Bar-coding bias in high-throughput multiplex sequencing of miRNA

Shahar Alon[1,*], Francois Vigneault[2,3,4*], Seda Eminaga[2], Danos C. Christodoulou[2], J.G. Seidman[2], George M Church[2,3] and Eli Eisenberg[5]

1 Department of Neurobiology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Israel
2 Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
3 Wyss Institute for Biologically Inspired Engineering, Boston, MA 02115, USA
4 Ragon Institute of MGH, MIT and Harvard, Massachusetts, MA 02129, USA
5 Raymond and Beverly Sackler School of Physics and Astronomy, Tel-Aviv University, Israel

* These authors contributed equally to this work.
Correspondence should be addressed to: Eli Eisenberg, elieis@post.tau.ac.il

SUPPLEMENTARY PROTOCOL

Ligation-Bias bar-coding for multiplex microRNA sequencing library protocol

DO NOT USE THIS PROTOCOL, use the No-Bias protocol instead (see below).

List of Oligonucleotides:
These oligos are used in the ligation-Bias bar-coding protocol. We recommend not using them.

name	Sequence (5'-3')
5BC_3'rApp-adapter	/5rApp/ATCTCGTATGCCGTCTTCTGCTTG/3ddC/
5BC_5'rApp-adapter-BC1	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rCrArUrG**rCrG
5BC_5'rApp-adapter-BC2	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rCrUrArG**rCrG
5BC_5'rApp-adapter-BC3	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rGrArUrC**rCrG
5BC_5'rApp-adapter-BC4	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rArCrUrG**rCrG
5BC_5'rApp-adapter-BC5	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rGrCrUrA**rCrG
5BC_5'rApp-adapter-BC6	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rGrUrArC**rCrG
5BC_5'rApp-adapter-BC7	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rArGrUrC**rCrG
5BC_5'rApp-adapter-BC8	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rGrUrCrA**rCrG
5BC_5'rApp-adapter-BC9	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rCrUrGrA**rCrG
5BC_5'rApp-adapter-BC10	rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC**rUrCrArG**rCrG
5BC_RT primer	CAAGCAGAAGACGGCATACGA
5BC_PCR1	AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA
5BC_PCR2	CAAGCAGAAGACGGCATACGA

All Oligo were ordered form IDT HPLC (or RNA-HPLC) purified. Bar-codes were designed so that base-balancing required for proper Illumina sequencing was taken into account.
To our knowledge, adenylation of the 3'adapter can only be ordered from Illumina, Bio Scientific or IDT, or can be enzymatically made from a standard phophorylated oligo as described previously (Vigneault, F. *et al. Nat Methods*, 2008).

Library Design Scheme:

```
miRNA sequencing library; 5' Ligation bar-coding

ligate 3' rApp-oligo
5' ------------------------------------------------------------ (miRNA) ATCTCGTATGCCGTCTTCTGCTTG 3'

ligate 5' RNA-oligo (bar-code introduction)
5' ----------------------GUUCAGAGUUCUACAGUCCGACGAUC[NNNN]CG (miRNA) ATCTCGTATGCCGTCTTCTGCTTG 3'

Reverse Transcription
5' ----------------------GUUCAGAGUUCUACAGUCCGACGAUC[NNNN]CG (miRNA) ATCTCGTATGCCGTCTTCTGCTTG 3'
                                                                    <---AGCATACGGCAGAAGACGAAC 5'

PCR Primer1

5' AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA------------- (miRNA) ------------------------ 3'
3' ------------------------------------------------------------ (miRNA) ------------------------ 5'

PCR Primer2
5' ------------------------------------------------------------ (miRNA) ------------------------ 3'
3' ------------------------------------------------------------ (miRNA) ---AGCATACGGCAGAAGACGAAC 5'

Final Library
5' AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGACGATC[NNNN]CG (miRNA) ATCTCGTATGCCGTCTTCTGCTTG 3'
3' TTACTATGCCGCTGGTGGCTGTCCAAGTCTCAAGATGTCAGGCTGCTAG[NNNN]GC (miRNA) TAGAGCATACGGCAGAAGACGAAC 5'

Sequencing single pass (Illumina standard small RNA primer)
                      CGACAGGTTCAGAGTTCTACAGTCCGACGATC--->
5' AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGACGATC[NNNN]CG (miRNA) ATCTCGTATGCCGTCTTCTGCTTG 3'
3' TTACTATGCCGCTGGTGGCTGTCCAAGTCTCAAGATGTCAGGCTGCTAG[NNNN]GC (miRNA) TAGAGCATACGGCAGAAGACGAAC 5'
```

Protocol Details:
All steps of the protocol for miRNA capture were conducted as described bellow, with the distinction of bar-coding during the 5'adapter ligation instead of during the PCR step.

Sample:
We used the following RNA sample in our experiments;

- Wild type mouse heart tissue RNA and cardiac disease mouse tissue RNA (extracted using the Ambion mirVana miRNA Isolation Kit AM1561)
- FirstChoice Human Brain Reference RNA (Ambion AM6050),

In both cases, each starting RNA sample was split into equal independent reactions to simulate different samples for bar-coding analysis.

No-Bias PCR bar-coding for multiplex microRNA sequencing library protocol

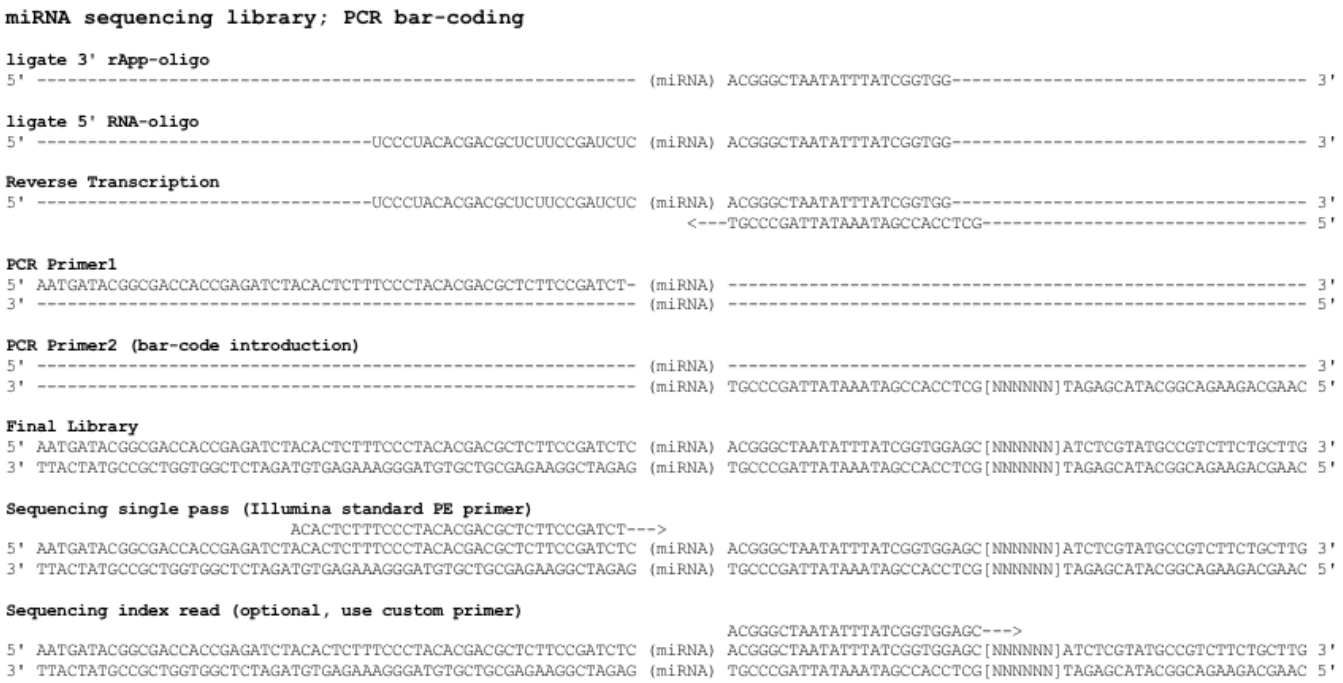
This library creates no significant bias, but necessitate that each sample is processed in parallel all the way to the end. Also this library can be sequenced either in a one pass of 75 bp, or using a custom indexing primer (so 36bp run + indexing), according to user preferences.

1. List of oligonucleotides

These oligos are used for the No-Bias PCR bar-coding protocol.
All oligos were ordered thought Integrated DNA Technologies (IDT; http://www.idtdna.com), with HPLC purification.

name	Sequence (5'-3')
BCPCR_3'rApp-adapter	/5rApp/ACGGG'CTAATATTTATCGGTGG/3SpC3/
BCPCR_5'RNA-adapter	rUrCrCrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrC
BCPCR_RT primer	GCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
BCPCR_PCR2-BC1	CAAGCAGAAGACGGCATACGAGATCGTGATGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC2	CAAGCAGAAGACGGCATACGAGATACATCGGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC3	CAAGCAGAAGACGGCATACGAGATGCCTAAGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC4	CAAGCAGAAGACGGCATACGAGATTGGTCAGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC5	CAAGCAGAAGACGGCATACGAGATCACTGTGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC6	CAAGCAGAAGACGGCATACGAGATATTGGCGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC7	CAAGCAGAAGACGGCATACGAGATGATCTGGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC8	CAAGCAGAAGACGGCATACGAGATTCAAGTGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC9	CAAGCAGAAGACGGCATACGAGATCTGATCGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC10	CAAGCAGAAGACGGCATACGAGATAAGCTAGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC11	CAAGCAGAAGACGGCATACGAGATGTAGCCGCTCCACCGATAAATATTAGCCCGT
BCPCR_PCR2-BC12	CAAGCAGAAGACGGCATACGAGATTACAAGGCTCCACCGATAAATATTAGCCCGT

2. Library design scheme



3. Particular Notes

Water: We use dH2O (such as Ambion AM9932 nuclease-free water) instead of DEPC, even when dealing with RNA since RNA is readily degraded when stored in DEPC water for longer time. But we are very careful about using fresh water aliquots and readily cleaning all surfaces and instruments with RNaseZap (Ambion AM9782).

Starting RNA: We recommend starting with at least 1 μg of good quality total RNA, but RNA enriched for small RNAs can also be used. We recommend verifying RNA quality using the Agilent Bioanalyzer RNA nano or pico chip.

Enzymes: Except for Superscript III from Invitrogen, we have found that using Enzymatics enzymes gave higher yield in all of our protocols (likely due to the purity of their product).

Incubations: All incubations are conducted on a thermal cycler, and when removed from the thermal cycler all samples are always carried on a PCR Iceless Cold Pack (Eppendorf 022510509).

Reaction volumes: here we detail the protocol reaction volumes for 1 sample and for a master mix of 15 samples (convenient for a 12 samples run). A " - " indicate that the reagent is independent to each sample and must be added uniquely to each reaction (such as RNA samples for example).

T4 RNA ligase: The use of T4 RNA ligase 2 truncated in conjunction with the 3' adenylated oligo is required for optimal microRNA capture by ligation since the reaction can be conducted in absence of ATP, therefore inhibiting the self-circularization of the microRNAs. Since the 3'adenylted oligo is blocked at its 3' end, the second ligation can be conducted in the presence of ATP so that the 5' RNA adapter can be efficiently ligated to the other extremity of the unknown microRNAs.

Illumina compatibility: These protocols and oligos are different but fully compatible with any Illumina sequencing instruments, and were designed for that specific purpose.

4. Ligate 3' adenylated adapter and 5' RNA adapter to microRNAs

4.1 Dilute the starting RNA to 200 ng/µl in dH2O if possible.

4.2 Set up the ligation reactions in a PCR tube as follow:

	1 rx	15 rx (add 5µl/tubes)	
• dH2O	0µl	0µl	
• 1µg total RNA in dH2O	5µl	-	
• 10x T4 RNA Ligase 2tr Buffer*	1µl	15µl	
• 10µM 3'rApp-adapter	1µl	15µl	
• 100% DMSO	1µl	15µl	

Heat at 90°C for 30 sec, then 4°C for 30sec

• RNAseInhibitor	0.5µl	7.5µl	(Enzymatics Y924L)
• T4 RNA ligase 2tr (200 U/µl)	1.5µl	22.5µl	(Enzymatics)

Incubate at 22°C for 1 hour

***Notes:** We recommend using Enzymatics buffer only, since its composition gave us significantly higher yield than other commercially available T4 RNA ligase 2 truncated buffers.*

4.3 Within 5 minutes remaining, prepare the 5' RNA adapter by incubating a small aliquot at 70°C for 2min, then 4°C for 30 sec.

4.4 Spin down and add the following reagents directly to the previous ligation mixture and mix well:

	1 rx	15 rx (add 3µl/tubes)	
• 10 mM ATP	1 µl	15µl	(Enzymatics N207-10-L)
• 10µM 5' RNA adapter	1 µl	15µl	
• T4 RNA ligase 1	1 µl	15µ	(Enzymatics L605L)

Incubate at 20°C for 1 hour

5. Reverse Transcription of the captured microRNAs

The previous steps result in a reaction volume of 13µl. Only 4µl is used in the subsequent RT-PCR step, and so the remaining can be stored (-80℃) as a backup (highly recommended) for two more runs, or processed in parallel to achieve a higher yield at the end of the protocol, post gel extraction.

5.1 Assemble the RT reaction as follow:

	1 rx	15 rx
• Ligated microRNAs	4µl	-
• 25 µM RT primer	1µl	-

Incubate at 65C°C for 10 min, than 4°C

5.2 Spin down and add the following on ice and mix well:

	1 rx	15 rx	
• 5x First strand buffer	2µl	30	
• 12.5mM dNTP mix	0.5µl	7.5	(Enzymatics N205L)
• RNAse Inhibitor	0.5µl	7.5	
• 100mM DTT	1µl	15	
• Superscript III	1µl	15	(Invitrogen 18080-044)

Incubate at 48°C for 30 min

6. Limited PCR

6.1 Assemble the PCR reaction as follow:

	1 rx	15 rx (39µl/tubes)	
• dH2O	27µl	405	
• Reverse Transcribed-miRNAs	10µl	-	
• 5x HF buffer	10µl	150	
• 25 µM PCR1	1µl	15	
• 25 µMPCR 2*	1µl	-	(this is the bar-coded primer)
• 25mM dNTP	0.5µl	7.5	
• Phusion hotstart DNA polymerase	0.5µl	7.5	(NEB F-540L)

**Notes: this is where you want to use a different bar-coded primer for each different sample.*

6.2 Thermal cycle as follow:

1- 98°C for 30 sec
2- 98°C for 10 seconds
3- 60°C for 20 seconds
4- 72°C for 20 seconds go to step 2, 11x*
5- 72°C for 5 min
6- 4°C pause

*****Notes:*** *the number of cycles can be varied according to the amount of microRNA present in the starting sample, but in our hand 12 cycles total generally resulted in the best yield for the number of cycles. One wants to also avoid too much cycles to limit the introduction of unnecessary PCR bias as with any other PCR experiments (so try to keep it under 15 cycles if possible). The use of realtime PCR is not recommended to monitor the yield since the adapter-adapter product (i.e. without a microRNA insert) will greatly overtake the qPCR signal and will be misleading.*

6.3 Purify using a MinElute PCR Purification column*(Qiagen 28004), ressuspend in 15µl of 10mM tris-HCl.

****Notes:*** *This step is optional but highly recommended for an easier subsequent gel extraction. Also these columns have a tendency to trap ethanol after the wash, so to avoid subsequent gel loading problem, after the wash, spin 1.5 min dry (instead of 1 min), then rotate the column so that the tip of the lid point in the opposite direction and spin for another 1.5min at max speed. Then transfer the column to a clean elution tube and let the column open for 3 min to air-dry prior adding the resuspension buffer.*

7. Denaturing PAGE separation of library

One of the key issues with microRNA libraries is the presence of the 5'adapter ligated directly to the 3'adapter without any microRNA inserts captured. This product is quite significant since these oligo are used in large excess to maximize microRNA capture. Since the microRNA represent only a small portion (~22bp) of the final library size (~135bp), the majority of the remaining sequence is common to any other library fragments and can therefore anneal during the PCR cycling as a mismatch for the microRNA captured region. Therefore, if one was to extract the library on a regular native TBE gel, fragments of 5' adapter – 3'adapter (without miRNA) will be carried by the proper full-length library fragment and therefore will act as undesirable contaminant during the subsequent sequencing. To bypass this issue, the sample can be run on a denaturing gel in order to efficiently remove such undesirable bands. If the amount is

too significant, one can also substitute the 10μM 3'and 5'oligo adapters for 1 μM versions (optimize PCR cycling condition accordingly).

7.1 Sample preparation

- Transfer the 15μl sample to a PCR tubes.
- Add 15μl of 2X Novex TBE-Urea Sample Buffer (Invitrogen LC6876) to each sample.
- Also prepare one tube of a 25 bp loading ladder and one of a 100 bp loading ladder (2μl of loading ladder +13μl of dH2O + 15μl of 2X Novex loading buffer)
- Denature at 95°C for 3 min, then incubate at 4°C (quickly mix and spin down prior loading, but keep the sample in the PCR Iceless Cold Pack as much as possible)

7.2 Gel preparation

- Pre-warm the TBE running buffer for 2 min in a microwave.
- Load each sample across two lanes* of a 10% TBE-Urea precast acrylamide gel (Invitrogen EC6875BOX).
- Load the 25bp ladder at one end and the 100bp ladder at the other end*.

***Notes:** Therefore one gel can be used to extract 4 different library samples total.*

- Run the gel at 15 watts* for 35 minutes

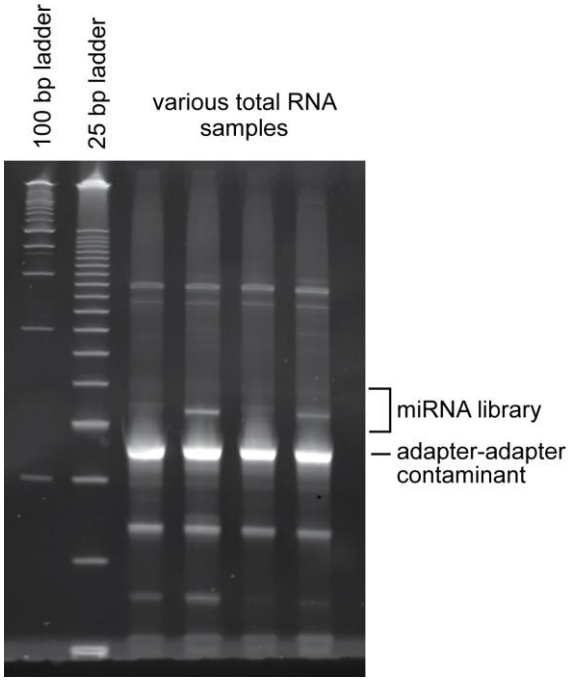
***Notes:** here we use constant watts instead of volts to try to maintain the high temperature of the gel (and therefore denaturation of the samples), if you run at 200 volts instead (as recommended by the manufacturer), the gel resolution will not be as good. If you are running two gels in the same box, run at 22 watts instead.*

7.3 Gel staining

- Pry apart the gel and incubate for 10 min in a small container with 200 ml of distilled water (just enough to cover depending on container size) and 10μl of GelRed stain* (Biotium 41002).

***Notes**: SybrGold can also be used (but GelRed seems to be a bit better). The results may vary greatly if using EtBr and the bands of interest might be impossible to see for the extraction.*

- Visualize the gel on a gel doc imager of your choice.
- The final library of microRNA should consist of one or two bands from 132 to 144 bp. The product of adapter-adapter ligation (without miRNA inserts) will be of 114 bp (you do not want this band)



***Notes:** we recommend taking a picture to help in identifying the bands, which will make subsequent visual gel extraction a lot easier. The bands of interest are just above the undesirable bright and thicker bands of adapter-to-adapter fragments. You can safely cut from 125 bp to 150 bp to capture the microRNAs.*

8. Gel extraction

8.1 Conduct gel extraction as follow;

- With a clean razor blade, cut the gel (ideally on a blue light illuminator, otherwise quickly on a UV illuminator) from 125bp to 150 bp (which should contain the double bands of small RNAs of interests). Be careful to not include the contaminating "no inserts library" bands of 114 bp.
- Place each gel band into a bottom-pierced 0.5ml tube, placed in a 1.5ml tube and centrifuge at maximum speed (~13 000 g) for 5 min (Ideally use one 1.5 ml tube for each 2-3 gel lanes maximum).
- Add 400µl of dH2O and 40µl of 3M NaOAc pH 5.5 to each tube, vortex briefly to mix.
- Incubate at 55°C for 60 min with frequent vortexing*

Notes: *We recommend using an Eppendorf thermomixer, since it seems to achieve a higher yield of elution by maintaining a constant vortexing action during the process. Also do not use non-stick and/or siliconized 1.5ml tubes for this incubation step, the extraction yield will be greatly affected, because of the heating.*

- Using a p1000 pipette tip cleanly cut by 4 mm at the end, transfer the liquid and gel slurry into nanosep columns (1 column per 400µl tube) (VWR 29300-642).
- Centrifuge for 5 min at max speed and transfer the eluate to a clean 1.5ml tube and discard the trapped acrylamide and the filter section of the column.

8.2 Ethanol precipitate

- Add 2µl of glycoblue* (Ambion AM9516) or glycogen (Ambion AM9510), and 1000µl of 100% ice cold EtOH (no need to re-add any salts for this precipitation). Inverse the sample 20 times to mix.
- Freeze at -80°C for 1hr (or -20°C o/n), centrifuge for 25 min at maximum speed, remove supernatant but leaving the pellet intact, wash with 500µl EtOH 70% (only if pellet is visible and strong), spin a max speed for 5 min and remove supernatant.
- Quick spin, and remove the last trace of EtOH, let to air dry open for 8 min.
- Ressuspend in 10µl each of 10mM tris-HCl or EB buffer.

***Notes:** If the samples are to be combined and sent to sequencing directly, the use of glycogen instead of glycoblue is preferable, because the dye in glycoblue will throw off any form of initial QC (Qubit, Nanodrop, Bioanalyzer). Nevertheless,* qPCR library quantitation against a phiX control will be fine.

9. Library QC and mixing

The objective is to mix each library to an equimolar concentration so that the final total read count of each sample post sequencing remains similar. To do so one can use real-time qPCR; we recommend, and have had great success using the KAPA Library Quant Kits (Kappa Biosystems KK4822) following the manufacturer protocol. Otherwise (and if on a budget) we have found that when the yield of the library preparation was good (i.e. you could easily see a band during the gel extraction step), then a Nanodrop reading seems to do the job.

9.1 Quantitate each sample on a Nanodrop spectrophotometer and mix an equimolar concentration of each sample into a single reaction tube.

9.2 Optional: concentrate using a Qiagen MinElute PCR Purification column* (with the modification to the protocol described above).

***Notes:** for this step only, add 10µl of 3M NaOAc and level to 150µl with TE to add some extra salt, which will improve DNA binding to the column at this point.*

10. **Optional; Gel diagnostic of library and/or second round of PAGE extraction;**

It is recommended to load an aliquot of the combined library onto another 10% TBE-Urea gel to insure proper removal of "no-insert" library (or to run a RNA pico chip on an Agilent Bioanalyzer). We have observed that sometime there is so much of this contaminating library compared to actual proper microRNA library, that when combining them into a single reaction, their level can remain quite significant. Thus, to maximize the cost spent on sequencing (i.e. to not waste sequencing reads on non-relevant library features), it is recommended to re-extract a second time the pooled samples as described previously (it seems like a lot of work, but it is insignificant compared to the time and cost of the downstream sequencing and data analysis, and definitively worth it in order to achieve the best quality of data).

10.1 Re-Extract the pooled samples together as described in step 7 and 8 with the following details;

- Load no more then 10µl of sample per lane (i.e. 20µ total if taking into account the 2x loading dye), split across a few lane as required. There is unlikely a need for more then 8 lanes since actual sequencing only required 10 nM of final library (We normally only load 4 lanes and extract across 2 tubes that we pooled back into one at the end).