OPEN ACCESS Freely available online

PLOS ONE

Sequencing mRNA from Cryo-Sliced *Drosophila* Embryos to Determine Genome-Wide Spatial Patterns of Gene Expression

Peter A. Combs[1]*, Michael B. Eisen[2,3]

1 Graduate Program in Biophysics, University of California, Berkeley, California, United States of America, **2** Department of Molecular and Cell Biology, University of California, Berkeley, California, United States of America, **3** Howard Hughes Medical Institute, University of California, Berkeley, California, United States of America

Abstract

Complex spatial and temporal patterns of gene expression underlie embryo differentiation, yet methods do not yet exist for the efficient genome-wide determination of spatial expression patterns during development. *In situ* imaging of transcripts and proteins is the gold-standard, but it is difficult and time consuming to apply to an entire genome, even when highly automated. Sequencing, in contrast, is fast and genome-wide, but is generally applied to homogenized tissues, thereby discarding spatial information. To take advantage of the efficiency and comprehensiveness of sequencing while retaining spatial information, we cryosectioned individual blastoderm stage *Drosophila melanogaster* embryos along the anterior-posterior axis and developed methods to reliably sequence the mRNA isolated from each 25 μm slice. The spatial patterns of gene expression we infer closely match patterns previously determined by *in situ* hybridization and microscopy. We applied this method to generate a genome-wide timecourse of spatial gene expression from shortly after fertilization through gastrulation. We identified numerous genes with spatial patterns that have not yet been described in the several ongoing systematic *in situ* based projects. This simple experiment demonstrates the potential for combining careful anatomical dissection with high-throughput sequencing to obtain spatially resolved gene expression on a genome-wide scale.

Citation: Combs PA, Eisen MB (2013) Sequencing mRNA from Cryo-Sliced *Drosophila* Embryos to Determine Genome-Wide Spatial Patterns of Gene Expression. PLoS ONE 8(8): e71820. doi:10.1371/journal.pone.0071820

Editor: Barbara Jennings, University College London, United Kingdom

Received May 21, 2013; **Accepted** July 4, 2013; **Published** August 12, 2013

Funding: This work was funded by a Howard Hughes Medical Institute investigator award to MBE, and National Institutes of Health training grant #T32 HG 00047 to PAC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: MBE is a cofounder and member of the Board of Directors of PLOS. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: peter.combs@berkeley.edu

Introduction

Analyzing gene expression in multicellular organisms involves a tradeoff between the spatial precision of imaging and the efficiency and comprehensiveness of genomic methods. RNA *in situ* hybridization (ISH) and antibody staining of fixed samples, or fluorescent imaging of live samples, provides high resolution spatial information for small numbers of genes [1–3]. But even with automated sample preparation, imaging, and analysis, *in situ* based methods are difficult to apply to an entire genome's worth of transcripts or proteins. High throughput genomic methods, such as DNA microarray hybridization or RNA sequencing, are fast and relatively inexpensive, but, at least for the small species worked with in most labs, the amount of input material they require has generally limited their application to homogenized samples, often from multiple individuals. Methods involving the tagging, sorting, and analysis of RNA from cells in specific spatial domains have shown promise [4], but remain non-trivial to apply systematically, especially across genotypes and species.

Recent advances in DNA sequencing suggest an alternative approach. With increasingly sensitive sequencers and improved protocols for sample preparation, it is now possible to analyze small samples without amplification. Several years ago we developed methods to analyze the RNA from individual *Drosophila* embryos [5]. As we often recovered more RNA from each embryo than was required to obtain accurate measures of gene expression, we wondered whether we could obtain good data from pieces of individual embryos, and whether we could obtain reliable spatial expression information from such data. To test this possibility, we chose to focus on anterior-posterior (A–P) patterning in the early *D. melanogaster* embryo, as the system is extremely well-characterized and the geometry of the early embryo also lends itself to biologically meaningful physical dissection by simple sectioning along the elongated A–P axis.

Results

To test whether we could consistently recover and sequence RNA from sectioned *D. melanogaster* embryos, we collected embryos from our laboratory stock of the line CantonS (CaS), aged them for approximately 2.5 hours so that the bulk of the embryos were in the cellular blastoderm stage, and fixed them in methanol. We examined the embryos under a light microscope and selected single embryos that were roughly halfway through cellularization (mitotic cell cycle 14; developmental stage 5). We embedded each embryo in a cryoprotecting gel, flash-froze it in liquid nitrogren, and took transverse sections along the anterior-posterior axis. For this initial trial we used 60 μm sections, meaning that we cut each approximately 350 μm embryo into six pieces. We placed each

piece into a separate tube, isolated RNA using Trizol, and prepared sequencing libraries using the Illumina Tru-Seq kit.

In early trials we had difficulty routinely obtaining good quality RNA-seq libraries from every section. We surmised that we were losing material from some slices during library preparation as a result of the small amount (approximately 15 ng) of total RNA per slice. To overcome this limitation, after the initial RNA extraction we added RNA from a single embryo of a distantly related *Drosophila* species to each tube to serve as a carrier.

We used RNA as a carrier, instead of a standard carrier like salmon sperm DNA or linear acrylamide, so that the carrier was present throughout the experiment, and we used RNA from multiple *Drosophila* species in particular so that the sequence reads from the carrier RNA would not be wasted. In this first experiment the carrier RNA was part of an experiment examining gene expression in early embryos of other *Drosophila* species. We only used embryos from species that were fully sequenced and sufficiently diverged from *D. melanogaster* to allow us to readily separate reads derived from the *D. melanogaster* slice and the carrier species computationally after sequencing.

With the additional approximately 100 ng of total RNA from the carrier in each sample, library preparation became far more robust. We independently sliced three CaS embryos, prepared libraries from the sliced RNA using the standard TruSeq RNA kit, and sequenced them using an Illumina HiSeq 2000, obtaining approximately 40 million 50 bp paired-end reads for each slice+carrier sample. We aligned these reads to the *D. melanogaster* and carrier genomes using TopHat [6,7], and identified between 1.7 and 31.4 percent of reads as having come unambiguously from *D. melanogaster* (see Table 1). We then used Cufflinks [8] to infer expression levels for all annotated mRNAs using the *D. melanogaster* reads alone.

The data for each slice within an embryo were generally highly correlated (Figure S1), reflecting the large number of highly expressed genes with spatially uniform expression patterns. The data for equivalent slices of embryos 2 and 3 were also highly correlated, while the slices for embryo 1 were systematically less well matched to their counterparts in embryos 2 and 3 (Figure S2), suggesting that it may have been sampled at a slightly different developmental stage.

To examine how well our data recapitulated known spatial profiles, we identified a panel of genes with known anterior-posterior patterns of gene expression and compared our data to their published expression patterns. Figure 1A shows RNA in-situ hybridization patterns from the Berkeley Drosophila Genome Project (BDGP) [2] alongside the expression data for that gene from our sliced embryos, demonstrating a close qualitative agreement between the visualized expression patterns and our sliced RNA-seq data.

In order to more quantitatively compare our data to existing patterns, we constructed a reference set of spatial expression patterns along the A-P axis using three-dimensional "virtual embryos" from the Berkeley Drosophila Transcription Network Project, which contain expression patterns for 95 genes at single-nucleus resolution [1]. We transformed the relative expression levels from these images into absolute values (FPKM) using genome-wide expression data from intact single embryos [5]. We compared the observed expression for these 95 genes from an average of each of our slices to all possible 60 μm slices of these virtual embryos (Figure 1B). High scores for most slices fell into narrow windows, with the best matches for each slice falling sequentially along the embryo with a spacing of about 60 μm, the same thickness as the slices.

We next used the program Cuffdiff [9] to identify 85 genes with statistically significant differences in expression between slices (Dataset S2; this is a very conservative estimate). We compared these genes to those examined by the BDGP, the most comprehensive annotation of spatial localization in *D. melanogaster*

Table 1. Sequencing statistics for sliced single-stage wild-type mRNA-Seq samples.

Replicate	Slice	Carrier Species	Barcode Index	Total Reads	Uniquely mapped *D. mel* reads (%)	Ambiguous Reads (%)
1	1	*D. per*	1	69,339,972	2,284,228 (3.2%)	1,634,055 (2.3%)
1	2	*D. per*	2	73,632,862	3,706,630 (5.0%)	1,603,444 (2.1%)
1	3	*D. per*	3	82,076,328	6,002,034 (7.3%)	1,774,485 (2.1%)
1	4	*D. per*	4	73,437,708	6,401,565 (8.7%)	1,592,665 (2.1%)
1	5	*D. per*	5	75,922,812	4,951,178 (6.5%)	1,559,097 (2.0%)
1	6	*D. per*	6	78,623,784	1,355,079 (1.7%)	1,574,067 (2.0%)
2	1	*D. wil*	7	59,813,036	4,066,295 (6.7%)	878,476 (1.4%)
2	2	*D. wil*	8	90,961,338	15,212,716 (16.7%)	1,301,095 (1.4%)
2	3	*D. wil*	9	73,201,902	14,855,374 (20.2%)	911,768 (1.2%)
2	4	*D. wil*	10	75,754,772	23,858,301 (31.4%)	1,136,031 (1.4%)
2	5	*D. wil*	11	84,497,566	10,026,713 (11.8%)	1,080,910 (1.2%)
2	6	*D. wil*	12	66,316,952	13,122,508 (19.7%)	898,776 (1.3%)
3	1	*D. moj*	13	75,847,986	12,496,248 (16.4%)	3,615,452 (4.7%)
3	2	*D. moj*	14	72,497,660	4,005,714 (5.5%)	803,381 (1.1%)
3	3	*D. moj*	15	77,532,368	11,138,154 (14.3%)	772,446 (0.9%)
3	4	*D. moj*	16	83,400,882	8,227,562 (9.8%)	861,839 (1.0%)
3	5	*D. moj*	18	83,608,454	2,630,069 (3.1%)	795,169 (0.9%)
3	6	*D. moj*	19	85,823,784	2,239,493 (2.6%)	829,382 (0.9%)

Counts are for read ends. Discordant read ends are always classed as ambiguous, but failure of one end to map does not disqualify the other.
doi:10.1371/journal.pone.0071820.t001

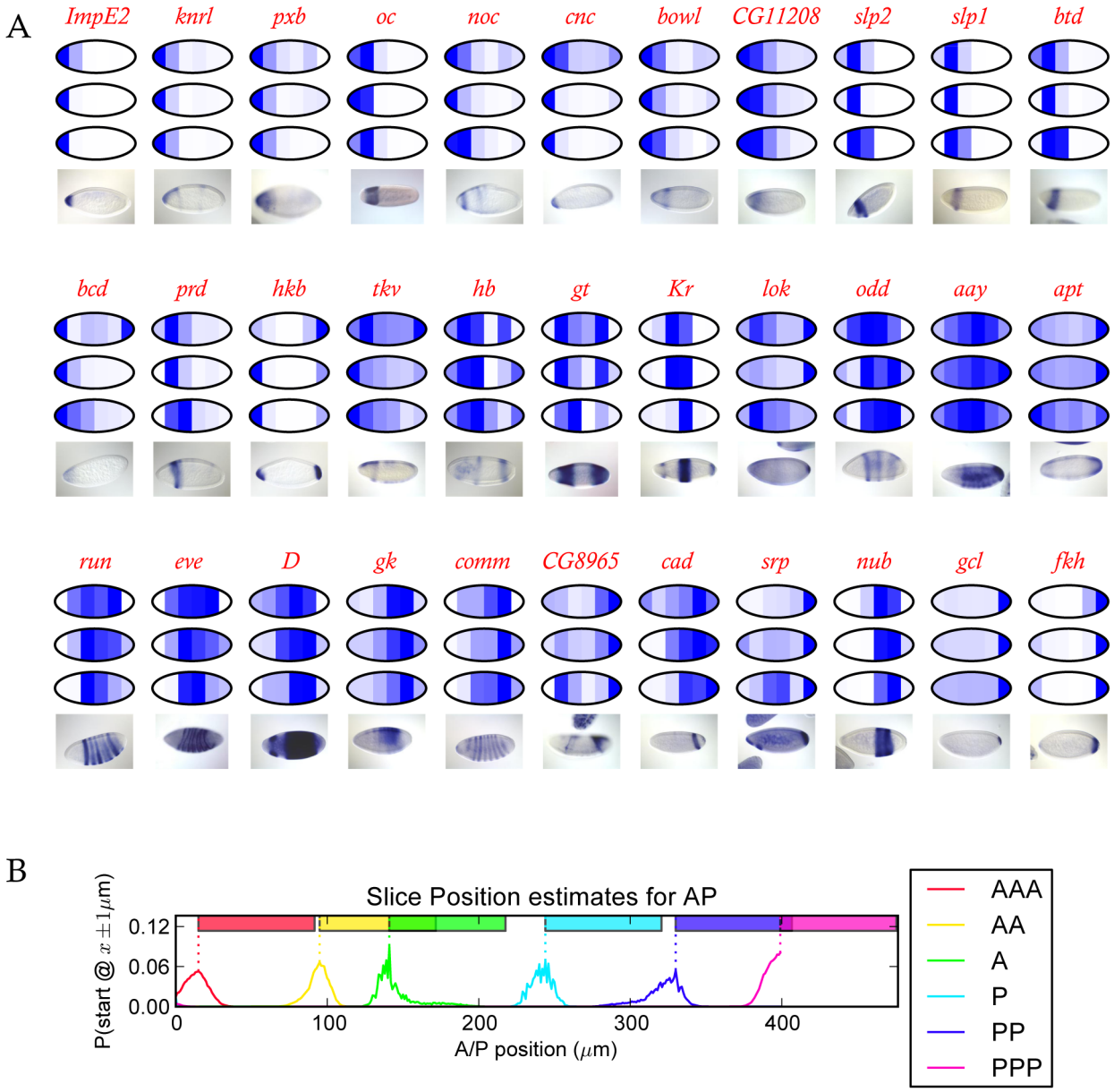

Figure 1. Expression in the slices closely matches published expression data. (A) Published *in situ* patterns for 33 genes are shown alongside reconstrucred spatial patterns for these genes from each of the three 60 μm sliced CaS embryos. The reconstructed patterns were each scaled to the slice with the highest expression level for each embryo individually. (B) To evaluate the overall quality of our reconstructed spatial expression patterns, we compared expression levels of 98 genes from each slice in our 60 μm data (averaged across the three embryos) to all possible 60 μm sections from a cellular resolution spatial atlas of gene expression from the Berkeley Drosophila Transcription Network Project [1] with absolute expression levels computed using data from [5]. We computed the posterior probability that a slice from our data corresponded to a slice from the BDTNP atlas using a simple Bayesian procedure that compares the level of each gene in a slice to the level of that gene in sections of the atlas. The line graphs are the posterior probabilities that each slice started at a given position in the atlas. Each slice has a clear peak and the ordering of the peaks corresponds to the ordering of the slices, as expected. The colored bars show the portion of the embryo spanned by the slice assuming it begins at the peak in the posterior probability distribution.
doi:10.1371/journal.pone.0071820.g001

development that we are aware of [2]. Of our differentially expressed genes, 21 had no imaging data available, and 33 were annotated as present in a subset of the embryo (the annotation term meant to capture patterned genes); the remaining 31 genes showed either clear patterns that were not annotated with the most general keyword, or no clear staining (Figure S3). There were 194 genes tagged by the BDGP as patterned that were not picked up as having statistically significant patterns in our data. However, most of these had primarily dorsal-ventral patterns, faint patterns, later staging in the images used for annotation, or had good qualitative agreement with our data but fell above the cutoff for statistical significance (Figure S4).

As a more sensitive approach to finding patterned genes, we applied k-means clustering to our data. We first filtered on expression level (at least one slice in one embryo with FPKM >10) and agreement between replicates (average Pearson correlation between embryos of >0.5), then clustered based on normalized expression (k = 20, centroid linkage; 20 was chosen empirically as smaller k's merged genes with different patterns and larger k's provided no additional useful information) [10]. We identified several broad classes of expression, including localization to each of the poles, and five different gap gene-like bands along the AP axis (Figure 2 and Figure S5). Of the 745 genes, only 349 had images in the BDGP set [2]. Staining for these genes is sometimes undetectable and well-matched stages are often missing from the databases, but where comparisons were possible, the BDGP image data agrees with our RNA-seq patterns (Figure S6).

To extend our dataset, we collected individual embryos from seven different time points based on morphology—stage 2, stage 4, and 5 time points within stage 5—and sliced them into 25 μm sections, yielding between 10 and 15 contiguous, usable slices per embryo. For these embryos we used total RNA from the yeasts *Saccharomyces cerevisiae* and *Torulaspora delbruckii* as carrier, which are so far diverged as to have fewer than 0.003% of reads ambiguously mapping.

These finer slices are better able to distinguish broad gap-gene domains, with several slices of relatively low expression between the multiple domains of *hb*, *kni*, and *gt*. Excitingly, we can also distinguish the repression between stripes of pair-rule genes like *eve* as well (Figure 3). Given the non-orthogonal orientation of the anterior-most and posterior-most *eve* stripes relative to the AP axis, we do not expect to see all 7 pair-rule stripes, but at least three can be unambiguously observed.

Putting the 60 μm and 25 μm slice datasets together, we find a large number of genes with reproducible patterns in the 60 μm slices whose formation over time can be clearly seen in the timed 25 μm slices, including many without previously described early patterns (Figure S7).

Discussion

The experiments reported here demonstrate that slicing and sequencing animal embryos is a practical and effective method to systematically characterize spatial patterns of expression. While we are by no means the first to dissect samples and characterize their RNAs—Ding and Lipshitz pioneered this kind of analysis twenty years ago [11]—to our knowledge we are the first to successfully apply such a technique to report genome-wide spatial patterns in a single developing animal embryo.

Given the degree to which the *D. melanogaster* embryo has been studied, and the presence of at least two large *in situ* based studies whose goals were to systematically identify and characterize genes with patterned expression in the embryo, we were surprised by the large number of genes we find as clearly patterned that had not been previously described as such. We note in particular a large

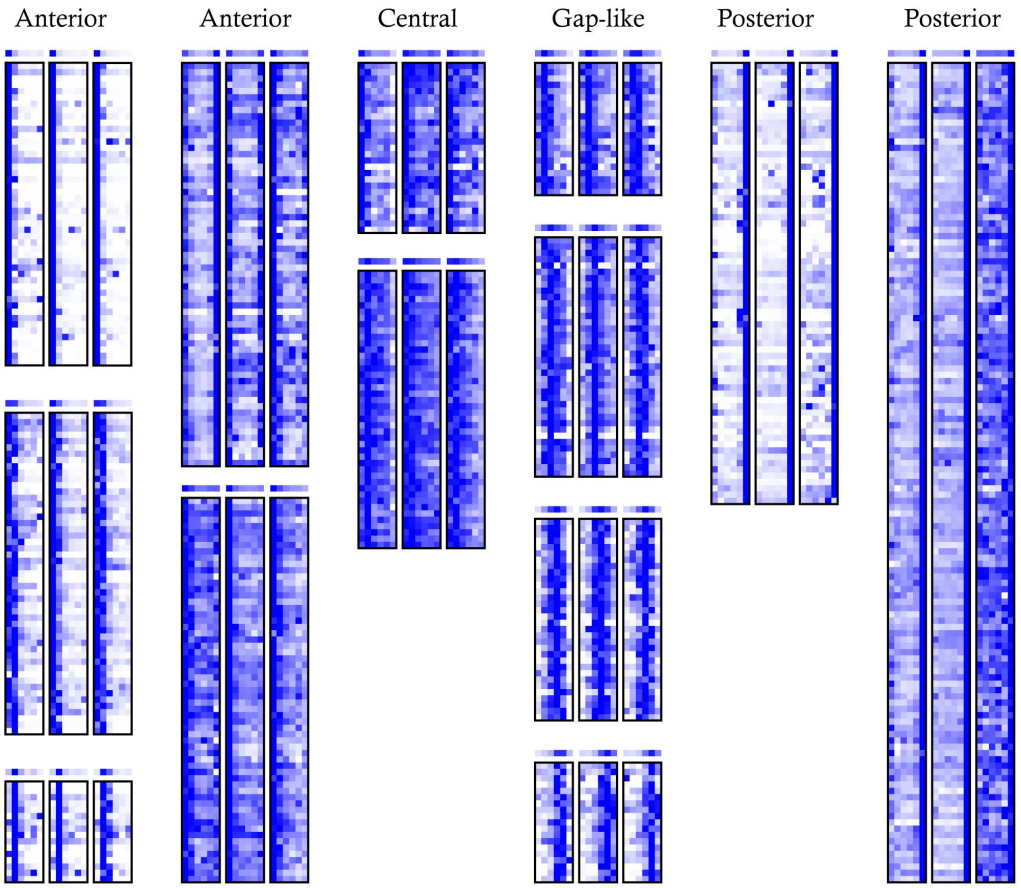

Figure 2. Heat maps of gene expression clusters. Of the k = 20 clusters, 13 with non-uniform patterns are shown. The expression levels for each gene was normalized for clustering and display so that the maximum expression of each gene in each embryo is dark blue. The plot above each cluster is the mean normalized expression level in that cluster. All clusters are listed in Dataset S3.
doi:10.1371/journal.pone.0071820.g002

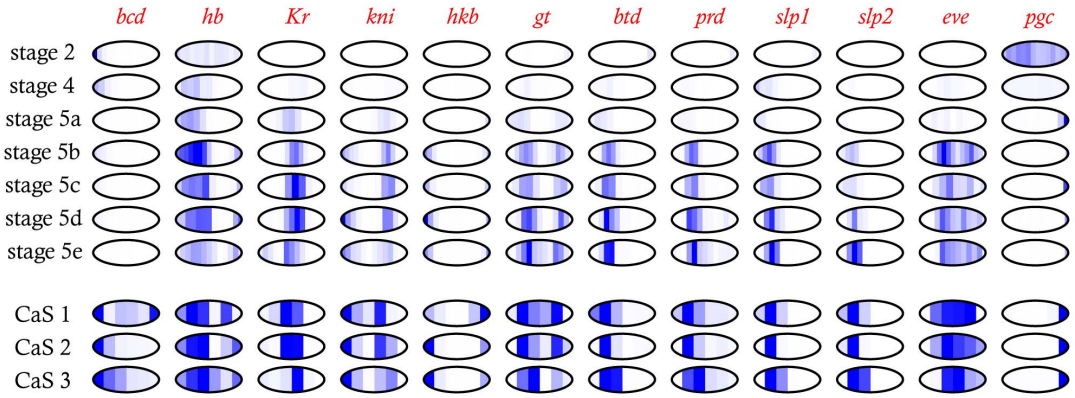

Figure 3. Expression of key patterning genes across early development. Expression levels in the 25 µm timeseries are normalized to the highest expression level at any time pioint. For slices with poor quality data (timepoint 4, slice 10; timepoint 6, slice 6; timepoint 7, slice 7; and timepoint 7, slice 8) data imputed from neighboring slices is shown. Expression levels for the 60 µm slice samples are normalized to the highest level in each embryo.
doi:10.1371/journal.pone.0071820.g003

number of genes with expression restricted to the poles, most with no known role in either anterior patterning or pole cell formation or activity. This emphasizes the potential for sequencing-based methods to replace *in situ* based studies in the systematic analysis of patterned gene expression, as they are not only simpler, cheaper, and easier to apply to different species and genetic backgrounds, but appear to be more sensitive.

The data we present here are far from perfect - the relatively small number of reads per slice (due to the presence and sequencing of carrier RNA) means that the slice by slice data are somewhat noisy. However the consistency between replicates and the agreement between the 25 µm and 60 µm data demonstrate that the experiment clearly worked, and additional sequencing depth and better methods for working with small samples should greatly reduce the noise as we move forward.

Obviously, to truly replace *in situ* based methods, sequencing based methods will need to achieve greater resolution than presented here. One can envision several basic approaches to achieving the ultimate goal of determining the location of every RNA in a spatially complex tissue. Sequencing RNAs in place in intact tissues would obviously be the ideal method, and we are aware of several groups working towards this goal. In the interim, however, methods to isolate and characterize smaller and smaller subsets of cells are our only alternative. One possibility is to combine spatially restricted reporter gene expression and cell sorting to purify and characterize the RNA composition of differentiated tissue—c.f. [4]. While elegant, this approach cannot be rapidly applied to different genetic backgrounds, requires separate tags for every region/tissue to be analyzed, and will likely not work on single individuals.

Sectioning based methods offer several advantages, principally that they can be applied to almost any sample from any genetic background or species, and allow for the biological precision of investigating single individuals. The 60 µm and 25 µm slices we used here represent reasonable tradeoffs between sequencing depth and spatial resolution given the current limits of sample preparation and sequencing methods, but with methods having been described to sequence the RNAs from single cells, and with sequencing costs continuing to plummet, it should be possible to obtain far better resolution in the near future. A rough estimate suggests that a single embryo contains enough RNA to sequence over 700 samples to a depth of 20 million reads. Thus it is theoretically possible to dice an embryo into 20 µm cubes and sequence each one to obtain genome-wide three-dimensional expression data, although this presents several difficult but likely solvable technical challenges, especially handling and tracking hundreds or thousands of tiny samples.

Materials and Methods

Fly Line, Imaging, and Slicing

We raised flies on standard media at 25° in uncrowded conditions, and collected eggs from many 3 to 10-day old females from our *Canton-S* lab stocks. We washed and dechorionated the embryos, then fixed them according to a standard methanol cracking protocol. Briefly, we placed embryos in 20 ml glass vials containing 10 ml of heptane and 10 ml of PEM (100 mM PIPES, 2 mM EGTA, 1 mM MgSO4) and mixed gently. We then removed the aqueous phase, added 10 ml of methanol, shook vigorously for 15–30 seconds, and collected the devitellinized embryos, which we washed several times in methanol to remove residual heptane. We then placed the fixed embryos on a slide in halocarbon oil, and imaged on a Nikon 80i with DS-5M camera. After selecting embryos with the appropriate stage according to depth of membrane invagination and other morphological features, we washed embryos with methanol saturated with bromophenol blue dye (Fisher, Fair Lawn NJ), aligned them in standard cryotomy cups (Polysciences Inc, Warrington, PA), covered them with OCT tissue freezing medium (Triangle Biomedical, Durham, NC), and flash froze them in liquid nitrogen.

We sliced frozen embryos on a Microm HM 550 (Thermo Scientific, Kalamozoo, MI) at a thickness of 60 µm or 25 µm. We adjusted the horizontal position of the blade after every slice to eliminate the possibility of carry-over from previous slices, and used a new blade for every embryo. We placed each slice in an individual RNase-free, non-stick tube (Life Technologies, Grand Island, NY).

RNA Extraction, Library Preparation, and Sequencing

We performed RNA extraction in TRIzol (Life Technologies, Grand Island, NY) according to manufacturer instructions, except with a higher concentration of glycogen as carrier (20 ng) and a higher relative volume of TRIzol to the expected material (1 mL, as in [5]). For the 60 µm slices, we pooled total RNA from each slice with total RNA from single *D. persimilis*, *D. willistoni*, or *D. mojavensis* embryos, then made libraries according to a modified

TruSeq mRNA protocol from Illumina. We prepared all reactions with half-volume sizes to increase relative sample concentration, and after AmpureXP cleanup steps, we took care to pipette off all of the resuspended sample, leaving less than 0.5 μL, rather than the 1–3 μL in the protocol. Furthermore, we only performed 13 cycles of PCR amplification rather than the 15 in the protocol, to minimize PCR duplication bias.

Libraries were quantified using the Kapa Library Quantification kit for the Illumina Genome Analyzer platform (Kapa Biosystems) on a Roche LC480 RT-PCR machine according to the manufacturer's instructions, then pooled to equalize index concentration. Pooled libraries were then submitted to the Vincent Coates Genome Sequencing Laboratory for 50bp paired-end sequencing according to standard protocols for the Illumina HiSeq 2000. Bases were called using HiSeq Control Software v1.8 and Real Time Analysis v2.8.

Mapping and Quantification

Reads were mapped using TopHat v2.0.6 to a combination of the FlyBase reference genomes (version FB2012_05) for *D. melanogaster* and the appropriate carrier species genomes with a maximum of 6 read mismatches [12,13]. Reads were then assigned to either the *D. melanogaster* or carrier genomes if there were at least 4 positions per read to prefer one species over the other. We used only the reads that mapped to *D. melanogaster* to generate transcript abundances in Cufflinks.

Data and Software

We have deposited all reads in the NCBI GEO under the accession number GSE43506. The processed data are available at the journal website (Dataset S1) and at http://eisenlab.org/sliceseq with a search feature for the 25 μm dataset. All custom analysis software is available https://github.com/eisenlab/SliceSeq, and is primarily written in Python [14–18]. Commit b0b115a was used to perform all analyses in this paper.

Supporting Information

Figure S1 Correlation of slices within embryos. Log-log plots of FPKM values between slices within each of the three 60 μm sliced embryos.
(PDF)

Figure S2 Correlation of slices between embryos. Log-log plots of FPKM values of corresponding slices between each of the three 60 μm sliced embryos.
(TIF)

Figure S3 Genes called as patterned by Cuffdiff lacking subset tag in BDGP database. Images are from BDGP; graphs are average of three CaS embryos. Many of these are known patterned genes, highlighting the incompleteness of available annotations.
(TIF)

Figure S4 Genes with subset tag in BDGP not called as patterned by Cuffdiff.
(TIF)

Figure S5 Figure 2 with gene names.
(PDF)

Figure S6 Images from BDGP for genes in clusters shows in Figure 2.
(TIF)

Figure S7 Data from 25 μm timecourse and 60 μm embryos for a large number of genes with manually curated patterns.
(PDF)

Dataset S1 Normalized read counts per gene for each individual slice.
(ZIP)

Dataset S2 Differential expression calls from Cuffdiff. Listing of significantly different gene expression between slices; Anterior-most slice is AAA, posterior most is PPP.
(TSV)

Dataset S3 Listing of all clusters. Clustering was performed with $k = 20$, on genes filtered for minimum FPKM >10 and Pearson correlation >0.5. Includes clusters in figure 2 and uniform expression clusters.
(KGG)

Acknowledgments

We thank peer reviewer Boris Adryan for helpful comments, and many readers who contributed feedback on a preprint of the manuscript posted on MBE's blog and the arXiv. We also thank members of the Eisen lab for their assistance, especially Susan Lott and Jackie Villalta.

Author Contributions

Conceived and designed the experiments: PAC MBE. Performed the experiments: PAC. Analyzed the data: PAC MBE. Contributed reagents/materials/analysis tools: PAC MBE. Wrote the paper: PAC MBE.

References

1. Fowlkes CC, Hendriks CLL, Keränen SVE, Weber GH, Rübel O, et al. (2008) A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell 133: 364–374.
2. Tomancak P, Berman BP, Beaton A, Weiszmann R, Kwan E, et al. (2007) Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biology 8: R145.
3. Lécuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, et al. (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell 131: 174–187.
4. Steiner FA, Talbert PB, Kasinathan S, Deal RB, Henikoff S (2012) Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. Genome Research 22: 766–777.
5. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, et al. (2011) Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through singleembryo RNA-seq. PLoS Biology 9: e1000590.
6. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359.
7. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, et al. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14: R36.
8. Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7: 562–578.
9. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, et al. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnology 31: 46–53.
10. de Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics (Oxford, England) 20: 1453–1454.
11. Ding D, Lipshitz HD (1993) A molecular screen for polar-localised maternal RNAs in the early embryo of Drosophila. Zygote (Cambridge, England) 1: 257–271.
12. McQuilton P, St Pierre SE, Thurmond J, the FlyBase Consortium (2011) FlyBase 101 - the basics of navigating FlyBase. Nucleic Acids Research 40: D706–D714.
13. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (Oxford, England) 25: 1105–1111.
14. Van Rossum G, Drake FL (2003) Python language reference manual.

15. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics (Oxford, England) 25: 1422–1423.
16. Hunter JD (2007) Matplotlib: A 2D graphics environment. Computing In Science & Engineering 9: 90–95.
17. Jones E, Oliphant T, Peterson P, others (2001) SciPy: Open source scientific tools for Python.
18. Perez F, Granger BE (2007) IPython: a system for interactive scientific computing. Computing In Science & Engineering.