Evaluation of Commercially Available RNA Amplification Kits at Subnanogram Input Amounts of Total RNA for RNA-seq

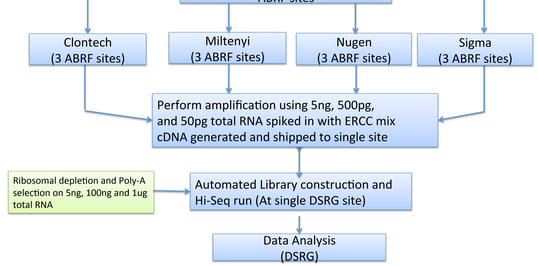
S. Shanker[†1]**, C. Nicolet**[†2]**, A. Paulson**[†3], Y. Alekseyev[1], N. Beckloff[2], D. Bintzler[†4], N. J. Bivens[3], R. R. Davis[4], R. J. Donnelly[5], H. J. Edenberg[6], A. F. Gillaspy[7], D. Grove[†5], N. Jafari[8], J. S. Kerley-Hamilton[9], K. Lashley[10], R. H. Lyons[†6], A. Peak[3], A. Perera[†3], J. Thimmapuram[†7], L. Wang[11], C. L. Wright[†8]

[†1]University of Florida, [†2]University of Southern California, [†3]Stowers Institute, [†4] DNA Analysis Inc. [†5]Penn State, [†6]University of Michigan, [†7]Purdue University, [†8]University of Illinois, [1]Boston University, [2]Case Western Reserve University, [3]University of Missouri, [4]UC Davis, [5]UMD in New Jersey, [6]Indiana University, [7]Oklahoma University Health Science Center, [8]Northwestern University, [9]Dartmouth, [10]National Cancer Institute, [11]University of Tennessee Health Science Center †Member of DNA Sequencing Research Group (DSRG) of ABRF

Abstract

NextGen sequencing is a powerful and cost-effective tool for ultra-high-throughput genome and transcriptome analysis. Multiple recent publications on RNA-Seq have demonstrated the power of next generation sequencing technologies in whole transcriptome analysis. It has greatly accelerated our understanding of both the quantitative and qualitative aspects of transcript biology in both prokaryotes and eukaryotes. The vendor specific protocols used for RNA library construction typically require at least 100 ng of total RNA. However, under certain conditions such as single cells, stem cells, difficult to isolate cell types, or fractionated cancer cells, only a small amount of material is available. In these cases, effective transcriptome profiling requires amplification of subnanogram amounts of RNA. Several RNA amplification kits are available for amplification prior to library construction and next generation sequencing but these kits have not been comprehensively field evaluated for accuracy and performance of RNA-Seq for picogram amounts of RNA.

Study Design



Materials

Experimental Kits

SMARTer™ Ultra Low input RNA-Clontech
- polyA based
- 10 pg to 10 ng input

SuperAmp™-Miltenyi Biotec
- polyA based
- 10 pg to 100 ng input

Ovation® RNA-Seq System V2-NuGEN Technologies
- No polyA requirement; using fragmented RNA not specifically addressed
- 500 pg to 100 ng input **NOTE: 50 pg data study point out of spec**

SeqPlex RNA-Sigma-Aldrich Corporation
- No polyA requirement, fragmented RNA OK as input
- 100 pg to 5 ng input **NOTE: 50 pg data study point out of spec**

Control Kits

Illumina RNA Sample Prep Kit V2 ("TS")
- polyA based
- 100 ng to 4 ug input **NOTE: 5 ng data study point out of spec**

Epicentre RiboZero Gold ("RZ")
- polyA based
- 100 ng to 1 ug input

Input RNAs

Human universal reference total RNA sample (Clontech)

ERCC control mix (Ambion/LifeTech)

Amplification Results from Clontech Kit

Sample Name	Total Mass (ng) as provided by participants	Method used for quantification	Total Mass (ng) based on Qubit Assay (DSRG)	Bioanalyzer Results (DSRG)	Difficulty level of the Amplification kit	Comments by Participants
P1C_5ng	18	Qubit	16.9	600bp-4,900bp with average fragment size of 1.79kb	1	Performed 17 cycles for 50pg, 14 cycles for 500pg and 12 cycles for 5ng input total RNA during amplification. Took less than a day
P1C_500pg	4.7	Qubit	-			
P1C_50pg	3.5	Qubit	-			
P2C_5ng	No information provided		-	Concentration of amplified products was too low to be detected		
P2C_500pg	No information provided		-			
P2C_50pg	No information provided		-			
P3C_5ng	19.65	Bioanalyzer	31.1	450bp-6,700bp with average fragment sizes of 1.79kb	6	Simple protocol, but not all steps can be practically carried out in a clean box. The labels on Advantage 2 PCR reagent tubes are not appropriate for freezer storage.
P3C_500pg	2.3	Bioanalyzer	-			
P3C_50pg	2.82	Bioanalyzer	-			

Amplification Results from Miltenyi Kit

Sample Name	Total Mass (ng) as provided by participants	Method used for quantification	Total Mass (ng) based on Qubit Assay (DSRG)	Bioanalyzer Results (DSRG)	Difficulty level of the Amplification kit	Comments by Participants
P1M_5ng	4,032	Qubit	2,650	>750bp with average fragment size of 1kb	2 for experienced biologist	Easy to perform protocol. Will be easier to use when starting directly from cells. The whole process took about nine and a half hours
P1M_500pg	3,996	Qubit	2,610			
P1M_50pg	4,236	Qubit	3,060			
P2M_5ng	11,300	Nanodrop	3,090	>800bp with average fragment size of 1.28kb	5	The amplification using uMAC magnetic stand is not quite suitable for multisampling or high-throughput handling. The entire process took about 1.5 days
P2M_500pg	11,700	Nanodrop	3,440			
P2M_50pg	11,500	Nanodrop	3,070			
P3M_5ng	No information provided		-	Too low to be detected	3	Change initial volume of incubation buffer from 6.5ul to 10ul so it is more manageable. Add more sealing solution during incubations (step 2,3 and 6) . Mention in the protocol that light color is green when at 37 degree and red at 42 degree
P3M_500pg	No information provided		-			
P3M_50pg	No information provided		-			

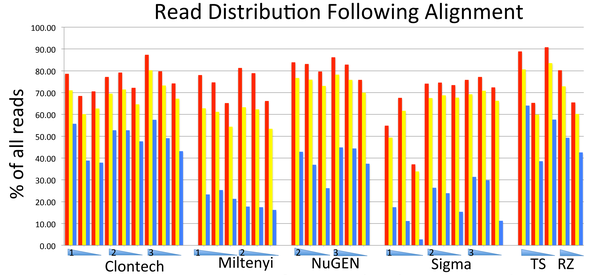
Amplification Results from Nugen Kit

Sample Name	Total Mass (ng) as provided by participants	Method used for quantification	Total Mass (ng) based on Qubit Assay (DSRG)	Bioanalyzer Results (DSRG)	Difficulty level of the Amplification kit	Comments by Participants
P1N_5ng	6611.6	Nanodrop	-	Too low to be detected		
P1N_500pg	4338.4	Nanodrop	-			
P1N_50pg	1941.6	Nanodrop	-			
P2N_5ng	7,820	Nanodrop	3,300	160bp-1,300bp with average fragment size of 575bp	1	It took about 5-6 hrs with very little hands on time. However, it is one full work day because there are not really good stopping points until the last day. The participant liked the Interactive Nugen quick guide to calculate the amounts of each reagent to use in the master mix at each step
P2N_500pg	5,960	Nanodrop	2,750			
P2N_50pg	2,746	Nanodrop	1,670			
P3N_5ng	4,990	Bioanalyzer	2,750	119bp-1250bp with average fragment size of 719	3	The kit is very useful for low input RNA. Took about 6-7 hours to complete the procedure
P3N_500pg	4,474	Bioanalyzer	1,950			
P3N_50pg	3,740	Bioanalyzer	1,100			

Amplification Results from Sigma Kit

Sample Name	Total Mass (ng) as provided by participants	Method used for quantification	Total Mass (ng) based on Qubit Assay (DSRG)	Bioanalyzer Results (DSRG)	Difficulty level of the Amplification kit	Comments by Participants
P1S_5ng	21,082	Nanodrop	3,550	Has primer peak (48bp) at very high conc. so the average fragment size decreased to 62-188bp	3	Some of the vendors instructions were not clear. It took two days to execute the protocol.
P1S_500pg	19,340	Nanodrop	4,910			
P1S_50pg	17,380	Nanodrop	1,100			
P2S_5ng	926	Nanodrop	815	Most fragments ranging from 47bp-1500bp with average fragment size of 200bp-340bp	3	
P2S_500pg	256	Nanodrop	208			
P2S_50pg	74	Nanodrop	21.2			
P3S_5ng	1,066	Qubit	1,200	Most fragments ranging from 130bp-1500bp with average fragment size ~350-450bp	2	Took two days to execute the protocol with minimal hands on time. Performed two purifications to completely get rid of primer
P3S_500pg	353	Qubit	337			
P3S_50pg	18	Qubit	15			

Read Distribution Following Alignment

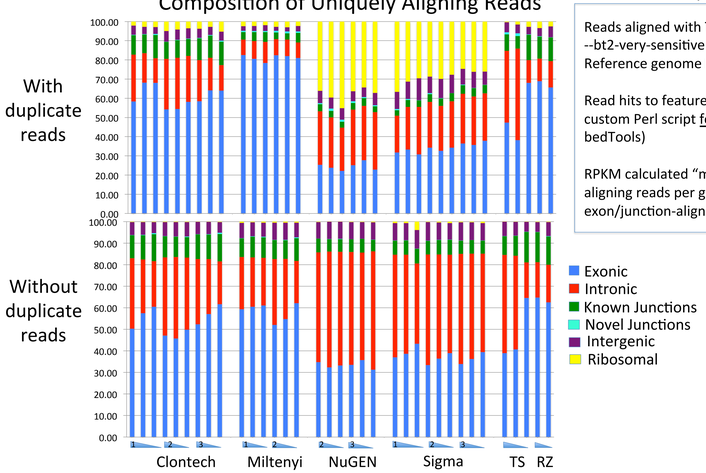


Starting Amounts

Blue wedges indicate increasing [RNA]
For kits, this corresponds to 50 pg, 500 pg, and 5 ng
For TruSeq (TS) it is 5 ng, 100 ng, and 1 ug
For RiboZero (RZ) it is 100 ng and 1 ug

All Aligning
Uniquely Aligning
Uniquely Aligning w/o Duplicates

Composition of Uniquely Aligning Reads



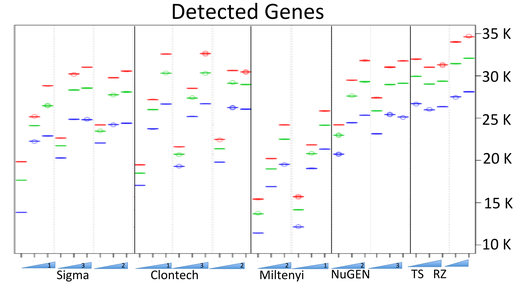
Analysis Methods

Reads aligned with Tophat 2.0.6 (+ Bowtie 2.0.2) --bt2-very-sensitive setting
Reference genome Ensembl 66 GTF, UCSC hg19

Read hits to features were quantitated using a custom Perl script featureMapper (based on bedTools)

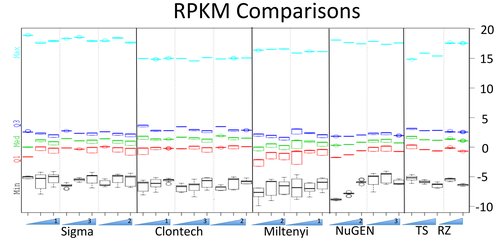
RPKM calculated "manually": R = exon/junction-aligning reads per gene, K = exonic bp, M = all exon/junction-aligning reads

Detected Genes



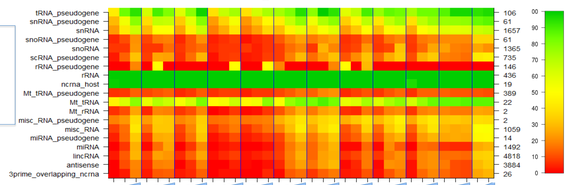
Genes detected with 5 (B), 2 (G), or 1 (R) reads
5 samplings at 10 M exonic reads/sampling

RPKM Comparisons



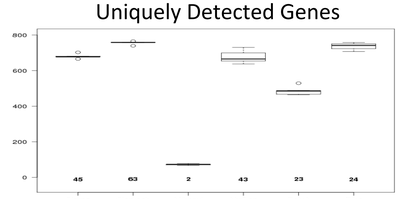
Values on RH Y axis are log 2 (RPKM) values

Non-coding RNA Identification



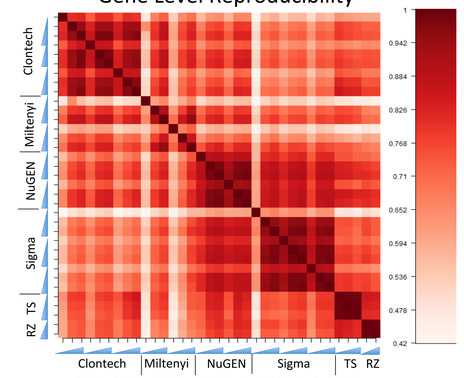
Scale indicates average % detected ncRNAs sampled at 10 million reads.
Values shown at right are the "known" numbers of these biotypes used in the percent calculation.

Uniquely Detected Genes



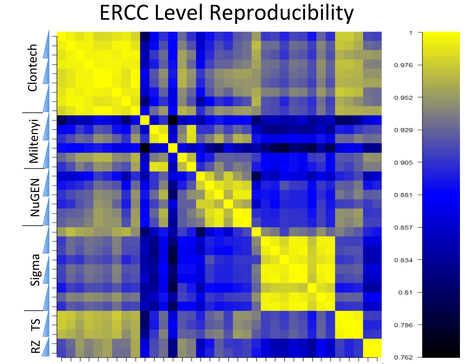
Number of genes uniquely detected by kit, determined by downsampling 10 million reads 5 times
The number of uniquely-detected genes **common to all 5 replicates** is shown at bottom

Gene Level Reproducibility



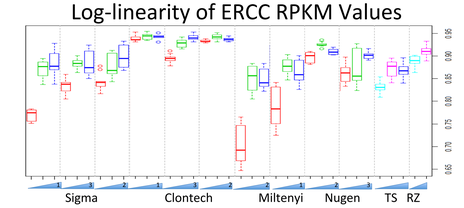
Pearson Correlations among all genes averaged replicate log2 RPKMs
Log2(Median Library FPKM) per gene.

ERCC Level Reproducibility



Pearson Correlations between averaged replicate log2 RPKMs

Log-linearity of ERCC RPKM Values



"Perfect" ERCC values would lie on a straight log-line, giving a correlation value of 1.0

Conclusions

1. Sub nanogram amounts of starting RNA can generate libraries displaying excellent performance metrics
2. All kits showed increased variability at the lowest [RNA]
3. PolyA and non-polyA based methods both work well
4. Site to site variability was minimal, particularly at higher [RNA]

Acknowledgements

As always the DSRG is particularly grateful for company support, since only that allows these studies to be carried out. We thank:

Illumina for sequencing reagents
Sigma-Aldrich, **NuGEN**, **Miltenyi Biotec** and **Clontech** for sample preparation kits
Clontech for the Universal Human RNA
LifeTech (**Ambion**) for the ERCC RNA standards
SeqGen for providing resources for incidentals
The DSRG is also particularly indebted to the above-and-beyond contributions of Anoja Perera, Ariel Paulson , and Alison peak in library construction, data generation, and analysis