Cell

Resource

Genome-wide RNA Tomography in the Zebrafish Embryo

Jan Philipp Junker,[1,6] Emily S. Noël,[1,6] Victor Guryev,[2] Kevin A. Peterson,[3] Gopi Shah,[4] Jan Huisken,[4] Andrew P. McMahon,[3] Eugene Berezikov,[2,5] Jeroen Bakkers,[1,7] and Alexander van Oudenaarden[1,5,7,*]

[1]Hubrecht Institute, KNAW and University Medical Center Utrecht, Cancer Genomics Netherlands, 3584 CT Utrecht, the Netherlands
[2]European Research Institute for the Biology of Ageing, University of Groningen, University Medical Center Groningen, 9713 AV Groningen, the Netherlands
[3]Department of Stem Cell Biology and Regenerative Medicine, Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California Keck School of Medicine, Los Angeles, CA 90089, USA
[4]Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany
[5]Skolkovo Institute of Science and Technology (Skoltech), Novaya Street 100, Skolkovo Moscow Region 143025, Russia
[6]Co-first author
[7]Co-senior author
*Correspondence: a.vanoudenaarden@hubrecht.eu
http://dx.doi.org/10.1016/j.cell.2014.09.038

SUMMARY

Advancing our understanding of embryonic development is heavily dependent on identification of novel pathways or regulators. Although genome-wide techniques such as RNA sequencing are ideally suited for discovering novel candidate genes, they are unable to yield spatially resolved information in embryos or tissues. Microscopy-based approaches, using in situ hybridization, for example, can provide spatial information about gene expression, but are limited to analyzing one or a few genes at a time. Here, we present a method where we combine traditional histological techniques with low-input RNA sequencing and mathematical image reconstruction to generate a high-resolution genome-wide 3D atlas of gene expression in the zebrafish embryo at three developmental stages. Importantly, our technique enables searching for genes that are expressed in specific spatial patterns without manual image annotation. We envision broad applicability of RNA tomography as an accurate and sensitive approach for spatially resolved transcriptomics in whole embryos and dissected organs.

INTRODUCTION

The formation of spatially distinct gene expression domains is a ubiquitous process during metazoan development and is fundamental for the meticulous patterning required to drive embryogenesis. Identifying genetic pathways and regulators that are active in well-defined regions of the embryo is crucial to understand the processes of embryonic axis formation, tissue specification, and organ development. As a consequence, many studies in different model organisms and tissues, and at different developmental stages, have focused on identifying spatial patterns of gene expression on a large scale (Fowlkes et al., 2008; Geffers et al., 2012; Gray et al., 2004; Lécuyer et al., 2007; Thisse et al., 2004; Yu et al., 2012). Because such studies rely on microscopy-based approaches like mRNA in situ hybridization or immunohistochemistry—that unavoidably investigate only one or a few genes per sample—screening spatial expression patterns of the entire transcriptome has so far been out of reach. Conversely, RNA sequencing has emerged as a powerful tool to study gene expression on the genome-wide level, but is unable to yield spatially resolved information. Gene expression analysis after cell sorting can be used to determine cell-specific transcriptomes, but the spatial resolution as well as the number of different cell types that can be screened are limited (Birnbaum et al., 2003). In situ sequencing of RNA in intact tissues has the potential to provide direct information about the spatial organization of the transcriptome (Ke et al., 2013; Lee et al., 2014). However, this method is currently restricted by low detection efficiency and has not yet been demonstrated to detect gene expression patterns in intact tissues.

With current technology, a promising approach for spatially resolved transcriptomics consists of dissecting the specimen of interest and preparing individual sequencing libraries from the different pieces of tissue. In cases where manual dissection is difficult or impossible, such as the small embryos of typical model organisms, cryosectioning followed by RNA extraction from individual slices has been shown to provide spatial resolution along the sectioning coordinate (Combs and Eisen, 2013). However, the requirement for high amounts of input RNA for preparing RNA sequencing (RNA-seq) libraries has so far limited the use of this approach for spatially resolved transcriptomics, because large amounts of carrier RNA have to be used for library preparation, severely reducing the number of useable sequencing reads. The recent development of several RNA amplification strategies for single-cell RNA sequencing (Hashimshony et al., 2012; Islam et al., 2011, 2014; Picelli et al., 2013; Ramsköld et al., 2012; Tang et al., 2009) is now beginning to alleviate the restrictions caused by high input material requirements.

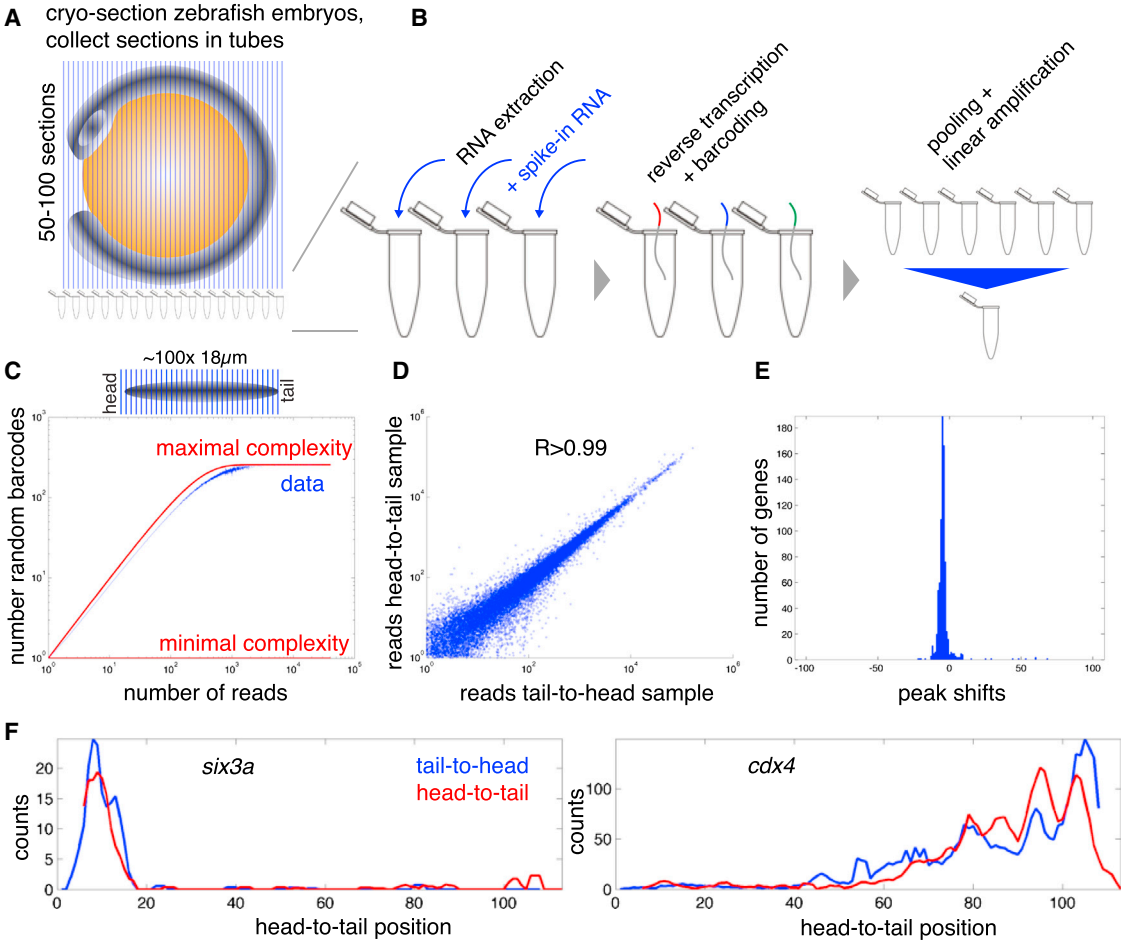


Figure 1. A Method for High Sensitivity Spatially Resolved Transcriptomics

(A) Schematic representation of segmentation stage zebrafish embryo (gray) and yolk sac (orange). Embryos were cryosectioned into 50–100 slices and sections were collected in individual Eppendorf tubes for spatially resolved transcriptomics.

(B) Summary of the experimental protocol: RNA extraction was performed for each individual section, followed by reverse transcription. The samples were then pooled and amplified by in vitro transcription (see main text and Experimental Procedures for details).

(C) Straightened zebrafish embryos at the 15 somites stage were sectioned from head to tail or vice versa into ~100 slices of 18 μm thickness. Graph shows number of different random barcodes versus number of reads for all genes and all sections. Blue data points are averages in bins of size 1. Red lines indicate theoretical boundaries determined by minimal complexity and maximal complexity libraries (Grün et al., 2014).

(D) Pearson correlation across all genes, summed over all sections for two biological replicates (sectioned tail-to-head and head-to-tail).

(E) Alignment of expression patterns for genes with detected peaks based on cross-correlation (peak calling criterion was one contiguous stretch of greater than five slices with *Z* score greater than two in each data set; expression traces were normalized to spike-ins).

(F) Examples for expression traces in both samples after alignment. *six3a* and *cdx4* are specifically expressed in the head and the tail, respectively.

See also Figure S1, Table S1, and Data S1.

Here, we present a method for spatially resolved transcriptomics based on cryosectioning that circumvents the need for carrier RNA and allows the generation of genome-wide spatial expression maps in 3D. In our approach, we cryosection individual zebrafish embryos into 50–100 thin slices, extract RNA from the individual sections, and make use of in vitro transcription for linear amplification of cDNA (Hashimshony et al., 2012) in order to minimize amplification biases (Figures 1A and 1B). By sectioning individual embryos in three different directions, we measure RNA profiles along the three main body axes. Using mathematical image reconstruction inspired by optical tomography techniques such as computed tomography, and taking into account the shape of the embryo as determined by microscopy, we then reconstruct spatial expression patterns in 3D on a transcriptome-wide level. We provide a comprehensive genome-wide 3D expression atlas at three different stages of early zebrafish development—shield stage, 10 somites stage, and 15 somites stage—and we show in a proof-of-principle experiment using mouse embryonic forelimbs that the protocol can be applied to other model organisms and to isolated tissues or organs. We demonstrate that our approach can identify spatial expression patterns in the zebrafish embryo and validate selected candidates by traditional in situ hybridization. Our method and database represent a powerful resource to identify

candidate genes expressed in any specific pattern without the need for anatomical annotation.

RESULTS

A Method for High Sensitivity Spatially Resolved Transcriptomics

Segmentation-stage zebrafish embryos are tightly wrapped around the yolk sac, with head and tail in close proximity (Figure 1A). To determine transcriptome-wide gene expression patterns along the anteroposterior axis, we removed the yolk and straightened zebrafish embryos at the 15 somites stage, allowing us to cryosection embryos from head to tail into ~100 slices of 18 μm thickness (see sketch in Figure 1C). We then extracted RNA from each individual section after adding a defined amount of synthetic spike-in RNA to control for efficiency differences during the downstream protocol (Figures S1A–S1D available online). cDNA was generated for the individual sections using barcoded primers, which allowed us to pool samples for RNA amplification (Figure 1B; Experimental Procedures; Table S1). In vitro transcription allows linear amplification of the cDNA, thereby minimizing amplification biases (Hashimshony et al., 2012). We then prepared RNA-seq libraries, sequenced ~30 million reads per embryo (50 bp paired-end), and mapped sequences to the zebrafish transcriptome (see Experimental Procedures). We detected ~23,000 different genes, with ~12,000 genes observed at more than four reads in at least two sections. We termed this procedure "tomo-seq" (the ancient Greek word "τόμος" (*tomos*) means "slice" or "section").

Because the tomo-seq protocol generates libraries primed from polyA tails of transcripts, it is possible to identify exact 3′ transcript boundaries if sequencing reads are sufficiently long to reach into the polyA regions of the library (Figure S1E). Sequencing of pooled tomo-seq libraries at 250 bp read length allowed us to substantially improve annotation of 3′ UTRs in zebrafish (see Extended Experimental Procedures; Figure S1F–S1H; Data S1). By using these improved gene annotations, we ensured that a maximum number of sequencing reads could be mapped to the transcriptome.

In addition to section-specific barcodes encoding spatial information, we used primers for reverse transcription that also contained a random barcode of 4 bp length, allowing for 4^4 = 256 different combinations. For sequencing reads mapping to the same gene, the random barcode reveals whether the reads are derived from the same molecule or from different molecules, based on whether or not the random barcode is identical (Grün et al., 2014; Kivioja et al., 2012). Comparing the number of different random barcodes to the number of reads per gene and section, we found that the complexity of our sequencing library was high, and we were not exhausting the library (Figure 1C), suggesting that additional lowly expressed genes might be detected at higher sequencing depth. We next evaluated the reproducibility of the tomo-seq technique across biological replicates. For two embryos at the 15 somites stage that were sectioned into ~100 slices in opposite directions—one from head to tail, the other from tail to head—the Pearson correlation across all genes (summed over all sections) was >0.99 (Figure 1D), demonstrating excellent reliability of the protocol.

To assess the ability of the tomo-seq protocol to determine spatial expression patterns, we then filtered the "head-to-tail" and "tail-to-head" data sets for genes exhibiting a spatially restricted peak of gene expression at any position along the anteroposterior axis. Cross-correlation of the head-to-tail and tail-to-head traces for all these genes revealed that peak positions aligned extremely well, indicating that tomo-seq allows reproducible detection of spatial expression patterns on the genome-wide scale (Figure 1E). Manual inspection of two marker genes expressed exclusively in the head or tail, *six3a* (Seo et al., 1998) and *cdx4* (Kudoh et al., 2001), further confirmed the reproducibility and precision of our protocol (Figure 1F).

Tomo-Seq Determines Gene Expression Patterns with High Spatial Resolution

We then proceeded to explore the ability of our method to identify different patterns of spatially-restricted gene expression by comparing our data to published expression patterns based on in situ hybridization data, sourced from the ZFIN expression database (Bradford et al., 2011). We found that tomo-seq correctly identified the expression patterns of selected genes (Figure S2A). Importantly, even intricate patterns such as the two stripes of *egr2b* (*krox20*) located in rhombomeres 3 and 5 were resolved correctly (Figure 2A), demonstrating the high spatial resolution of our approach. However, it is important to note that the resolution of tomo-seq is determined by section thickness, placing limits on our ability to resolve small structures such as individual somites (Figure S2B). In summary, tomo-seq allows detection of a wide range of spatial expression patterns on the genome-wide level. The full data set is provided in Table S2.

We next aimed to examine the resolution limits of our approach by comparing genes with very similar expression patterns. For this purpose, we removed the yolk, straightened and sectioned the posterior end of an embryo at the 18 somites stage, which we then sequenced at ~10-fold higher depth compared to the previous samples. We selected *mesp* genes, which are expressed in the most newly-formed somites (Cutty et al., 2012), as a test case for resolving subtle expression differences. Our data identify a strong peak for both *mespaa* and *mespab* corresponding to high expression in the newest-formed somite. Interestingly, tomo-seq further identified a small secondary peak posterior to the main peak for *mespaa*, but not for *mespab* (Figure 2Bi). We were able to confirm this predicted variation in expression domains between the two genes by traditional in situ hybridization (Figure 2Bii). This example clearly demonstrates the ability of the tomo-seq technique to resolve gene expression patterns with high sensitivity and great precision.

To explore whether tomo-seq can be applied to different model organisms and isolated organs, we decided to investigate gene expression patterns in mouse forelimbs at E10.5. At this stage, expression of *Shh* in the posterior forelimb establishes a hedgehog signaling gradient along the AP axis (Litingtung et al., 2002). We found that tomo-seq correctly reproduced spatial expression patterns of hedgehog targets and pathway components (Figure 2C). Importantly, even relatively subtle expression changes in *Gli3*$^{-/-}$ forelimbs were detected reliably,

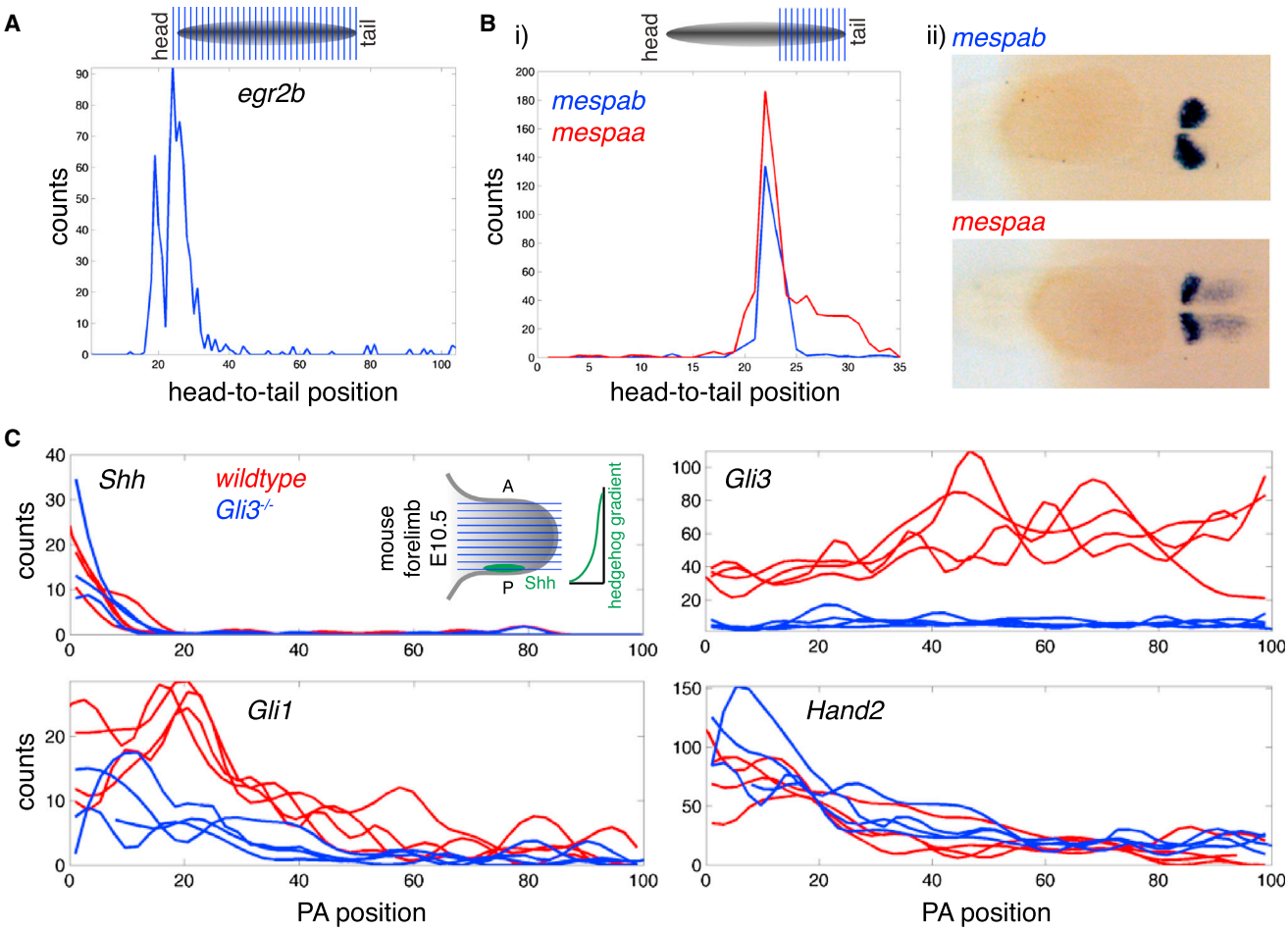


Figure 2. Tomo-Seq Determines Gene Expression Patterns with High Spatial Resolution

(A) Expression trace for *egr2b* in tail-to-head sample.

(B) Tomo-seq data for posterior end of an 18 somites stage embryo sequenced at higher depth. (i) Expression patterns of *mespab* and *mespaa* are very similar, with the exception of a shoulder on the posterior side of the main peak for *mespaa*. (ii) mRNA in situ hybridization confirms these differential expression domains (view of the tail, posterior to the right). See also Figure S2 and Table S2.

(C) Tomo-seq traces for selected genes along the posterior-to-anterior axis of E10.5 mouse forelimbs (right forelimbs of four wild-type and four *Gli3*$^{-/-}$ littermates). Forelimbs were sectioned as shown in the inset in the top left panel in order to resolve gradients of hedgehog target genes and pathway components. Tomo-seq traces and expression changes in Gli3$^{-/-}$ limbs are in full agreement with published whole-mount in situ data (Litingtung et al., 2002).

suggesting that tomo-seq enables spatially resolved differential expression analysis.

Genome-wide Determination of Genes with Similar Expression Patterns

One important advantage of our approach is that spatially resolved data are immediately available in a quantitative format, without the need for additional processing steps. As a consequence, tomo-seq data can be easily searched for genes that satisfy any desired spatial expression criteria, such as for example all genes that are coexpressed with a known marker for a specific organ or cell type. To explore this feature, we ranked all genes at the 15 somites stage by similarity to *mespab*. As a measure for expression pattern similarity, we calculated the Euclidean distance of *Z* score transformed expression traces. Figure 3A shows the expression trace for *mespab* as well as the 10 genes with most similar expression patterns, demonstrating that our approach can correctly identify the most similar patterns in the data set. Importantly, we noticed that expression patterns were gradually becoming more dissimilar to *mespab* as we descended the ranked list of similar genes. *dmrt2a*, the tenth most similar gene, was already expressed in a clearly different pattern, its expression forming a gradient with highest levels found in the tailbud, in contrast to the single somite domain observed for *mespab*. Ranking all genes by similarity to *dmrt2a*, we found that the top ten most similar genes showed significantly broader expression patterns (Figure 3B), indicating that our measure for pattern similarity is sensitive even to minor differences in gene expression patterns.

The lists of genes with expression most similar to *mespab* and *dmrt2a* consisted of genes with known and unknown expression patterns. We therefore decided to validate tomo-seq

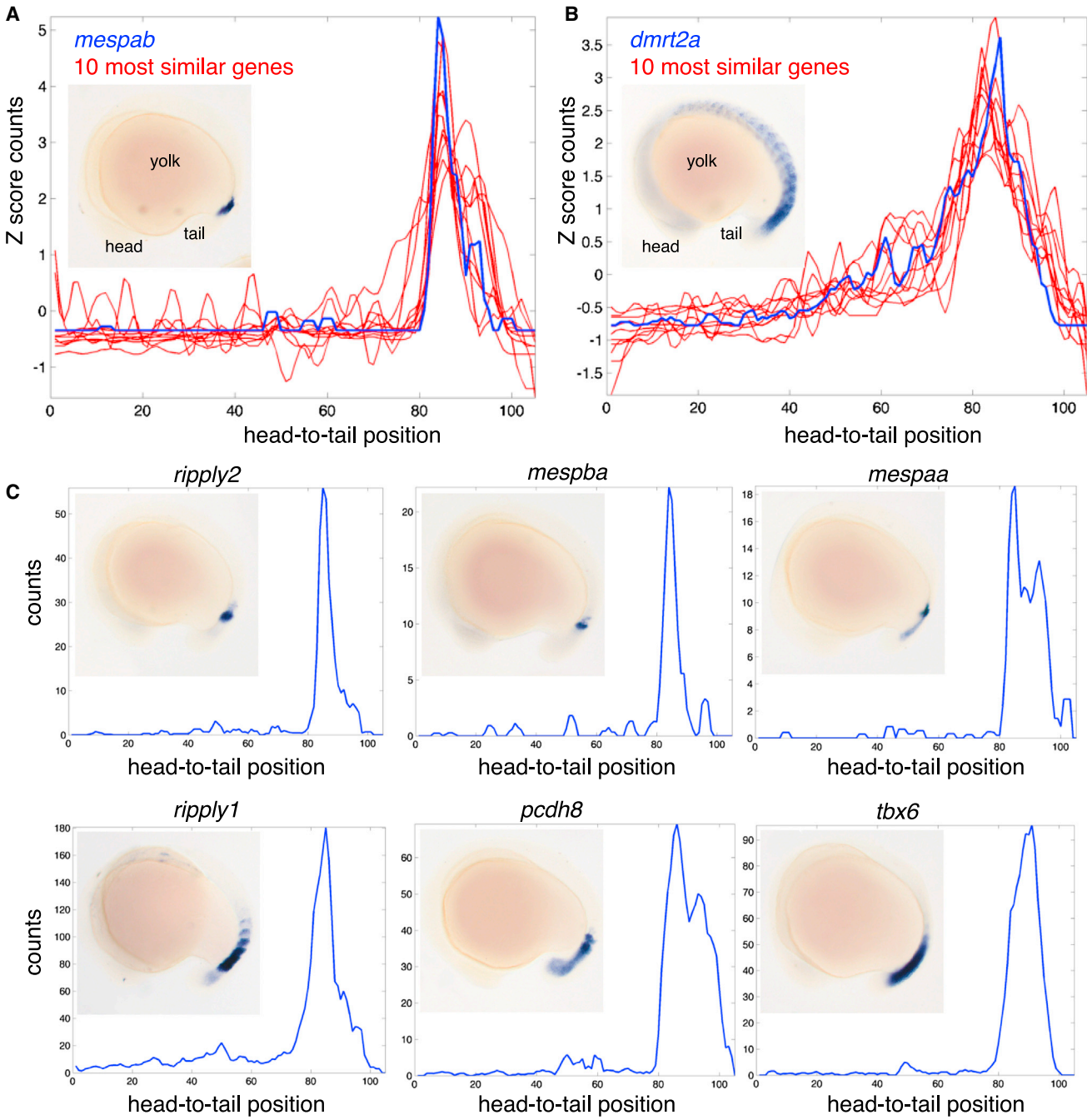


Figure 3. Genome-wide Determination of Genes with Similar Expression Patterns

(A) Tomo-seq expression patterns for *mespab* and 10 most similar genes as determined by Euclidean distance of *Z* score transformed expression traces (summed over head-to-tail and tail-to-head samples). Inset shows mRNA whole mount in situ hybridization expression pattern of *mespab*.

(B) Tomo-seq expression patterns for *dmrt2a* and ten most similar genes. Inset shows mRNA whole mount in situ hybridization expression pattern of *dmrt2a*.

(C) Tomo-seq expression traces and whole-mount in situ hybridization images for six genes that are similar to *mespab*. All in situ hybridizations are lateral views, anterior to left.

See also Figure S3.

measurements for selected genes similar to *mespab* (Figure 3C) and *dmrt2a* (Figure S3) by in situ hybridization. In situ results agreed very well with the sequencing data and even minor features such as more posterior expansion of *pcdh8* and *tbx6* expression compared to *mespab* were distinguished correctly (Figure 3C). Expression patterns of the uncharacterized genes

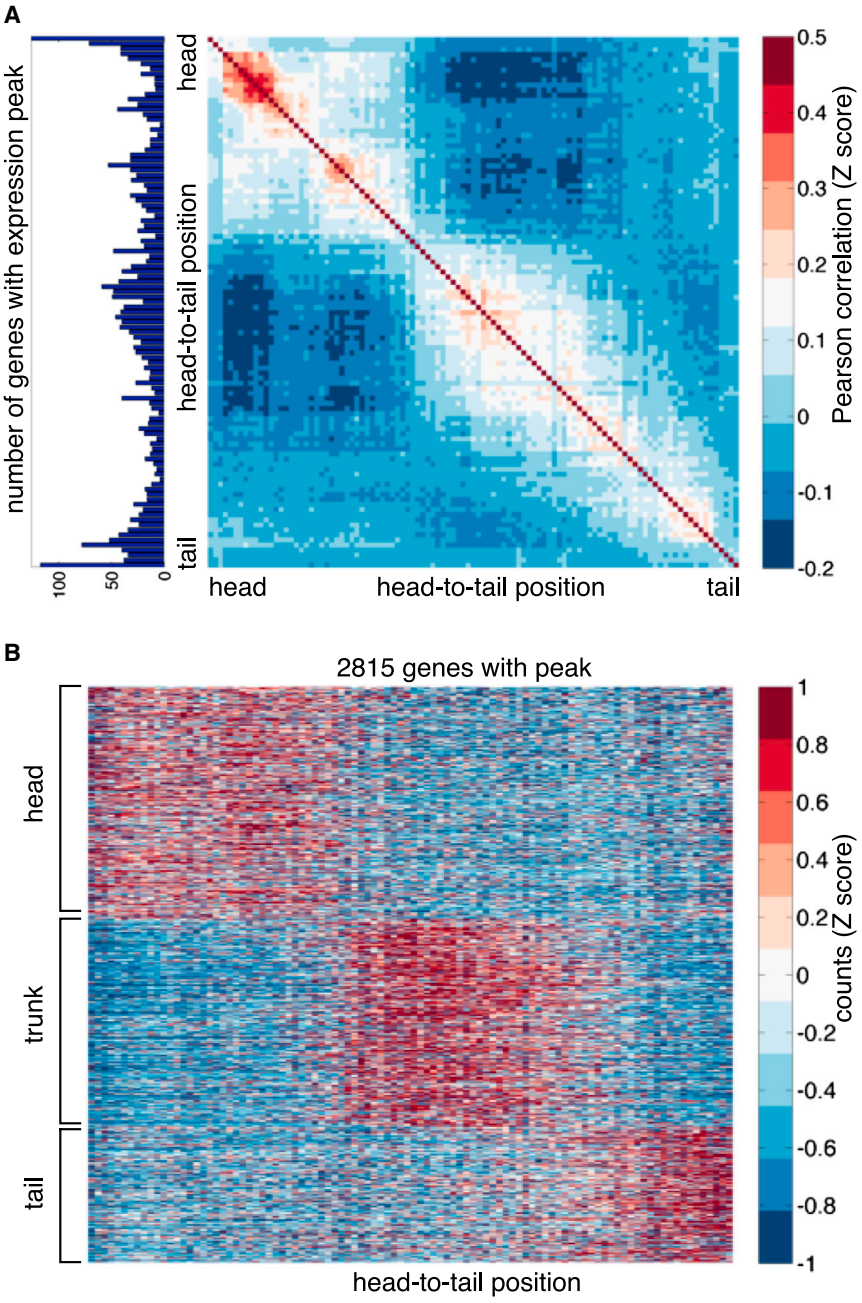


Figure 4. A Large Fraction of Genes Exhibit Spatially Patterned Expression

(A) Pairwise correlation for all sections across all genes detected at greater than four reads in greater than one section. Basic anatomical features (head, trunk, tail) can be observed as blocks of correlated sections. Plot on the left: the number of genes exhibiting a peak of expression is maximal in the regions corresponding to head, trunk, and tail. The peak calling threshold was five contiguous sections with a *Z* score of >1.

(B) K-means clustering of expression traces for all genes satisfying the peak calling conditions using Euclidean distance as metric. The number of major clusters was determined by gap statistics, and the stability of the clusters was determined by bootstrapping.

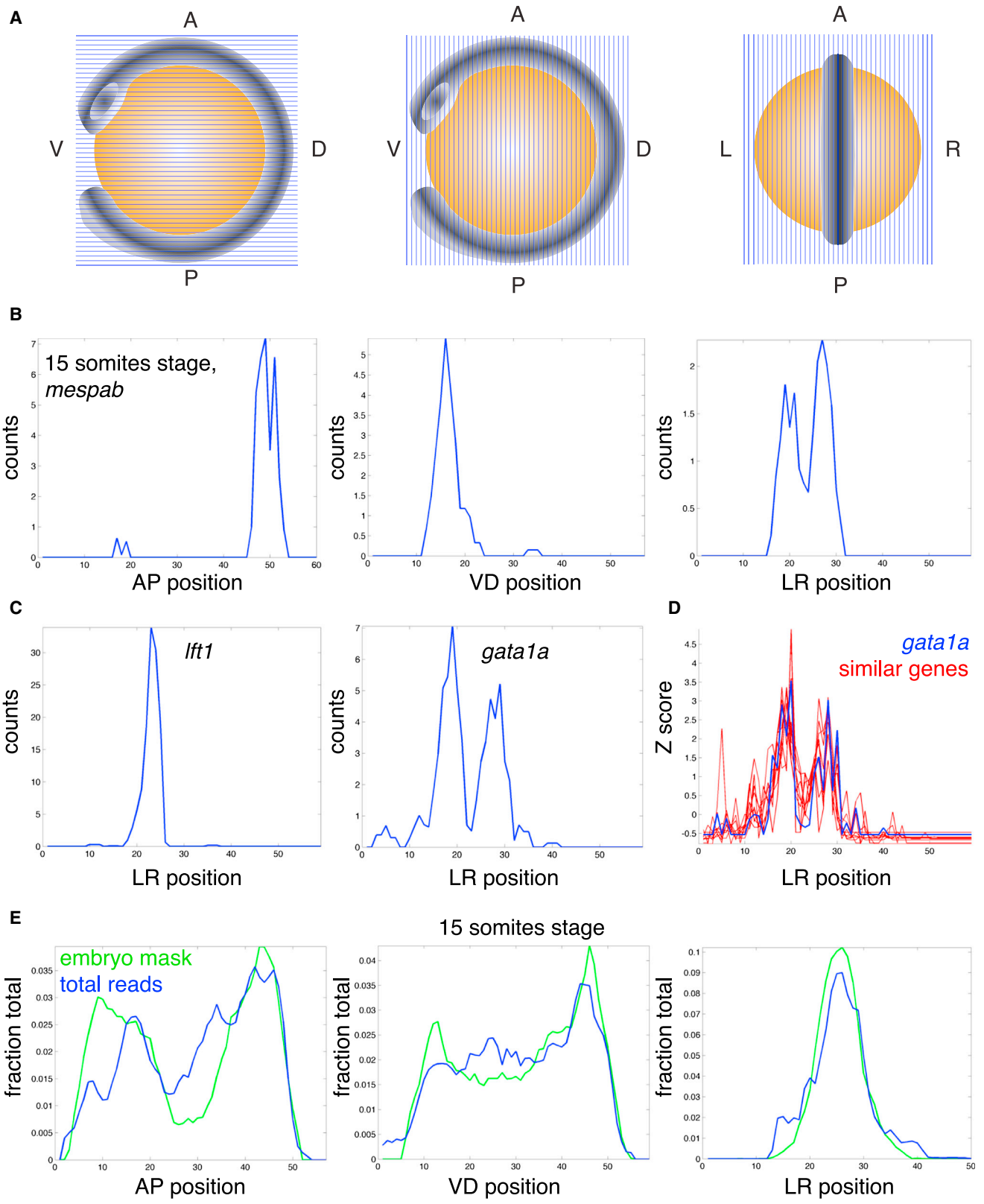
See also Figure S4 and Table S3.

si:ch211-105d18.8, *si:rp71-36n21.1* and *CU633857.1* were also detected correctly (Figure S3), proving that tomo-seq allows identification of novel genes with spatially patterned expression.

A Large Fraction of Genes Exhibit Spatially Patterned Expression

We next aimed to investigate global patterns of gene expression in our data set. To this purpose, we calculated the Pearson correlation across all genes for each pair of sections (Figure 4A). As expected, adjacent slices were generally more correlated than sections that are spaced far apart. Interestingly, however, we observed three blocks of contiguous sections that were positively correlated among each other. The spatial positions roughly corresponded to the head, the trunk and the tail of the zebrafish embryo. We also found that more genes exhibited expression peaks inside the blocks than in the transition zone (Figure 4A, left panel), suggesting that the three blocks are defined by a limited number of genes that exhibit maximal expression in either the head, the trunk or the tail. To further analyze the global spatial structure of gene expression in the zebrafish embryo, we computed a t-SNE map ("t-distributed stochastic neighbor embedding") for all genes that have an expression peak (Figure S4A). t-SNE maps project high-dimensional data onto a 2D surface while retaining distance information between individual objects (Amir et al., 2013; van der Maaten and Hinton, 2008). We found that genes peaking at similar spatial positions were grouped closely together in the t-SNE map, validating relative peak position as a measure for expression pattern similarity.

To further understand global gene expression patterns, we focused on all genes exhibiting an expression peak (~20% of the genes that were detected at greater than four reads in greater than one section). K-means clustering of genes identified three main clusters with expression maxima in the head, the trunk, and the tail, respectively (Figure 4B). Gap statistics suggested that three clusters captured most of the expression patterns (Tibshirani and Walther, 2006) (Figure S4B). However, a continued increase of the gap value at higher cluster numbers suggested that the 3 main clusters might contain a finer substructure. Hierarchical clustering of the three main clusters



(legend on next page)

indeed revealed considerable diversity of patterns within all three clusters, most notably in the head (Figure S4D). Based on assignment of marker genes to subclusters we were able to identify different anatomical structures such as forebrain, heart, somites and tail bud. The groups of genes that are specifically expressed in these organs are listed in Table S3. The genes for which we did not detect a peak showed no significant patterns after hierarchical clustering (Figure S4C), with the exception of two groups of genes that seemed to be mildly upregulated in the head and the trunk and which had narrowly failed to pass our peak detection filter. These findings show that many genes exhibit spatially structured expression, including a large number of genes with unknown function.

In order to estimate the agreement between tomo-seq and published in situ hybridization data on a global level, we randomly selected 100 genes from the head, trunk, and tail clusters, as well as 100 ubiquitously expressed genes, and compared the expression traces to the ZFIN database. We found that 131 of the 400 genes (33%) were not present in the ZFIN database, illustrating the potential for discovering genes with uncharacterized expression patterns. For the genes that were present in the ZFIN database, expression patterns agreed with tomo-seq for a large majority of genes. We found only seven genes (2%) that exhibited a clear discrepancy between the two techniques. Interestingly, for six of these seven genes (*hoxd9a*, *crygn2*, *zgc:153662*, *pgam2*, *im:7136729*, *negr1*) ZFIN reported no expression or ubiquitous expression. Furthermore, expression data for these six genes are based on a single submission. Hence, it is possible that the in situ probes were not working. For the remaining gene, *hoxb5a*, ZFIN reports a major peak in the tail and a minor peak in the trunk. According to our tomo-seq data set, however, the peak in the trunk is higher than the peak in the tail. We speculate that this discrepancy might potentially have been caused by minor differences in embryo staging. In summary, tomo-seq data agree extremely well with in situ data.

Tomo-Seq at Three Developmental Time Points along the Main Body Axes

The above analyses demonstrate that tomo-seq offers exquisite spatial resolution of expression patterns on the transcriptome-wide level. However, studying straightened embryos from which the yolk has been removed can potentially lead to artifacts because some lateral tissue might be lost when detaching the yolk. To address this issue and extend our analysis to all three main body axes, we set out to section and sequence entire embryos including the yolk sac along each body axis using sections of 18 μm thickness (see sketch in Figure 5A). To further expand our data set, we also decided to extend our approach to three different developmental stages—shield stage, the 10 somites stage, and the 15 somites stage—creating a large resource of transcriptome-wide expression patterns.

We validated our approach using entire embryos by manually comparing tomo-seq patterns of selected genes along all three axes to published expression patterns (Bradford et al., 2011) (Figures 5B, S5A, and S5B). Patterns were highly reproducible across replicates (not shown) and agreed very well with published in situ hybridization mRNA expression data. Because sectioning along the left-right axis is particularly challenging due to the small expanse of tissue along this dimension, we aimed at further validating the left-right data set at the 15 somites stage. Our data correctly showed *lft1* expression in the midline and *gata1a* expression in more lateral positions (Figure 5C). Ranking genes by similarity to *gata1a* in the left-right data set (Figure 5D), we noticed that genes with related function were enriched among the genes with the most similar spatial expression patterns. *gata1a* has been shown to determine myeloid-erythroid progenitor cell fate in zebrafish (Lyons et al., 2002). Functional annotation analysis using DAVID (Huang et al., 2009) revealed enrichment of genes with a known role in blood formation and vasculogenesis, such as *alas2*, *hbae3*, *hbbe3*, *hbae1*, *znfl2a*, *yrk*, *foxc1b*, and *egfl7* among the 50 genes with most similar expression to *gata1a* along the left-right axis. This finding provides another confirmation for the high accuracy and sensitivity of our technique. The full data set (three time points, sectioned along three axes in duplicates) is available in Table S4.

RNA Tomography Allows Construction of a Spatially Resolved Transcriptome-wide Atlas of Gene Expression in 3D

Because of the curved shape of the zebrafish embryo at somitogenesis stages, tomo-seq data obtained from entire embryos are harder to interpret compared to straightened embryos from which the yolk has been removed. For this reason we sought to image entire zebrafish embryos in 3D by using selective plane illumination microscopy (SPIM) (Huisken et al., 2004) in order to allow mapping of gene expression patterns onto the image (Figures S5C and S5D). We found that the total number of reads per section agreed well with the slice volume as determined by microscopy (Figure 5E), suggesting that the number of reads per section is approximately proportional to the number of cells per slice. This observation corroborated the quantitative nature of our method and allowed aligning of sequencing data and microscopy data by cross-correlation. Importantly, the agreement between microscopy and sequencing data also indicated that there were no major amounts of mRNA left in the yolk at the developmental stages we examined.

Figure 5. Tomo-Seq at Three Developmental Time Points along the Main Body Axes

(A) Schematic of sectioning approach for 3D analysis. Intact embryos including the yolk sac were sectioned along the three main body axes. Yolk sac, orange; embryo, gray; A, anterior; P, posterior; V, ventral; D, dorsal; L, left; R, right.

(B) Expression patterns for *mespab* at 15 somites stage along AP, VD, and LR axes.

(C) Expression trace of *lft1* and *gata1a* along LR axis in 15 somites stage embryo.

(D) Top ten genes with expression patterns similar to *gata1a* along LR axis at 15 somites stage.

(E) Total reads after spike-in normalization and slice volume as determined by microscopy as a function of AP, VD, and LR position at 15 somites stage.

See also Figure S5 and Table S4.

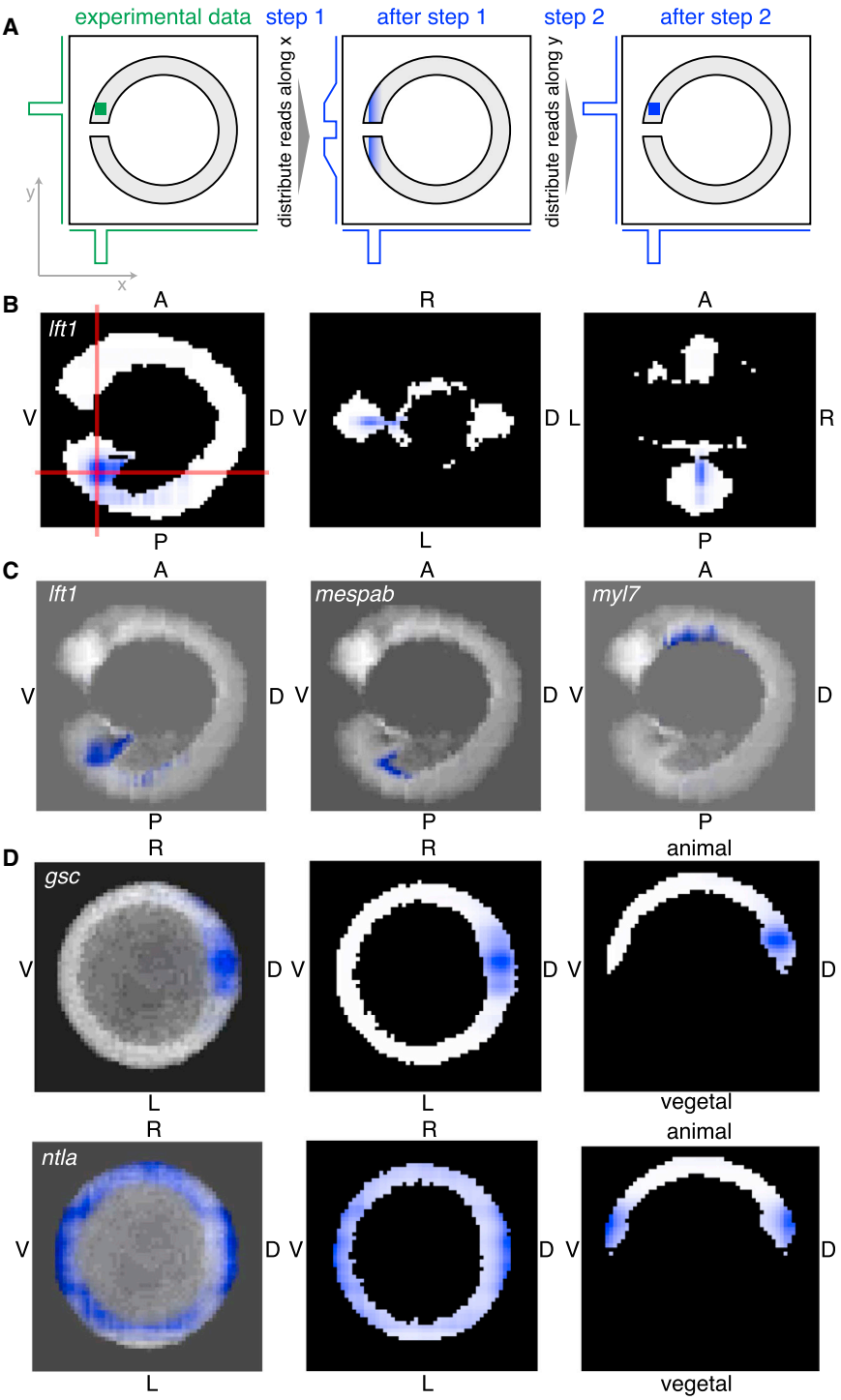


Figure 6. RNA Tomography Allows Construction of a Spatially Resolved Transcriptome-wide Atlas of Gene Expression in 3D

(A) Mathematical approach for 3D image reconstruction based on 1D projections. The expression pattern of a hypothetical gene is shown as a green square in the left panel. The middle and right panel show reconstructed expression patterns after sequential optimization steps in blue. Measured 1D tomo-seq traces are depicted as green lines, projections of computationally reconstructed image on x and y axis are shown as blue lines to the left of and under the panels. The steps shown here were iterated 100 times for optimal image reconstruction.

(B–D) Normalized expression patterns of selected genes (blue) after 3D image reconstruction. The shape of the embryos is shown in black and white. (B) Left: expression of *lft1* at 15 somites stage in a plane perpendicular to LR axis through the middle of the embryo. Middle and right: *lft1* expression in the two planes indicated by red lines in the left panel. (C) Projection images (summed over LR axis) for *lft1*, *mespab*, and *myl7* expression patterns at 15 somites stage. (D) Left: projection along animal-vegetal axis for *gsc* and *ntla* at shield stage. Middle: expression in a plane perpendicular to animal-vegetal axis at the level of the margin. Right: expression in a plane perpendicular to LR axis through the middle of the embryo.

See also Figure S6 and Data S2.

To facilitate an intuitive understanding of the measured expression patterns, we decided to combine the data sets measured along the three main body axes and the shape of the embryo determined by microscopy into a composite 3D image in which the projection along each axis matches the experimental data. Our approach for mathematical image reconstruction is based on sequential image optimization along the different sectioning axes using iterative proportional fitting (Fienberg, 1970) (Figure 6A). For a given gene, we first distributed the sequencing reads in the embryo such that the projection along the x axis matched the RNA tomography expression trace for the dorsoventral axis. At this stage, the reads were distributed homogenously in the embryo along the y and z axis. In the second step, we redistributed the reads along the y axis in order to also agree with the RNA tomography expression traces along the anteroposterior axis, while leaving the x position of each read unchanged. The same approach was then repeated along the z axis to optimize the read distribution along the left-right axis. This algorithm was iterated in order to optimize image reconstruction (see Extended Experimental Procedures for details). The resulting images consist of 3D expression patterns in the embryo and can be browsed by scrolling through the individual planes. We termed this approach for genome-wide 3D image reconstruction "RNA tomography," inspired by optical tomography methods such as computed tomography. Such methods determine the original 3D image based on projections measured under different orientations, similar to our approach (Gordon et al., 1970).

As shown in Figures 6B–6D, the resulting images reproduced spatially restricted expression patterns correctly in 3D: *lft1* is expressed exclusively in the posterior part of the 15 somites stage embryo while, along the left-right axis, it is restricted to the midline (Wang and Yost, 2008) (Figure 6B). Similarly, expression of *mespab* and *myl7* at the 15 somites stage, and of *dand5*, *egr2b* and *cdx4* at the ten somites stage, was placed correctly in the embryo (Bradford et al., 2011) (Figures 6C and S6A). RNA tomography is particularly suited for shield stage embryos, because the morphology of zebrafish embryos at that stage—the embryo forms a hemispherical "cap" on the yolk sac (see sketch in Figure S5B)—does not allow yolk removal and embryo straightening. Hence, 1D analysis would yield a very incomplete representation of expression patterns at shield stage. In Figures 6D and S6B we demonstrate that RNA tomography correctly identified known expression patterns for selected genes at the shield stage: *gsc* is restricted to the organizer—a small field of mesodermal cells at the dorsal side of the embryo, close to the margin. *ntla* is expressed round the entire margin, while *bmp4* forms a ventral-to-dorsal gradient and is expressed weakly in the dorsal organizer. These examples suggest that RNA tomography can correctly identify a range of different expression patterns at different developmental stages.

It is important to note that RNA tomography can potentially give rise to image reconstruction artifacts: As shown in Figure S6C, genes that are expressed in more than one contiguous region can lead to ambiguous solutions. This effect is caused by the fact that the number of pixels is larger than the number of data points. With ~50 sections per axis, we obtain 3 × 50 = 150 independent measurements, whereas the number of pixels is 50^3 = 125,000. As a consequence, the system is underconstrained, and solutions to the image reconstruction problem are in general not unique. The iterative proportional fitting algorithm aims to identify the solution with maximal entropy, ensuring that the least amount of unjustified information is introduced into the reconstructed image. Importantly, by taking the 3D shape of the embryo into account, we could effectively eliminate all pixels outside the embryo, reducing the number of pixels by ~90%. Hence, using microscopy data strongly alleviates problems caused by artifacts. Furthermore, for many practical applications genes with simple expression patterns, such as expression limited to a specific organ or tissue, are particularly relevant. Such genes do not only constitute the majority of genes with spatially heterogeneous expression (Figures 4B and S4C), but are also less prone to causing artifacts. However, for genes with complex expression patterns reconstruction artifacts cannot be fully ruled out.

Determining Genes with Similar Expression Patterns in 3D

The ability to identify novel genes with specific expression patterns is arguably the most important application of spatially resolved transcriptomics in developmental biology. We hence set out to extend our approach for identifying genes with similar expression patterns to 3D, focusing on shield stage embryos for which 1D analysis is difficult to apply. We ranked all genes by similarity to *gsc*, a marker for the zebrafish organizer (Stachel et al., 1993). Similar to the 1D approach, we computed the Euclidean distance of *Z* score transformed expression patterns across all pixels inside the embryo as a measure for similarity. 3D expression patterns of the 15 genes that are most similar to *gsc* are shown in Movies S1 and S2. Manual inspection of the ranked list revealed a large number of previously unknown genes in addition to well-known markers for the organizer among the most similar genes. We selected 12 of these novel genes as candidates for validation by in situ hybridization (Figures 7A, 7B, and S7A).

Expression of *chpfa*, *insb*, *net1*, *tbr1b*, and *ripply1* was clearly restricted to the organizer, providing an impressive confirmation of RNA tomography as a tool to identify candidate genes with spatially restricted expression in 3D (Figures 7B and S7A). Strikingly, the transcription factor *uncx4.1* was only expressed in a very small number of cells in the organizer.

magi1b, which shows homogenous staining in situ experiments, is the only gene for which RNA tomography data do not agree with in situ hybridization (Figure S7A). For two other genes, *CU074314.1* and *si:dkey-13n23.3*, in situ probes confirm upregulation in the organizer, but in contrast to tomo-seq also indicate basal expression throughout the entire embryo. It is unclear whether these disagreements reflect artifacts of RNA tomography or unspecific staining by in situ hybridization.

In order to explore whether our technique is equally successful in identifying genes with similar 3D expression patterns at the two later developmental stages, we ranked all genes by similarity to the heart marker *myl7* (Yelon et al., 2000) at the 15 somites stage. The eight most similar genes *cmlc1*, *vmhc*, *tnnt2a*, *tbx20*, *tnni1b*, *atp2a2a*, *fhl2a*, and *tnnc1a* (Figure S7B) all have a heart-related function (Bradford et al., 2011). Searching for genes with expression patterns similar to *dand5* at the ten somites stage, a gene expressed exclusively around the Kupffer's vesicle (the zebrafish equivalent of the node) (Hashimoto et al., 2004), we identified a mixture of known and unknown genes expressed in the Kupffer's vesicle or, more broadly, in the tip of the tail (Figure S7C). These findings provide further robust validation of the accuracy and predictive power of genome-wide RNA tomography.

DISCUSSION

A Sensitive and Reliable Method for Spatially Resolved Transcriptomics

The tomo-seq approach for spatially resolved transcriptomics consists of three steps: cryosectioning the tissue of interest, reverse transcription and barcoding, and amplification of the pooled cDNA by in vitro transcription. Tomo-seq can be applied to whole embryos or isolated organs of different model organisms. We demonstrated the high spatial resolution and unprecedented sensitivity of our approach by validating expression patterns of selected genes with in situ hybridization. We have been able to extend our approach to 3D by sectioning embryos in three different orientations and combining this data set with microscopy data and mathematical image reconstruction. Furthermore, we introduced several mathematical approaches for unbiased analysis of tomo-seq data sets that can be used for identifying global patterns of gene expression or for finding genes that are specifically expressed in a region of interest.

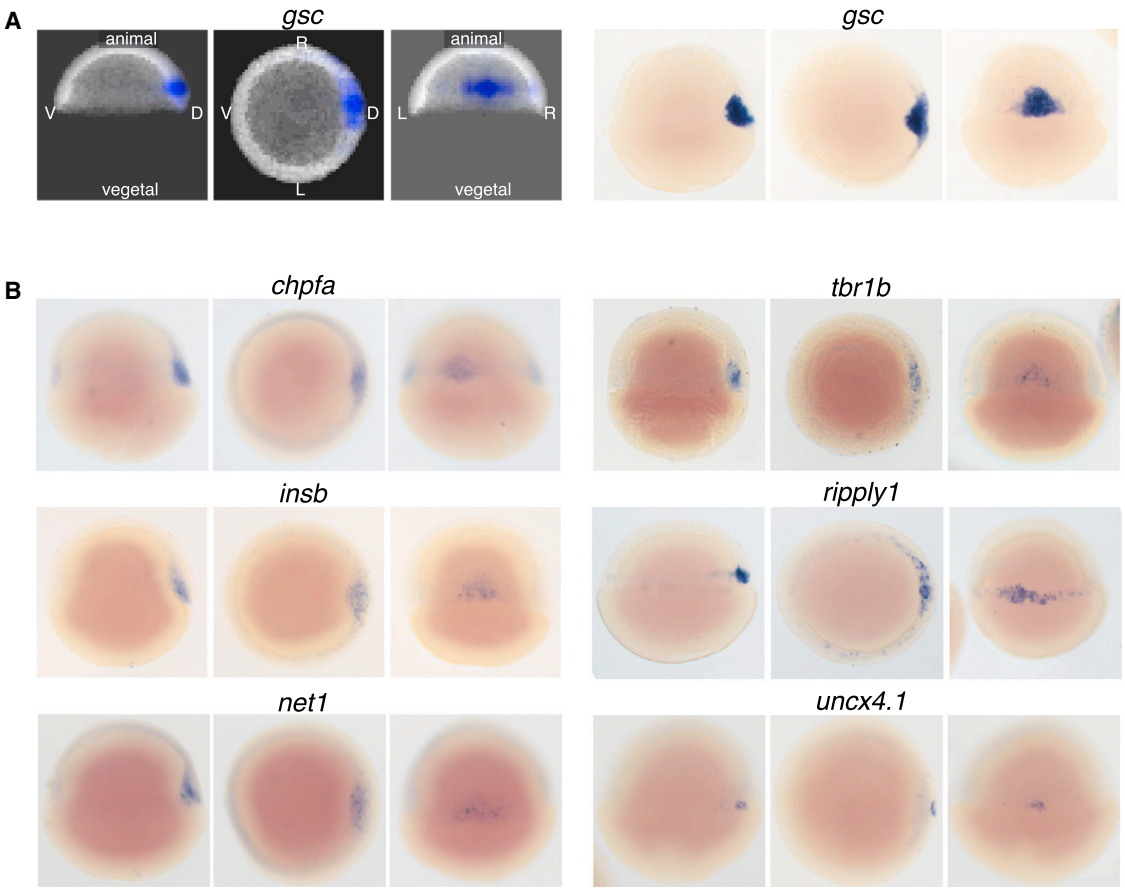


Figure 7. Determining Genes with Similar Expression Patterns in 3D

(A) RNA tomography data and in situ validation for *gsc*. Images show projections of reconstructed expression patterns along different body axes of shield stage embryos and microscopy images taken in the corresponding orientations. Left: lateral view, visualized from left. Middle: animal pole view. Right: dorsal view.

(B) Validation of novel genes with inferred expression patterns similar to *gsc* at shield stage. Panels show microscopy images along different body axes of shield stage embryos.

See also Figure S7 and Movies S1 and S2.

Tomo-seq is easy to implement because it is based on standard laboratory equipment and commercially available reagents and is hence broadly applicable to whole embryos or isolated organs of various model organisms. Many questions in developmental biology, such as pattern formation along specific body axes, are essentially 1D, which allows using the simplest form of our method without computational 3D reconstruction. 3D image reconstruction is more demanding than 1D analysis because three embryos have to be staged and oriented with high precision in order to obtain a correct composite image. Cryotome drift and damage during dissection are also potential causes for tomo-seq artifacts and should be minimized as much as possible.

For large embryos, SPIM or similar techniques are necessary for determining the shape of the embryo because they alleviate the detrimental effects of light scattering. 3D RNA tomography can be a cost-effective approach for determining complex spatial patterns because the number of data points (and hence the number of individual reverse transcription reactions) can be reduced below the number of pixels in the final image. In order to minimize costs and reduce the risk of image reconstruction artifacts, we recommend focusing on isolated organs rather than whole embryos whenever possible. Increasing the sequencing depth will reduce Poissonian sampling noise and hence allow more reliable quantification of lowly expressed genes.

While automated imaging-based techniques have the potential to interrogate large portions of the transcriptome (Battich et al., 2013; Thisse and Thisse, 2008), such approaches typically measure only one gene per sample. In many cases, in particular when working with vertebrate mutants where sample numbers can be small, the number of genes that can be screened is therefore limited. Hence, measuring the entire transcriptome in a single sample is an important advantage of our technique. Because tomo-seq libraries can be used to improve 3′ UTR annotations (Figure S1E–S1H), our approach also allows studying organisms with incomplete gene annotations. Furthermore, our method will allow researchers to determine expression differences between individual embryos on a genome-wide level, paving the way to a deeper understanding of developmental plasticity and phenotypic variation.

A Genome-wide 3D Resource for Spatial Expression Patterns in the Early Zebrafish Embryo

We determined genome-wide expression patterns at three developmental stages (shield stage, 10 somites stage, 15 somites stage), providing a resource that can be used to select candidate genes for future studies. By validating novel genes, we also demonstrated that our data set may be used to identify genes that are expressed in a spatially restricted manner.

Large in situ hybridization screens have to date provided the scientific community with a valuable database of gene expression for many genes. While such databases are searchable for tissue of interest, extensive and accurate manual annotation of embryonic structures is relied on in order to reveal whether or not a gene is expressed in the relevant tissue. Hence, one crucial advantage of our approach lies in the ability to easily search for genes displaying specific expression patterns without the need for manual anatomical annotation. In addition, in situ hybridization may not be sensitive enough to pick up genes expressed in low copy numbers and is dependent on probe efficiency and protocol optimization.

It is important to keep in mind that tomo-seq data offer lower spatial resolution than microscopy-based techniques, and reconstructed 3D images can potentially contain artifacts. For this reason, we believe that RNA tomography should not be considered a replacement for in situ protocols, but rather as a method to generate lists of candidate genes. In fact, we recommend validating selected candidates by whole-mount or section in situ hybridization. Tissues composed of different cell types may also benefit from methods with higher spatial resolution such as single-molecule RNA FISH (Lyubimova et al., 2013; Raj et al., 2008).

Because the number of detected reads should be proportional to the number of mRNA molecules, tomo-seq is a very quantitative technique that can be used to compare expression levels of different genes. This is a crucial advantage in comparison to in situ hybridization, where different probes often have different efficiencies and staining intensity is dependent on the exact hybridization protocol. Furthermore, tomo-seq has a high dynamic range and can correctly identify expression patterns of lowly and highly expressed genes at the same time, unlike microscopy-based approaches that typically exhibit detection thresholds and saturation effects.

Our analysis reveals that ~20% of all expressed zebrafish genes show spatially restricted expression, including a large number of genes with uncharacterized function. For such genes, spatial information can be an important cue regarding biological function. Recent publications suggest that a considerable number of regulators of spatial patterning during embryonic development might still be unknown (Pauli et al., 2014). We expect that RNA tomography will greatly accelerate the search for novel regulators by identifying their zone of expression, and we anticipate that spatially resolved transcriptomics will emerge as a powerful method to unravel the design principles of developmental pattern formation. Moreover, we expect that RNA tomography will facilitate genome-wide comparison of patterning mechanisms across species, including human embryos, yielding fascinating novel insights into the mechanisms underlying vertebrate development.

EXPERIMENTAL PROCEDURES

Preparation of Sequencing Libraries

Live *TL* embryos were embedded in Jung tissue freezing medium (Leica), oriented and rapidly frozen on dry ice, and stored at −80°C prior to cryosectioning. Embedded embryos were cryosectioned at 18 μm thickness, collected in Eppendorf LoBind tubes, and immediately transferred to dry ice. RNA was extracted from the individual sections using TRIzol reagent (Ambion) according to the manufacturer's manual. After RNA extraction, pellets were resuspended with barcoded primers. Primers consisted of a 24 bp polyT stretch, a 4 bp random barcode, a unique 8 bp section-specific barcode, the 5′ Illumina adaptor (as used in the TruSeq small RNA kit) and a T7 promoter for in vitro transcription (Grün et al., 2014). The RNA samples were then reverse transcribed, pooled, and in vitro transcribed for linear amplification with the MessageAmpII kit (Ambion) according to the CEL-seq protocol (Hashimshony et al., 2012). Illumina sequencing libraries were prepared with the TruSeq small RNA sample prep kit (Illumina) and sequenced paired-end at 50 bp read length on HiSeq. We determined library complexity based on random barcodes (Figure 1C) in order to prevent oversequencing. With the exception of Figure 2B, random barcodes were not used as unique molecular identifiers (Kivioja et al., 2012). See Extended Experimental Procedures for a detailed experimental protocol. Primer sequences are provided in Table S1.

All experiments were performed in biological replicates and yielded reproducible results. Sectioning experiments were repeated in the opposite direction (e.g., A → P and P → A) to rule out potential artifacts caused by carry-over of RNA to subsequent sections on the blade. Replicates along the same sectioning coordinate were aligned and averaged for 3D image reconstruction. All studies involving vertebrate animals were performed with institutional approval in compliance with institutional guidelines.

Data Analysis

Paired-end reads obtained by Illumina sequencing were aligned to the transcriptome using BWA (Li and Durbin, 2010). The 5′ mate of each pair was mapped to the improved gene models described in the Extended Experimental Procedures, discarding all reads that mapped equally well to multiple loci. The 3′ mate was used for barcode information.

Read counts were normalized to total counts per section or by linear fits to spike-in RNA input/output plots (Figures S1A–S1D), depending on whether or not we wanted to retain differences in total reads per section caused by embryo geometry. The data were then renormalized to the median of total reads across sections in order to ensure that count numbers roughly corresponded to the number of mapped reads. For the 1D analyses described in Figures 2, 3, and 4 we used total read normalization in order to prevent biases caused by embryo geometry. For the 3D analyses shown in Figures 5, 6, and 7, on the other hand, we preferred spike-in normalization in order to be able to align microscopy data to sequencing data by cross-correlation based on total read profiles (Figure 5E). Expression traces were smoothed with a moving average filter with a width of three data points.

All data analysis including normalization, alignment, 3D image reconstruction, and identification of expression patterns was performed in MATLAB (MathWorks) using custom-written code. Our mathematical approach for image reconstruction based on iterative proportional fitting (Fienberg, 1970) is described in more detail in Data S2, and the corresponding MATLAB script is provided in the Extended Experimental Procedures.

ACCESSION NUMBERS

RNA-seq data are deposited on Gene Expression Omnibus, accession number GSE59873. The analyzed data set is available on our web page http://zebrafish.genomes.nl/tomoseq/. The web interface allows displaying data traces for specific genes, identifying related genes, and browsing through individual planes of 3D images.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, two data files, four tables, and two movies and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.09.038.

AUTHOR CONTRIBUTIONS

J.P.J. and A.v.O. conceived and designed the project with input from E.S.N. and J.B. J.P.J. developed the protocol, did tomo-seq experiments, and analyzed the data. E.S.N. prepared zebrafish samples and performed in situ hybridizations. V.G. and E.B. improved gene annotations. E.B. created web interface. K.A.P. and A.P.M. prepared mouse samples. G.S. and J.H. provided SPIM data. J.P.J. wrote the manuscript with input from all other authors. All authors discussed and interpreted results.

ACKNOWLEDGMENTS

We thank D. Grün and B. Spanjaard for help with computational analysis. This work was supported by a European Research Council Advanced grant (ERC-AdG 294325-GeneNoiseControl) and by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) Vidi and Vici awards. Work in A.P.M.'s laboratory was supported by a grant from the NIH (NS033642).

Received: May 20, 2014
Revised: July 24, 2014
Accepted: September 18, 2014
Published: October 23, 2014

REFERENCES

Amir, A.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat. Biotechnol. *31*, 545–552.

Battich, N., Stoeger, T., and Pelkmans, L. (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution. Nat. Methods *10*, 1127–1133.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science *302*, 1956–1960.

Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Howe, D.G., Knight, J., Mani, P., Martin, R., Moxon, S.A.T., et al. (2011). ZFIN: enhancements and updates to the Zebrafish Model Organism Database. Nucleic Acids Res. *39*, D822–D829.

Combs, P.A., and Eisen, M.B. (2013). Sequencing mRNA from cryo-sliced Drosophila embryos to determine genome-wide spatial patterns of gene expression. PLoS ONE *8*, e71820.

Cutty, S.J., Fior, R., Henriques, P.M., Saúde, L., and Wardle, F.C. (2012). Identification and expression analysis of two novel members of the Mesp family in zebrafish. Int. J. Dev. Biol. *56*, 285–294.

Fienberg, S.E. (1970). An iterative procedure for estimation in contingency tables. Ann. Math. Stat. *41*, 907–917.

Fowlkes, C.C., Hendriks, C.L.L., Keränen, S.V.E., Weber, G.H., Rübel, O., Huang, M.-Y., Chatoor, S., DePace, A.H., Simirenko, L., Henriquez, C., et al. (2008). A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm. Cell *133*, 364–374.

Geffers, L., Herrmann, B., and Eichele, G. (2012). Web-based digital gene expression atlases for the mouse. Mamm. Genome *23*, 525–538.

Gordon, R., Bender, R., and Herman, G.T. (1970). Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. J. Theor. Biol. *29*, 471–481.

Gray, P.A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.-I., Tsung, E.F., Cai, Z., et al. (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. Science *306*, 2255–2257.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640.

Hashimoto, H., Rebagliati, M., Ahmad, N., Muraoka, O., Kurokawa, T., Hibi, M., and Suzuki, T. (2004). The Cerberus/Dan-family protein Charon is a negative regulator of Nodal signaling during left-right patterning in zebrafish. Development *131*, 1741–1753.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Reports *2*, 666–673.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J., and Stelzer, E.H.K. (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. Science *305*, 1007–1009.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods *10*, 857–860.

Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2012). Counting absolute numbers of molecules using unique molecular identifiers. Nat. Methods *9*, 72–74.

Kudoh, T., Tsang, M., Hukriede, N.A., Chen, X., Dedekian, M., Clarke, C.J., Kiang, A., Schultz, S., Epstein, J.A., Toyama, R., and Dawid, I.B. (2001). A gene expression screen in zebrafish embryogenesis. Genome Res. *11*, 1979–1987.

Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P., and Krause, H.M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. Cell *131*, 174–187.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360–1363.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Litingtung, Y., Dahn, R.D., Li, Y., Fallon, J.F., and Chiang, C. (2002). Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. Nature *418*, 979–983.

Lyons, S.E., Lawson, N.D., Lei, L., Bennett, P.E., Weinstein, B.M., and Liu, P.P. (2002). A nonsense mutation in zebrafish gata1 causes the bloodless phenotype in vlad tepes. Proc. Natl. Acad. Sci. USA *99*, 5454–5459.

Lyubimova, A., Itzkovitz, S., Junker, J.P., Fan, Z.P., Wu, X., and van Oudenaarden, A. (2013). Single-molecule mRNA detection and counting in mammalian tissue. Nat. Protoc. *8*, 1743–1758.

Pauli, A., Norris, M.L., Valen, E., Chew, G.-L., Gagnon, J.A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., et al. (2014). Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science *343*, 1248636.

Picelli, S., Björklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Seo, H.C., Drivenes, Ellingsen, S., and Fjose, A. (1998). Expression of two zebrafish homologues of the murine Six3 gene demarcates the initial eye primordia. Mech. Dev. *73*, 45–57.

Stachel, S.E., Grunwald, D.J., and Myers, P.Z. (1993). Lithium perturbation and goosecoid expression identify a dorsal specification pathway in the pregastrula zebrafish. Development *117*, 1261–1274.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. Nat. Protoc. *3*, 59–69.

Thisse, B., Heyer, V., Lux, A., Alunni, V., Degrave, A., Seiliez, I., Kirchner, J., Parkhill, J.-P., and Thisse, C. (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. Methods Cell Biol. *77*, 505–519.

Tibshirani, R., and Walther, G. (2006). Estimating the number of clusters in a data set via the gap statistic. J. R. Statist. Soc B *63*, 1–1411–4233.

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Wang, X., and Yost, H.J. (2008). Initiation and propagation of posterior to anterior (PA) waves in zebrafish left-right development. Dev. Dyn. *237*, 3640–3647.

Yelon, D., Ticho, B., Halpern, M.E., Ruvinsky, I., Ho, R.K., Silver, L.M., and Stainier, D.Y. (2000). The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. Development *127*, 2573–2582.

Yu, J., Valerius, M.T., Duah, M., Staser, K., Hansard, J.K., Guo, J.J., McMahon, J., Vaughan, J., Faria, D., Georgas, K., et al. (2012). Identification of molecular compartments and genetic circuitry in the developing mammalian kidney. Development *139*, 1863–1873.