Liquid biopsy: monitoring cancer-genetics in the blood. *Nature Reviews. Clinical Oncology*, *10*(8), 472–484. doi:10.1038/nrclinonc.2013.110

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. doi:10.1093/bioinformatics/bts635

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. doi:10.1093/bioinformatics/btw354

Fan, H. C., Blumenfeld, Y. J., Chitkara, U., Hudgins, L., & Quake, S. R. (2008). Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(42), 16266–16271. doi:10.1073/pnas.0808319105

Fan, H. C., Gu, W., Wang, J., Blumenfeld, Y. J., El-Sayed, Y. Y., & Quake, S. R. (2012). Non-invasive prenatal measurement of the fetal genome. *Nature*, *487*(7407), 320–324. doi:10.1038/nature11251

Ghaemi, M. S., DiGiulio, D. B., Contrepois, K., Callahan, B., Ngo, T. T. M., Lee-McMullen, B., … Aghaeepour, N. (2019). Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy. *Bioinformatics*, *35*(1), 95–103. doi:10.1093/bioinformatics/bty537

GitHub - Opentrons/otone_hardware: 3D models and Bill of Materials for the opensource OT.One lab automation platform. (n.d.). Retrieved July 30, 2020, from https://github.com/Opentrons/otone_hardware

Koh, W., Pan, W., Gawad, C., Fan, H. C., Kerchner, G. A., Wyss-Coray, T., … Quake, S. R. (2014). Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), 7361–7366. doi:10.1073/pnas.1405528111

Kowarsky, M., Camunas-Soler, J., Kertesz, M., De Vlaminck, I., Koh, W., Pan, W., … Quake, S. R. (2017). Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), 9623–9628. doi:10.1073/pnas.1707009114

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project

Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi:10.1093/bioinformatics/btp352

Liu, M. C., Oxnard, G. R., Klein, E. A., Swanton, C., Seiden, M. V., Cummings, S. R., … Berry, D. A. (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology*. doi:10.1016/j.annonc.2020.02.011

Lo, Y. M. D., Chan, K. C. A., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M. F., … Chiu, R. W. K. (2010). Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science Translational Medicine*, *2*(61), 61ra91. doi:10.1126/scitranslmed.3001720

Ngo, T. T. M., Moufarrej, M. N., Rasmussen, M.-L. H., Camunas-Soler, J., Pan, W., Okamoto, J., … Quake, S. R. (2018). Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science*, *360*(6393), 1133–1136. doi:10.1126/science.aar3819

Pan, W., Ngo, T. T. M., Camunas-Soler, J., Song, C.-X., Kowarsky, M., Blumenfeld, Y. J., … Quake, S. R. (2017). Simultaneously Monitoring Immune Response and Microbial Infections during Pregnancy through Plasma cfRNA Sequencing. *Clinical Chemistry*, *63*(11), 1695–1704. doi:10.1373/clinchem.2017.273888

Sharon, E., Shi, H., Kharbanda, S., Koh, W., Martin, L. R., Khush, K. K., … De Vlaminck, I. (2017). Quantification of transplant-derived circulating cell-free DNA in absence of a donor genotype. *PLoS Computational Biology*, *13*(8), e1005629. doi:10.1371/journal.pcbi.1005629

Snyder, T. M., Khush, K. K., Valantine, H. A., & Quake, S. R. (2011). Universal noninvasive detection of solid organ transplant rejection. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(15), 6229–6234. doi:10.1073/pnas.1013924108

Song, C.-X., Diao, J., Brunger, A. T., & Quake, S. R. (2016). Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(16), 4338–4343. doi:10.1073/pnas.1600223113

Song, C.-X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., … Quake, S. R. (2017). 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Research*, *27*(10), 1231–1242. doi:10.1038/cr.2017.106

Sorrentino, S. (2010). The eight human "canonical" ribonucleases: molecular diversity, catalytic properties, and special biological actions of the enzyme proteins. *FEBS Letters*, *584*(11), 2194–2200. doi:10.1016/j.febslet.2010.04.018

Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and protein extraction: the past and the present. *Journal of Biomedicine & Biotechnology*, *2009*, 574398. doi:10.1155/2009/574398

Tsang, J. C. H., Vong, J. S. L., Ji, L., Poon, L. C. Y., Jiang, P., Lui, K. O., … Lo, Y. M. D. (2017). Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(37), E7786–E7795. doi:10.1073/pnas.1710470114