1   **Title:** Miniaturization and optimization of 384-well compatible metagenomic sequencing library

2   preparation

3

4   **Authors:** Madeline Y Mayday[1][¶]; Lillian M Khan[2][¶]; Eric D Chow[2]; Matt S Zinter[1]; Joseph L

5   DeRisi[2,3]

6

7   **Affiliations:**

8   [1] UCSF School of Medicine, Benioff Children's Hospital, Department of Pediatrics, Division of

9   Critical Care, San Francisco, CA, USA

10   [2] UCSF School of Medicine, Department of Biochemistry & Biophysics, San Francisco, CA,

11   USA

12   [3] Chan Zuckerberg Biohub, San Francisco, CA, USA

13

14   ¶ These authors contributed equally to this work.

15

Abstract

Preparation of high-quality sequencing libraries is a costly and time-consuming component of metagenomic next generation sequencing (mNGS). While the overall cost of sequencing has dropped significantly over recent years, the reagents needed to prepare sequencing samples are likely to become the dominant expense in the process. Furthermore, libraries prepared by hand are subject to human variability and needless waste due to limitations of manual pipetting volumes. Reduction of reaction volumes, combined with sub-microliter automated dispensing of reagents without consumable pipette tips, has the potential to provide significant advantages. Here, we describe the integration of several instruments, including the Labcyte Echo 525 acoustic liquid handler and the iSeq and NovaSeq Illumina sequencing platforms, to miniaturize and automate mNGS library preparation, significantly reducing the cost and the time required to prepare samples. Through the use of External RNA Controls Consortium (ERCC) spike-in RNAs, we demonstrated the fidelity of the miniaturized preparation to be equivalent to full volume reactions. Furthermore, detection of viral and microbial species from cell culture and patient samples was also maintained in the miniaturized libraries. For 384-well mNGS library preparations, we achieved a savings of over 80% in materials and reagents alone, and reduced preparation time by 90% compared to manual approaches, without compromising quality or representation within the library.

Introduction

Metagenomic next-generation sequencing (mNGS) is becoming an increasingly useful tool in the field of biology and clinical medicine. Its applications are almost limitless – any nucleic acid can be turned into a library, amplified, and sequenced, making mNGS an appealing technology for labs and hospitals alike. As sequencers such as the Illumina NovaSeq increase throughput, hundreds to thousands of libraries can be sequenced in a single run. Although the per-base cost of sequencing has become less expensive over the last several decades, the cost and time associated with sample preparation remain disproportionately high [1,2].

Manual library preparation is tedious and is often the bottleneck for many sequencing projects. Numerous library preparation protocols have been adapted for automation through the use of various positive displacement tip-based liquid handler instruments, including the Beckman Coulter Biomek, Hamilton Star, Agilent Technologies Bravo, TTP LabTech Mosquito, and others [3-5]. Though these provide more hands-off time during the library preparation process, the overall cost can often exceed that of hand-prepared libraries due to the increased dead volume of reagents and the large number of expensive, sometimes proprietary tips required for liquid handlers. Furthermore, sub-microliter miniaturization is a challenge for the majority of positive displacement based liquid handlers.

Recently, acoustic liquid handlers with sub-microliter precision have become commercially available and have been used for a large range of applications, including RT-qPCR, mass spectrometry, drug discovery, and compound dosing assays [6-9]. To date, few end-to-end protocols for miniaturized mNGS preparation have been available. Given the cost and time

59 limitations of current library preparation techniques, we sought to adapt our mNGS library

60 preparation protocol into a high-throughput protocol by leveraging the small dispensing volumes

61 of the Echo 525. Here, we describe a detailed protocol that provides high-fidelity, miniaturized,

62 automated, cost- and time-efficient 384-well library preparation together with its quality and

63 performance metrics.

64

Methods and results

66 All next generation sequencing workflows begin with isolation of nucleic acid; however, such

67 protocols are highly dependent on the source of the sample and desired product, such as DNA,

68 RNA, mRNA, cell-free RNA, and so on. The protocol described herein is independent of nucleic

69 acid isolation methods, and for the purpose of this work, we assume an input of isolated total

70 RNA. To optimize miniaturization of our laboratory's current library preparation protocol, we

71 prepared libraries from varying concentrations of HeLa RNA using the New England Biolabs

72 Ultra II Library Prep Kit (E7770S/L). The miniaturized and automated workflow is presented in

73 **Fig 1** and described in detail below and at http://doi.org/10.17504/protocols.io.tcaeise.

74

75 **Fig 1. Overall workflow.**

76 **Legend:** Workflow of the miniaturized, automated library preparation protocol. The complete

77 protocol is available at: http://doi.org/10.17504/protocols.io.tcaeise.

78

Sample dehydration

80 To achieve miniaturized volumes, nucleic acid samples must first be dehydrated. Several

81 vacuum evaporators were tested to dehydrate input RNA: Thermo Savant AES2010, Thermo

82   Savant SC110A, Thermo Savant DNA110, and GeneVac EZ-2. 5uL of extracted RNA were

83   loaded into a 384-well PCR plate (E&K Cat No EK-75009) and were spun at low (no heat/room

84   temperature), medium (40ºC), or high (60ºC) temperatures until all wells were completely dry.

85   Room temperature drying was prohibitively slow; at medium and high temperatures, drying

86   times were comparable. At 40ºC, sample plates dried fastest with the GeneVac E-Z2 (25

87   minutes) and ranged from 35-50 minutes using other machines. Of note, variable drying times

88   were observed between brands of 384-well PCR plates. After drying RNA at 40ºC for 25

89   minutes, samples were rehydrated in 5uL water. Parallel capillary electrophoresis (Agilent

90   Bioanalyzer) RNA Integrity Number showed no significant difference between the resuspended

91   sample and the original sample (T-test p = 0.306, **Fig 2**), demonstrating that dehydration does not

92   compromise RNA quality within these parameters.

93

94   **Fig 2. Dehydrated RNA demonstrates preserved integrity.**

95   **Legend:** Bioanalyzer traces and RNA Integrity Numbers (RINs) of biological replicates of HeLa

96   RNA. (A) Before being dried in a vacuum evaporator. (B) After being dried for 30 minutes at

97   40ºC. (C) After being dried for 25 minutes at 65ºC. RINs indicate that RNA quality is not

98   compromised during the dehydration process.

99

100   # Intra-sample positive control ERCC spike-ins

101   Internal spike-in control nucleic acids are useful indicators of potential library preparation errors.

102   Furthermore, carefully designed spike-in controls, such as the External RNA Controls

103   Consortium (ERCC) collection [10-12], which consists of 92 variable-length archaeal templates

104   present at a pre-defined range of concentrations, may be used to establish the relationship

105   between read count and input RNA concentration. For this mNGS protocol, 25 picograms of

106   ERCC RNA (Thermo Fisher Scientific Cat. No 4456740) were added to each sample prior to

107   library preparation.

108

Library preparation

110   Our laboratory has previously described a mNGS library prep protocol for RNA using the New

111   England Biolabs Ultra II Library Prep Kit [13-15]. In this protocol, RNA was quantified using

112   QuBit, fragmented in a magnesium-based buffer at 94ºC, primed with random hexamers, and

113   reverse transcribed to form cDNA. Libraries were made Illumina-compatible by blunting ends

114   of cDNA and adding non-templated d-A tails. Loop adaptors were ligated and cleaved with

115   uracil-specific excision reagent (USER) enzyme before PCR enrichment was performed with

116   sample-specific octamer primers.

117

118   To miniaturize and adapt this procedure to the Echo 525, we tested several protocol

119   modifications involving reduced volume of reagents (**Table 1**) using variable HeLa RNA input

120   (0, 0.1, 0.5, 1, and 5 ng in 5uL of water). Reagents for each step, including ERCC spike-ins,

121   were prepared as a master mix and miniaturized volumes were dispensed using the Echo 525.

122   The ideal number of PCR cycles is input-dependent and should be optimized depending on

123   sample characteristics; for this miniaturized protocol, 19 PCR cycles were used to achieve

124   adequate amplification of very low input samples.

125

Miniaturization	Sample input (uL)	RNA Fragmentation (uL)	First Strand Synthesis (uL)	Second Strand Synthesis (uL)	End Prep (uL)	Adaptor Ligation (uL)	USER/ PCR (uL)

100% (full volume)	5	10	10	60	60	33.5	53
20%	1	2	2	12	12	6.7	15.9*
12.5%	0.625	1.25	1.25	7.5	7.5	4.2	15.9*
10%	0.5	1	1	6	6	3.3	15.9*

126   **Table 1. Library synthesis reactions at varying miniaturizations.**

127   **Legend:** Total volumes of master mixes for each step of the NEB Ultra II RNA Library Prep for

128   each experimental miniaturization: 100%, 20%, 12.5% and 10% of the regular hand-prepped

129   volume. *Maximum miniaturization was 30% volume in order to retain reaction efficacy.

130

131   Although all dispensing steps utilized acoustic liquid handling in this protocol, magnetic bead-

132   cleaning steps still required pipetting automation. To facilitate high-throughput bead-cleaning,

133   we programmed the Beckman Coulter Biomek NX[P] with a 384 multichannel head to perform a

134   simultaneous 384-well bead clean and size selection using AMPure XP beads (Beckman Coulter

135   Cat. No A63881). Bead volumes were chosen to select for a final library size of ≥200 base pairs

136   and to remove adaptor- and primer-dimers. An Alpaqua 384 Post Magnet Plate (SKU A001222),

137   adjusted with a 3D-printed plastic adaptor (available for download at

138   http://derisilab.ucsf.edu/index.php?3D=225), was used to minimize library elution volume. The

139   adaptor raises the PCR plate when placed on the magnet plate, thereby lowering the height of the

140   bead-DNA pellet on the side of each well and allowing for complete resuspension of beads in as

141   low as 6uL of eluent.

142

143   To check the quality of libraries prepared using the miniaturized protocol, we used a parallel

144   capillary electrophoresis assay to process up to 95 samples simultaneously (Advanced Analytical

145   Fragment Analyzer). Libraries were assessed for distribution of cDNA fragment size, estimated

7

146   molarity and concentration, and for the presence of primer- and adaptor-dimers. Library size

147   distributions were consistent between libraries prepared with both full and miniaturized volumes

148   (average fragment size = 471±37bp, 86±7% between 200-1000bp).

149

Sequencing results

151   Final HeLa RNA libraries were sequenced on the Illumina MiSeq to an average depth of 2.5

152   million paired-end reads. Resultant data were processed through a pipeline for pathogen

153   detection developed in our laboratory involving subsequent removal of duplicate reads, reads

154   with low quality, and reads aligning to phage [13,16,17]. Original FASTQ files are available at

155   BioProject Accession #PRJNA493096.

156

157   Analysis of reads aligning to ERCC transcripts showed strong correlation to their original molar

158   spike-in concentrations, which spanned six orders of magnitude, indicative of a successful library

159   preparation and reflected uniform PCR amplification across fragment size (R^2 = 0.93, **Fig 3A**).

160   The linear association between input RNA concentration and sequencing output results was used

161   to calculate the approximate amount of HeLa RNA present in the original sample by solving the

162   following equation: $\frac{ERCC\ mass\ (pg)}{Total\ mass\ (pg)} = \frac{ERCC\ reads}{Total\ reads}$. The mass of input RNA measured by

163   fluorometric quantification (Qubit HS RNA) correlated strongly with estimations using the

164   ERCC back-calculation method (R^2 = 0.995, **Fig 3B**) indicating that ERCC controls are an

165   effective way to assess the mass of input nucleic acid, even when present at sub-nanogram

166   levels.

167

8

168 **Fig 3. ERCC reads to determine library preparation quality and back-calculate RNA**

169 **input mass**.

170 **Legend:** (A) After normalizing to RNA input mass, reads aligning to the 92 ERCC spike-in

171 transcripts correlate linearly with ERCC spike-in concentration across six orders of magnitude in

172 all libraries prepared with the miniaturized protocol ($R^2 = 0.932$). (B) The initial sample input

173 mass can be calculated using the ratio of ERCC reads to total sequencing reads in each sample

174 using the equation: $\frac{ERCC\ mass\ (pg)}{Total\ mass\ (pg)} = \frac{ERCC\ reads}{Total\ reads}$ Back-calculated masses of HeLa libraries

175 correlated strongly with QuBit quantification ($R^2 = 0.995$).

176

177 The integrated portion of the HPV18 genome that is inherent to HeLa cells was used as a proxy

178 for a human infection as would be expected in metagenomic sequencing of patient samples

179 [18,19]. Libraries were analyzed for percent and depth of coverage of the reference sequence of

180 integrated HPV18, as well as the human transcriptome. At 1ng and 5ng of input, HPV18

181 coverage was essentially indistinguishable for reactions performed with full or miniaturized

182 volumes of reagents, indicating that using as low as 10% of the standard reagent volume

183 provided equivalent detection of the viral genome (**Table 2**). Assessment of human

184 transcriptome coverage in the HeLa libraries demonstrated high correlation between full volume

185 and miniaturized reactions (for 5ng RNA input: Spearman's $\rho = 0.79$, $p<0.001$; for 1ng RNA

186 input: Spearman's $\rho = 0.69$, $p<0.001$, **Fig 4**). Additionally, the miniaturized protocol

187 demonstrated strong correlation between replicate 1ng HeLa RNA samples (Spearman's $\rho = $

188 0.84, $p<0.001$). Together, these metrics show that libraries prepared using the miniaturized

189 protocol are equally sensitive for detection of both pathogen and human reads.

190

9

191   **Table 2) HPV coverage in HeLa samples**

HeLa RNA	Full (F) Miniaturized (M)	Unique paired-end reads	Average depth	Integrated HPV Coverage
5ng	F	1,609,470	23.6±10.9	99.6%
1ng	F	930,857	8.0±5.0	91.5%
5ng	M	2,770,339	23.0±15.9	98.9%
1ng (rep)	M	2,597,928	17.5±9.7	100%
1ng (rep)	M	31,114,923	59.4±34.8	100%
0.5ng	M	2,427,233	14.3±10.0	99.5%
0.1ng	M	1,642,888	7.6±5.0	96.9%

192   **Legend:** Coverage of the HPV genome was similar for libraries prepped with both the regular

193   and miniaturized protocols with various RNA input concentrations. This indicates that pathogen

194   detection is maintained despite the reduction in reagent volume. Miniaturized libraries were

195   prepared using 10% of the original volumes, with the exception of PCR, as described above.

196

197   **Fig 4. HeLa transcriptome coverage is comparable in full volume and miniaturized volume**

198   **preparations.**

199   **Legend:** Rank-rank plots of the human transcriptome show strong correlation between the full-

200   volume hand prepared protocol and the miniaturized, automated protocol for both (A) 1ng and

201   (B) 5ng of HeLa RNA input (5ng RNA input: Spearman's $\rho = 0.79$, p<0.001; 1ng RNA input:

202   Spearman's $\rho = 0.69$, p<0.001).

203

204   # Improving high-throughput library pooling

1

205   With advances in sequencing technology, the number of reads provided by a single run on high-

206   throughput sequencers such as the HiSeq or NovaSeq is driving the creation of larger sample

207   pools to take advantage of lower sequencing costs. Samples can be pooled to occupy equal

208   portions of a flow cell lane by capillary electrophoresis, fluorimetry, or qPCR. These processes

209   are costly, tedious, and error-prone due to imprecise estimations and inaccurate pipetting,

210   especially when pooling large numbers of libraries together.

211

212   To overcome these difficulties, we employed a two-step process to optimize the pooling of

213   hundreds of samples (**Fig 5**). First, we used the Echo to dispense *equal volumes* (500nL) from

214   each sample of a set of 265 libraries. This pool was then sequenced on the Illumina iSeq to a

215   total combined depth of 4.5 million reads. Use of the Echo to dispense nanoliter volumes

216   allowed preservation of the bulk of each original sample library. The read distribution of the 265

217   libraries resulted in a normal distribution with each library occupying a mean of 0.377% ±

218   0.125% of the total reads per sample (**Fig 5**).

219

220   **Fig 5. Pooling by iSeq correlates with pooling by NovaSeq.**

221   **Legend:** 265 libraries were pooled at equal volumes (0.5uL per library) using the Echo 525 and

222   sequenced on the Illumina iSeq to a total combined depth of approximately 4.5 million reads

223   (standard deviation indicated by green dashed lines). Resultant reads were assigned to each

224   barcoded library to calculate the percent of total reads occupied by each sample. These ratios

225   were used to re-pool each library in appropriate volumes (ranging from 160nL to 3800nL) using

226   the Echo 525 to achieve equal read representation across all 265 samples. Libraries pooled at

227   equal volumes and sequenced on the iSeq occupied a mean of 0.378% ± 0.125% of total

1

228   sequencing reads (standard deviation indicated by purple dashed lines). Re-pooled based on iSeq

229   ratios, libraries sequenced on the Illumina NovaSeq demonstrated a significantly tighter standard

230   deviation, with each library occupying a mean of 0.376% ± 0.087% of total sequencing reads.

231

232   By virtue of the internal ERCC controls, the number of reads belonging to each library was

233   proportional to the partial concentration occupied by each library. This enabled estimation of the

234   partial concentrations for each of the original 265 libraries. Next, the Echo 525 was used to

235   dispense calculated *equimolar* volumes of each library (ranging from 160nL to 3800nL) into a

236   final pool. The pool was sequenced on the Illumina NovaSeq to a total combined depth of

237   approximately 11 billion paired-end reads. Sequencing results from the NovaSeq yielded a mean

238   library proportion of 0.376% ± 0.087% of the total reads per sample (in this case, 41.7 million ±

239   9.6 million paired-end reads per sample). The significantly tighter standard deviation produced

240   by this step and shown in **Figure 5** demonstrate that hundreds of libraries can be pooled quickly

241   and within close range of each other using this method. As expected, samples with low read

242   counts on the original iSeq calibration run possessed the greatest variability when pooled for

243   sequencing on the NovaSeq.

244

Cost and time comparison

246   This miniaturized, high-throughput protocol significantly reduces the cost and time associated

247   with library prep. The materials cost for each library preparation using the manual protocol,

248   including consumables such as reagents and tips, was approximately $43 (~$16,648 for 384

249   samples). Using this miniaturized protocol, the cost per sample dropped to approximately $8 per

250   sample (~$3,161 for 384 samples), resulting in cost savings of over 80% (**Table 3A**). Similarly,

251   the automation of this library prep resulted in significant time savings. To complete 384 samples

252   by hand with the manual protocol would have consumed an estimated 166 hours (assuming 16

253   samples are prepared at a time); whereas the time to complete the same number using the

254   automated miniaturized protocol was approximately 10 hours (**Table 3B).** This significant time

255   savings further increases the cost savings of this protocol, as the cost estimates above do not take

256   labor into account.

257

258   **Table 3A) Cost comparison of full-volume and miniaturized protocols for 384 samples**

	Full volume protocol	Miniaturized protocol
Reagents	$13,427	$1824
Ampure beads	$2024	$282
Tips	$1085	$1020
Plates/tubes	$112	$14
Source plate	N/A	$21
TOTAL	**$16,648**	**$3,161**

259   **Legend:** Approximate cost comparison between the full volume protocol (done by hand) and the

260   miniaturized, automated protocol (performed using the Gene-Vac EZ-2, Labcyte Echo 525, and

261   the Beckman Coulter Biomek NX[P]). The miniaturized, automated protocol costs approximately

262   19% of the regular full volume hand prep. When accounting for employee salary for the time

263   required to complete 384 libraries, the cost of the miniaturized protocol drops significantly.

264

265   **Table 3B) Comparison of time at the bench for full-volume and miniaturized protocols for**

266   **384 samples**

	Full volume protocol	Miniaturized, automated protocol
Calculations/planning	120 minutes	10 minutes
Aliquoting RNA	120 minutes	120 minutes

1

Drying RNA	N/A	25 minutes
Master mix preparation	120 minutes	10 minutes
Dispensing master mixes	215 minutes	45 minutes
Incubations	5808 minutes	242 minutes
Bead cleans	2880 minutes	120 minutes
Quality check	Using the fragment analyzer and iSeq, QC time is equivalent for 384 samples	
Pooling	720 minutes	20 minutes
TOTAL	**9,883 minutes (~166 hours)**	**592 minutes (~10 hours)**

267   **Legend:** Approximate bench-time comparison between the full volume protocol (done 16 at a

268   time by hand) and the miniaturized, automated protocol (performed using the Gene-Vac EZ-2,

269   Labcyte Echo 525, and the Beckman Coulter Biomek NX^P).   The miniaturized, automated

270   protocol can complete 384 libraries in approximately 6% of the time it would take to complete

271   the same number by hand using the regular protocol.

272

Use-case example: mNGS for pathogen detection

274   As a demonstration of the applicability of this protocol, we used mNGS for pathogen detection in

275   clinical samples. Sequencing libraries were prepared, as described above, from RNA isolated

276   from two clinical samples. **Patient #1:** Endotracheal tube aspirate from a patient with

277   respiratory failure was collected with IRB approval and consent. An aliquot was sent to the

278   UCSF hospital clinical lab and another aliquot was placed immediately on dry ice. Standard

279   bacterial culture of this sample produced heavy *Enterobacter cloacae* growth.   Sequencing

280   libraries produced by hand with full reagent volume, and again as described in the miniaturized

281   protocol above, both demonstrated robust detection of *E. cloacae* RNA (109 unique rpm vs 103

282   unique rpm, respectively).   **Patient #2:** bronchoalveolar lavage sample was collected from

283   another patient with respiratory failure, with IRB approval and consent. Standard bacterial

284   culture of this sample produced heavy *Haemophilus influenzae* growth.   Sequencing libraries

1

285   produced concurrently by hand with full reagent volume, and as described in the miniaturized

286   protocol above, demonstrated comparable detection of *H. influenzae* (160 unique rpm vs 126

287   unique rpm, respectively) in both preparations. Human-stripped FASTQ files are available at

288   BioProject Accession #PRJNA493096.

289

290   **Discussion**

291   In this study, we present a miniaturized, highly automated, high-throughput protocol for

292   preparation of high-quality mNGS sequencing libraries. Compared to the standard full-volume

293   protocol, this protocol is faster, less expensive, produces data of equivalent quality for pathogen

294   detection and human transcriptome coverage in HeLa preparations, and, for use-case patient

295   samples, demonstrates correlation with clinical microbiology test results. To our knowledge, this

296   is the first application of the Echo for metagenomic and metatranscriptomic sequencing analysis

297   and adds to existing Echo-based protocols for DNA synthesis and plasmid sequencing [20,21].

298

299   This protocol has several advantages over alternative library preparation approaches. First, since

300   the Labcyte Echo does not use pipette tips, tip contamination is eliminated as is the pipetting

301   error inherent to transferring small volumes by volumetric displacement [22].   Second,

302   simultaneous processing of up to 384 samples significantly reduces variance between library

303   preps and exposes all samples to the same environmental and reagent contaminants, minimizing

304   batch effects. Third, use of ERCC spike-in controls allows the determination of the original

305   RNA input quantity, which is highly variable in human patient samples, and is often difficult to

306   measure by traditional spectroscopy. The ERRC controls also help establish the degree of

307   linearity between read counts and input concentrations.   The latter is particularly important to

1

308   avoid calling false negatives in clinical metagenomics. Lastly, the robots used in this

309   miniaturized and automated library prep protocol enable rapid and simultaneous processing of a

310   up to 384 samples at a time. This results in significant cost and time savings and can allow

311   large-scale projects to be completed at an expedited timeline.

312

313   There are several limitations to this protocol. First, although this protocol reduces the time

314   needed to prepare libraries for sequencing, it does not take into account the necessary upstream

315   procedures to isolate nucleic acid [23]. Second, Labcyte Echo source plates have a limited

316   working volume which results in dead volume loss of reagents in each source well; however,

317   because the overall volume of each master mix is greatly reduced by the miniaturization, this

318   dead volume does not greatly affect the overall cost of the prep. Third, the bead-based clean-up

319   steps described here require aspiration with traditional tip based liquid handlers. Finally, the

320   capital expense cost of robots used in this protocol is significant. Access to a core lab facility

321   with these machines will greatly reduce the initial start-up cost of this protocol.

322

323   In conclusion, we present an automated, miniaturized, high-throughput protocol to prepare RNA

324   sequencing libraries using the NEB Ultra II RNA Library Prep Kit. With this workflow, it is

325   possible to prepare 384 high-quality sequencing libraries with just 10% of the regular reagent

326   volume, at less than 20% of the cost and in less than 10% of the time compared to the regular

327   hand-prep. The workflow presented here may support the further advancement of clinical

328   metagenomics as well as large scale sequencing projects.

329   # Acknowledgements

1

334

References

335

336   1. Callejas S, Alvarez R, Benguria A, Dopazo A. AG-NGS: a powerful and user-friendly

337   computing application for the semi-automated preparation of next-generation sequencing

338   libraries using open liquid handling platforms. BioTechniques. 2014;56: 28-35.

339   2. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing:

340   higher than you think! Genome Biol. 2011;12: 125-2011.

341   3. Farias-Hesson E, Erikson J, Atkins A, Shen P, Davis RW, Scharfe C, et al. Semi-automated

342   library preparation for high-throughput DNA sequencing platforms. J Biomed Biotechnol.

343   2010;2010: 617469.

344   4. Ware JL, Moran L, Lin CL, Slatko B. Implementation of automation in a small-scale DNA

345   sequencing core facility. J Biomol Tech. 2000;11: 151-154.

346   5. Mora-Castilla S, To C, Vaezeslami S, Morey R, Srinivasan S, Dumdie JN, et al.

347   Miniaturization Technologies for Efficient Single-Cell Library Preparation for Next-Generation

348   Sequencing. J Lab Autom. 2016;21: 557-567.

349   6. Roberts K, Callis R, Ikeda T, Paunovic A, Simpson C, Tang E, et al. Implementation and

350   Challenges of Direct Acoustic Dosing into Cell-Based Assays. J Lab Autom. 2016;21: 76-89.

351   7. Sinclair I, Stearns R, Pringle S, Wingfield J, Datwani S, Hall E, et al. Novel Acoustic Loading

352   of a Mass Spectrometer: Toward Next-Generation High-Throughput MS Screening. J Lab

353   Autom. 2016;21: 19-26.

1

354   8. Agrawal S, Cifelli S, Johnstone R, Pechter D, Barbey DA, Lin K, et al. Utilizing Low-Volume

355   Aqueous Acoustic Transfer with the Echo 525 to Enable Miniaturization of qRT-PCR Assay. J

356   Lab Autom. 2016;21: 57-63.

357   9. Yin X, Scalia A, Leroy L, Cuttitta CM, Polizzo GM, Ericson DL, et al. Hitting the target:

358   fragment screening with acoustic in situ co-crystallization of proteins plus fragment libraries on

359   pin-mounted data-collection micromeshes. Acta Crystallogr D Biol Crystallogr. 2014;70: 1177-

360   1189.

361   10. Lee H, Pine PS, McDaniel J, Salit M, Oliver B. External RNA Controls Consortium Beta

362   Version Update. J Genomics. 2016;4: 19-22.

363   11. Pine PS, Munro SA, Parsons JR, McDaniel J, Lucas AB, Lozach J, et al. Evaluation of the

364   External RNA Controls Consortium (ERCC) reference material using a modified Latin square

365   design. BMC Biotechnol. 2016;16: 54-016.

366   12. External RNA Controls Consortium. Proposed methods for testing and selecting the ERCC

367   external RNA controls. BMC Genomics. 2005;6: 150-2164.

368   13. Zinter MS, Dvorak CC, Mayday MY, Iwanaga K, Ly NP, McGarry ME, et al. Pulmonary

369   Metagenomic Sequencing Suggests Missed Infections in Immunocompromised Children. Clin

370   Infect Dis. 2018.

371   14. Doan T, Wilson MR, Crawford ED, Chow ED, Khan LM, Knopp KA, et al. Illuminating

372   uveitis: metagenomic deep sequencing identifies common and rare pathogens. Genome Med.

373   2016;8: 90-016.

374   15. Wilson MR, Suan D, Duggins A, Schubert RD, Khan LM, Sample HA, et al. A novel cause

375   of chronic viral meningoencephalitis: Cache Valley virus. Ann Neurol. 2017;82: 105-114.

376   16. Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, et al.

377   Chronic Meningitis Investigated via Metagenomic Next-Generation Sequencing. JAMA Neurol.

378   2018;75: 947-955.

379   17. Langelier C, Zinter MS, Kalantar K, Yanik GA, Christenson S, O'Donovan B, et al.

380   Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant

381   Patients. Am J Respir Crit Care Med. 2018;197: 524-528.

382   18. Badal S, Badal V, Calleja-Macias IE, Kalantari M, Chuang LS, Li BF, et al. The human

383   papillomavirus-18 genome is efficiently targeted by cellular DNA methylation. Virology.

384   2004;324: 483-492.

385   19. Schwarz E, Freese UK, Gissmann L, Mayer W, Roggenbuck B, Stremlau A, et al. Structure

386   and transcription of human papillomavirus sequences in cervical carcinoma cells. Nature.

387   1985;314: 111-114.

388   20. Shapland EB, Holmes V, Reeves CD, Sorokin E, Durot M, Platt D, et al. Low-Cost, High-

389   Throughput Sequencing of DNA Assemblies Using a Highly Multiplexed Nextera Process. ACS

390   Synth Biol. 2015;4: 860-866.

391   21. Kanigowska P, Shen Y, Zheng Y, Rosser S, Cai Y. Smart DNA Fabrication Using Sound

392   Waves: Applying Acoustic Dispensing Technologies to Synthetic Biology. J Lab Autom.
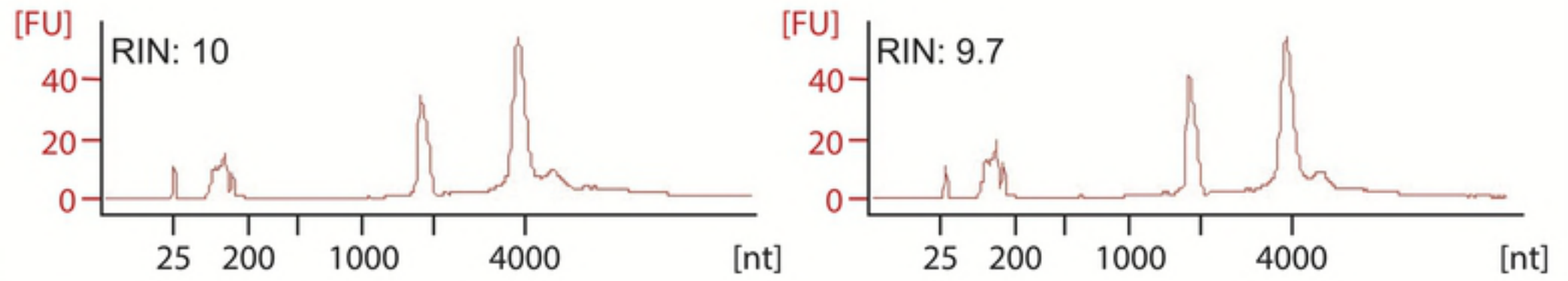
393   2016;21: 49-56.

394   22. Astle TW. Small Volume Pipetting. JALA: Journal of the Association for Laboratory

395   Automation. 1998;3: 62-64.

396   23. Schiebelhut LM, Abboud SS, Gomez Daglio LE, Swift HF, Dawson MN. A comparison of

397   DNA extraction methods for high-throughput DNA analyses. Mol Ecol Resour. 2017;17: 721-
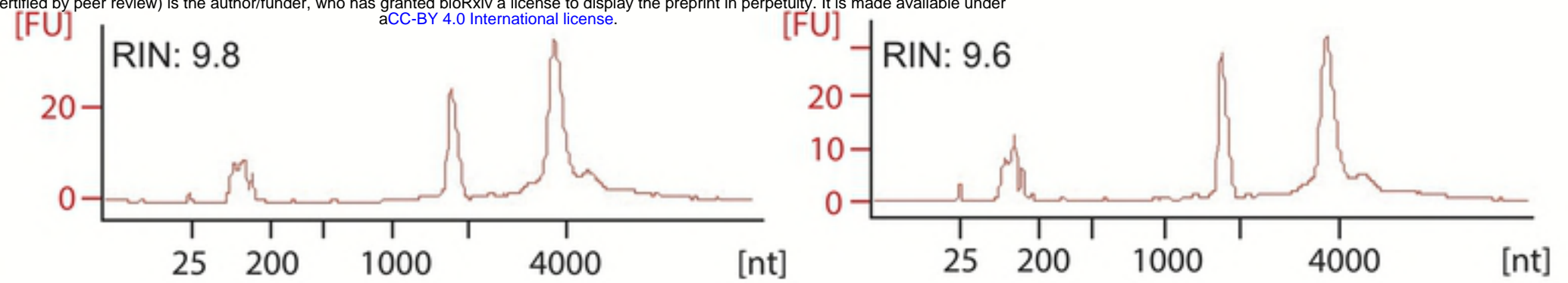
398   729.

399

| 25 mins | Dehydrate RNA |
| 10 mins | Prep master mixes |

Echo dispense master mix reagent

| 5 mins | Fragmentation and Priming Spike-In ERCCs |

12 mins   Incubate

Echo dispense master mix reagent

| 5 mins | First Strand Synthesis |

35 mins   Incubate

Echo dispense master mix reagent

| 10 mins | Second Strand Synthesis |

60 mins   Incubate

| 40 mins | Bead clean |

Echo dispense master mix reagent

| 5 mins | End Prep |

60 mins   Incubate

Echo dispense master mix reagent

| 10 mins | Adaptor Ligation |

15 mins   Incubate

| 40 mins | Bead clean |

Echo dispense master mix reagent

| 10 mins | USER/PCR |

60 mins   Incubate

| 40 mins | Bead clean |

| Varies | QC using BA/FA |

Equal volume pooling using the Echo

| 10 mins | Pooling Quantification |

Use iSeq to determine pooling ratios

| Varies | Pooling Quantification |

Equimolar pooling using the Echo

| 10 mins | Pooling |

Load onto NovaSeq

| Varies | Sequence libraries |

| | Library and ERCC Analysis |

Liquid Handlers and Sequencers

| | GeneVac E-Z2 | | Labcyte Echo 525 |
| | BC Biomek NXᴾ | | Illumina iSeq | | Illumina NovaSeq |

A | Undried HeLa RNA

RIN: 10

RIN: 9.7

B | HeLa RNA dried at medium heat

RIN: 9.8

RIN: 9.6

C | HeLa RNA dried at high heat

RIN: 9.5

RIN: 9.1

Finished and quantified sequencing libraries

Equal volume pooling of each library using Echo

Sequence library on iSeq

Pooling Quantification

Determine equimolar pooling using iSeq ratios

Equimolar pooling using Echo

Load onto NovaSeq

Final sequencing libraries

Labcyte Echo 525   iSeq (equal volume pooling)   NovaSeq (equimolar pooling)