# A Comparison of Single-cell RNA-seq with Gene Expression Microarrays


Paul Scheid[1], Clay Spencer[2,3], Michelle Gutwein[1], Ashish Agarwal[1], Kristin C. Gunsalus[1], David Miller III[3]

1. Center for Genomics and Systems Biology, Department of Biology, New York University, New York, USA
2. Current Address: Department of Neurosciences, Case Western Reserve University, Cleveland, Ohio, USA
3. Department of Cell and Developmental Biology, Vanderbilt University, Nashville, Tennessee, USA


## Abstract

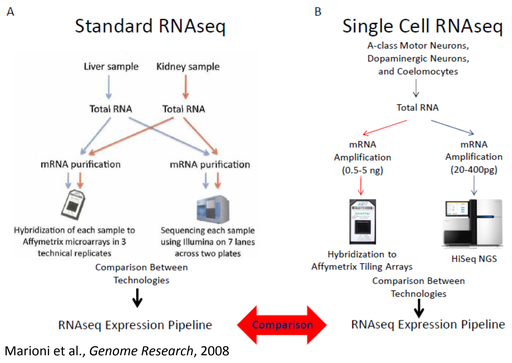

Single-cell RNA-sequencing (SCRS) is a powerful technique to address biological variation by profiling expression in single cells and samples with low RNA input. Previous studies have shown that gene expression is highly correlated across array and standard RNA-sequencing technologies. However, no comparative studies utilizing SCRS and the same starting sample across platforms have been reported.

We compared expression data from tiling arrays and SCRS on RNA harvested from embryonic cholinergic motor neurons (dorsal A), embryonic coelomocytes (macrophage-like cell), and larval dopaminergic neurons from *C. elegans*. Picogram quantities of total RNA from each sample were amplified using a single-cell protocol to generate double stranded cDNAs, which were then sequenced with the Illumina HiSeq platform. The same RNA samples from each cell type were previously amplified using the NuGEN WT-Ovation Pico protocol and hybridized to Affymetrix tiling arrays[1].

We compared log2 FPKM counts for each gene with the corresponding RMA-normalized log2 array expression values. These two independent measures of transcript expression are highly correlated (Spearman correlation = 0.75 for coelomocytes, 0.62 for A-class motor neurons, and 0.68 for dopaminergic neurons). Moreover, SCRS data showed several hundred genes that are significantly enriched in each cell type in comparison with existing whole animal RNA-seq from the same developmental stage, and these significantly overlap genes detected as enriched from the tiling array data[1] ($p < 5.38e\text{-}36$ for all three sets).

In sum, the correlation of SCRS to tiling array is high, similar to published comparisons between microarrays and standard RNA-seq where at least one thousand-fold more starting material was used. These results suggest that single-cell RNA-sequencing is a robust tool for gene expression quantification and transcriptome profiling when input material is limiting.

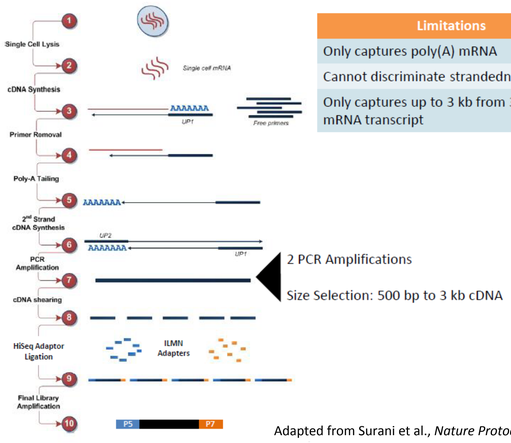

## Study Design




Marioni et al., *Genome Research*, 2008

Graphical representation of study design. (A) Experimental design was based on framework published by Marioni et. al. that compared RNA-seq to expression arrays hybridized with the same biological samples (liver and kidney). (B) Picogram quantities of RNA harvested from embryonic cholinergic motor neurons, embryonic coelomocytes (macrophage-like cell), and late larval dopaminergic neurons were profiled using tiling arrays and SCRS.

## Single-cell RNA-seq




| Limitations |
|---|
| Only captures poly(A) mRNA |
| Cannot discriminate strandedness |
| Only captures up to 3 kb from 3' end of mRNA transcript |

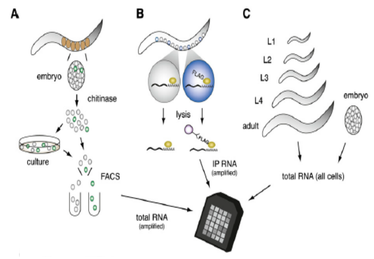
2 PCR Amplifications

Size Selection: 500 bp to 3 kb cDNA

Adapted from Surani et al., *Nature Protocols*, 2010

Schematic of single-cell RNA-sequencing library construction for transcriptome profiling of single cells and samples with low RNA input. Picogram quantities of mRNA are reverse transcribed into cDNAs using an oligo d(t) primer with universal anchor sequence UP1; undigested primers are removed enzymatically. Poly(A) tails are added to the 3' end of the first cDNA strand and the second cDNA strand is synthesized using an oligo d(t) primer with anchor sequence UP2. The double stranded cDNAs are amplified by PCR using UP1/UP2 primers, fragmented, and subject to standard Illumina library construction.

## Sample Summary

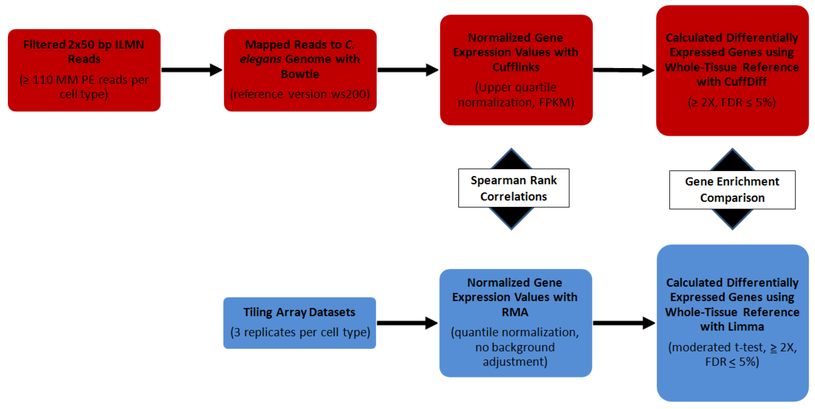



Spencer & Zeller, et al. *Genome Research*, 2011

| Cell Population | Description |
|---|---|
| Dopaminergic Neuron (DN) | Larval (L3/L4) dopaminergic neuron RNA obtained by mRNA tagging. |
| Cholinergic Motor Neurons (CMN) | Embryonic A-class motor neuron RNA obtained from FACS-isolated cells. |
| Coelomocytes (CL) | Embryonic Coelomocyte RNA obtained from FACS-isolated cells. |

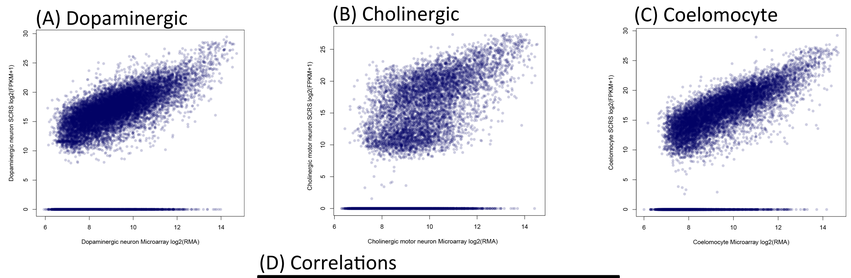
RNA harvest strategies for transcriptome profiling with gene expression arrays and SCRS. (A) Embryonic cells are dissociated by chitinase treatment and are either sorted immediately or cultured 24 hours before FACs (CMN and CL). (B) Epitope-tagged (FLAG) polyA binding protein (PAB-1) is expressed in specific larval cells; the PAB-1:RNA complex is immunoprecipitated (DN). (C) Total RNA is isolated from embryonic and larval worms (used as whole-tissue reference to determine enriched gene populations).

## Analysis Workflow




Schematic of analysis workflow used to compare SCRS datasets to tiling array correlates.

## Expression Comparison Across Platforms

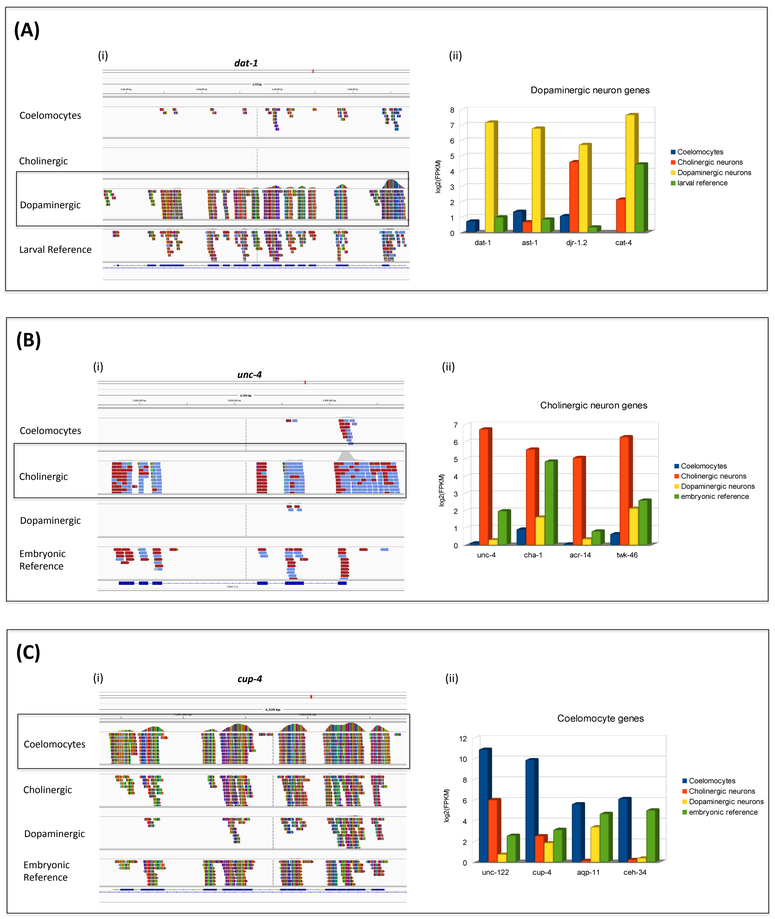



(D) Correlations

| Cell Population | Spearman Correlation |
|---|---|
| DN-Array vs. DN-SCRS | 0.75 |
| CMN-Array vs. CMN-SCRS | 0.62 |
| CM-Array vs. CM-SCRS | 0.68 |

Comparison of log2 FPKM+1 counts from SCRS (Y-axis) to log2 RMA normalized expression values from arrays (X-axis) for (A) DN, (B) CMN and (C) CM cell types. (D) Correlations across platforms similar to published comparisons between microarrays and RNA-seq where standard starting inputs were used.

## Gene Enrichment Comparison




**(D)**

| Cell-type | Gene | Function | Enrichment (log2 ratio) |
|---|---|---|---|
| Dopaminergic neurons | dat-1 | dopamine transporter | 8.19 |
| | djr-1.2 | DJ-1/PARK7 homolog | 7.53 |
| | ast-1 | transcription factor | 6.96 |
| | cat-4 | GTP cyclohydrolase | 6.30 |
| Cholinergic motor neurons | unc-4 | transcription factor | 10.72 |
| | acr-14 | acetylcholine receptor | 10.12 |
| | twk-46 | potassium channel | 7.34 |
| | cha-1 | choline acetyltransferase | 5.63 |
| Coelomocytes | unc-122 | transmembrane protein | 9.80 |
| | aqp-11 | aquaporin | 8.17 |
| | ceh-34 | transcription factor | 8.00 |
| | cup-4 | ion channel | 7.22 |

Examples of genes reported as significantly enriched by SCRS versus whole-tissue reference for (A) DN, (B) CMN and (D) CL. (Ai, Bi, Ci) Coverage examples for one enriched gene from each cell type. (Aii, Bii, and Cii) Log2FPKM values of genes identified as significantly enriched in the corresponding tiling array dataset. (D) Table detailing gene function and enrichment ratios from each cell type.

Overall, SCRS data showed several hundred genes that are significantly enriched in each cell type in comparison with existing whole animal RNA-seq from the same developmental stage, and these significantly overlap genes detected as enriched from the tiling array data ($p < 5.38e\text{-}36$ for all three sets).

## Conclusions

1. Correlation across platforms is high despite different amplification procedures used by Affymetrix tiling array and Illumina HiSeq SCRS (NuGEN WT-PICO versus Single-Cell Amplification, respectively)
2. Data shows strong enrichment for genes expected to be highly expressed in each cell type.
3. Data suggests SCRS is a robust tool for gene expression quantification and transcriptome profiling when input material is limiting.

## References


1. Spencer & Zeller, et al. *Genome Research*, 2011 Feb;21(2):325-41.
2. Tang et al. *Nature Protocols,* 2010;5(3):516-35.
3. Marioni et al. *Genome Research*, 2008 Sep;18(9):1509-17



Acknowledgement: The authors wish to thank Jane Carlton and NYU CGSB for providing sequencing support and AGBT conference funds.