The *Drosophila* embryo at single-cell transcriptome resolution

Nikos Karaiskos,1,∗ Philipp Wahle,2,∗ Jonathan Alles,1 Anastasiya Boltengagen,1 Salah Ayoub,1 Claudia Kipar,2 Christine Kocks,1 Nikolaus Rajewsky,1† Robert P. Zinzen2†

1Systems Biology of Gene Regulatory Elements, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany. 2Systems Biology of Neural Tissue Differentiation, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125 Berlin, Germany.

†Corresponding author. Email: nikolaus.rajewsky@mdc-berlin.de (N.R.); robert.zinzen@mdc-berlin.de (R.P.Z.)

By the onset of morphogenesis, *Drosophila* embryos consist of about 6000 cells that express distinct gene combinations. Here, we used single-cell sequencing of precisely staged embryos and devised DistMap, a computational mapping strategy to reconstruct the embryo and to predict spatial gene expression approaching single-cell resolution. We produce a virtual embryo with about 8000 expressed genes per cell. Our interactive “*Drosophila*-Virtual-Expression-eXplorer” (DVEX) database generates three-dimensional virtual in situ hybridizations and computes gene expression gradients. We used DVEX to uncover patterned expression of transcription factors and long noncoding RNAs, as well as signaling pathway components. Spatial regulation of Hippo signaling during early embryogenesis suggests a mechanism for establishing asynchronous cell proliferation. Our approach is suitable to generate transcriptomic blueprints for other complex tissues.

Intricate gene regulatory networks produce and maintain complex assemblies of specialized cells such as tissues and organs. To unravel the underlying gene expression dynamics, significant efforts have been made to compare tissue-specific materials (1–4). Cell culture often constitutes a poor proxy for in vivo complexity, and dissected tissues are comprised of heterogeneous cell populations (3–10). An alternative is isolation of specific cell types via cell sorting (11–14); however, pooled cells obscure heterogeneity and expression in rare populations of cells may not be detectable. Furthermore, transcriptional relationships at the single-cell level, such as exclusivity and concomitancy of expression of groups of genes, cannot be distilled. This restricts our ability to infer gene regulatory relationships, to predict what functional roles individual cells play and how they integrate with their spatial environment. With the advent of single-cell expression profiling, it has become possible to assess the transcriptomic landscape of complex cell mixtures with single-cell resolution, thereby allowing insights into differentiation trajectories, cell fate decisions, spatial relationships, and rare cell types (15–21).

The *Drosophila melanogaster* embryo has been an exquisite model for the patterning principles that shape cellular identities. The fertilized egg undergoes 13 rapid nuclear divisions resulting in a syncytial embryo of ~6000 nuclei. By developmental stage 5, nuclei have moved to the embryo periphery, become surrounded by cell membranes, and spatial gene expression patterns emerge as cells translate anteroposterior and dorsoventral positional information into transcriptional responses [e.g., (22)]. Stage 6 is marked by the first morphogenetic movements after cellularization completes and gene expression around this stage has been extensively assayed in whole embryos (e.g., (23)), in mutants converting entire embryos to germ layers (4) and in dissected slices (24). Available in situ databases present systematically annotated spatial gene expression (25, 26), but they often stop short of single-cell resolution, direct comparison of several genes per cell, genome wide profiling of all transcripts including noncoding RNAs, and quantitative assessment of gene expression.

To overcome these problems, we used massively parallel droplet-based single-cell sequencing (Drop-seq) (19) and quantified gene expression across >10,000 fixed cells from dissociated embryos (27) at a median depth of thousands of genes per cell. Computational analysis of the high-resolution in situ patterns of 84 genes (28) indicated that most if not all cells of the fly embryo have a unique transcriptional identity, highlighting the need to resolve the embryo at single-cell resolution. Previous efforts to map sequenced embryonic cells back to their origin [e.g., (27)] did so by reducing mapping complexity (e.g., by binning the entire zebrafish embryo composed of thousands of cells into 128 expression regions) and these methods could not correctly map our data at the required resolution. Therefore, we de-
Deconstructing the transcriptomic state of the embryo

In order to assess transcriptome diversity genome-wide with single-cell resolution, we hand-selected embryos at the onset of gastrulation (stage 6, Fig. 1A). Cells were extracted from >5000 precisely staged embryos, methanol fixed (27) and sequenced in seven Drop-seq runs corresponding to five biological replicates (table S2) resulting in a total of approximately 7975 sequenced D. melanogaster cells (table S3 and Fig. 1A, left). The vast majority of the sequenced cells (>90%) represented single-cell transcriptomes, as assessed by mixing cells from stage-matched D. melanogaster and virilis embryos (fig. S1A and tables S2 and S3). Gene expression correlated well across Drop-seq replicates (R > 0.94) and between Drop-seq and stage-matched, unfixed, whole embryos (R > 0.88) (fig. S1B), and was consistent with absolute quantification in individual stage-matched embryos (29) (fig. S1C).

To concentrate on cells with unambiguous patterning information, we excluded pole cells, yolk nuclei, and cell doublets from further analysis. Pole cells constitute a discrete lineage and contribute only to the germ line (30), whereas yolk nuclei function primarily in energy metabolism (31). Pole cells and yolk nuclei readily separated in a principal component analysis (fig. S1E). Similarly, t-SNE representation (32) correctly groups cells according to cell type, while a central cluster of doublets (cells expressing markers of distinct dorsoventral territories, i.e., mesoderm; neuroectoderm; dorsal ectoderm; see table S4) emerges centrally (fig. S1E, inset, and fig. S1D). Furthermore, we considered only cells with ≥12500 unique transcripts and expressing >5 genes of the BDTNP reference atlas (see below). The remaining ~1300 high quality cells had a median unique transcript number of >20,800 mapping to >3100 genes (fig. S2C). These cells separated along the first two principal components by dorsoventral identities (fig. S2D), but not by biological replicates (fig. S2E). The embryos contained a DsRed reporter transgene under control of a ventral neurogenic ectodermal vnd enhancer (33) and DsRed transcripts were primarily detected in a subset of cells that also score highly for broad neuroectodermal markers (fig. S2D). We “in silico dissected” the embryo by merging the transcriptomes of cells expressing specific markers for various dorsoventral territories and found that the merged transcriptomes accurately reflected gene expression in the respective domains (fig. S2, A and B). This dissection procedure is versatile and represents a distinct advantage of single-cell analysis over traditional bulk analyses (see Supplementary Note 1: “In silico dissection”).

We concluded that individual cells were sequenced without significant batch effects or bias for particular cellular identities, and that our Drop-seq data accurately reflect the transcriptomic state of individual embryonic cells.

Spatial reconstruction of the embryo

High quality cells could be grouped into 9 prominent cell clusters in the stage 6 embryo (fig. S2F). Genes well-known for their roles in embryonic patterning and tissue specification were readily identified (table S5; see Supplementary Note 2: “Clustering Analysis”). Spatial expression of cluster-specific driver genes as assessed by in situ hybridization (25, 26) was similar within clusters and spatial coherence within clusters was further supported by gene ontology (GO) term enrichments (fig. S3 and Supplementary Note 3: “GO terms”).

As might be expected, transcription factors were prevalent among the genes that drive cluster identity (table S5; e.g., fkh, kni, grn, hb, Abd-B, ken oc, and toy). Several of the highly variable genes remain unstudied in early development (see table S5; e.g., DNaseII, Z600, Meltrin, mttr, Atx-1, and several genes without names (“CG’s”), though their cluster-specific expression suggests functional roles in early embryogenesis. Furthermore, numerous long noncoding RNAs (lncRNAs) and a microRNA precursor are among the most variable genes (table S5; e.g., CR44683, CR43302, CR43279, CR41257, CR45185), which indicates an unrecognized role for lncRNAs in early embryonic patterning and development.

We sought to map cells back to their position of origin to produce a virtual embryo with single-cell transcriptome resolution by using a database of known in situ markers (Fig. 1B). The Berkeley Drosophila Transcription Network Project (BDTNP) generated in situ hybridization data for 84 genes, resulting in a quantitative high-resolution gene expression reference atlas with substantial combinatorial complexity (28). To correlate these 84 marker genes with our single-cell transcriptomes, we binarized the BDTNP atlas by manually choosing thresholds for each gene (25, 26) (fig. 2A.I). The combinatorial expression of these 84 binarized BDTNP markers sufficed to uniquely classify almost every position within the embryo (fig. S5A). Attempts to map our single-cell data using previously published algorithms (20, 21) was unsuccessful. We therefore designed a mapping strategy based on distributed mapping scores (DistMap, see Supplement). We binarized the Drop-seq data (fig. 2A.II; see Supplement), then compared the profiles of each cell against each bin, collected the (mis)matches into confusion matrices and computed Matthews’s correlation coefficients (MCC) for every cell-bin combination (fig. 2A.III). The result is a dis-
tributed mapping score for any sequenced cell across all embryonic positions (fig. 2A.IV). Figure 2B explores DistMap’s efficacy and demonstrates that sequenced cells mapped to a few cell diameters with high confidence (red) approaching the accuracy of the binarized bins themselves (green), whereas random mapping positions are spread throughout the whole embryo (blue). Due to the high transcriptional complexity, we reasoned that mapping a cell to multiple likely positions would be more meaningful than assigning it to a single-location. DistMap covers most of the embryo (Fig. 2C), assigns cells confidently (fig. S5C) and allows the quantification of many more genes per bin than initially detected in individual sequenced cells, so that most bins exhibit 6500 to 8500 expressed genes (Fig. 2D).

To compute the spatial expression of a gene we combined normalized gene expression per cell with the MCC scores for every cell-bin pair (see Supplement). This allows querying any given gene across all cells of the virtual embryo and produces a “virtual in situ hybridization” (vISH). To assess prediction quality, we computed vISHs of the 84 BDTNP-mapped genes using all high quality cells, as well as from subsamples and compared the resulting vISHs against the BDTNP database. The discrepancies saturate by ~750 cells (fig. S5B), so that the mapping would only be marginally improved by including more than our full set of ~1300 high quality cells.

To uncover the spatial signature of the nine single-cell clusters described above (Fig. 3A), we calculated the average mapping scores per bin across all cells per cluster (Fig. 3B). The concordance of the spatial signatures of these 9 principle clusters with regionally confined developmental fates [e.g., (34, 35)] is striking. Although cluster 4 largely encompasses the primordium of the mesoderm, cluster 3 corresponds to the future neuroectoderm and ventral epidermis, and cluster 6 corresponds to the dorsal epidermis and extraembryonic tissues. Cluster 9 in anterioventral regions corresponds the future esophagus and pharynx, whereas cluster 2 and 7 will give rise to the anterior- and posterior midgut upon invagination. Furthermore, this spatial cluster mapping is in agreement with GO term enrichments (see Supplementary Note “GO terms”).

**The virtual embryo predicts spatial gene expression**

With single-cell transcriptomes confidently mapped onto the embryo, each of the positional bins can be individually queried for gene expression. Our online *Drosophila* Virtual Expression eXplorer (DVEX, www.dvex.org) allows generation of vISHs for single genes and combinations. Predictions are displayed on a virtual embryo in multiple orientations (Fig. 4A) and expression gradients can be estimated along the anteroposterior and dorsoventral axes (e.g., Fig. 4B). Furthermore, DVEX provides an interactive environment to explore the t-SNE representation, gene expression in clustered cells, and genes driving clustering (e.g., Fig. 4C).

We observed close concordance between vISH predictions and expression detected by RNA in situ hybridization for genes expressed in a wide variety of patterns (see Fig. 5 and fig. S4), including vnd:DsRed reporter expression in the ventral neuroectoderm. Many of the genes shown were not previously known to be patterned at stage 6. Especially striking were predictions restricted to small patches of cells – expression limited to a few cells is often undetectable in traditional transcriptome studies, but is resolved by vISH (Fig. 5 and fig. S4).

vISH can be used to identify genes with distinct spatial patterns. We predicted the spatial expression of the 476 most highly variable genes, clustered their correlation matrix and identified 10 parental branches, which generate archetypal expression patterns when averaged (fig. S5E). These archetypes reflect the predominant transcriptional patterning responses, identify gene sets that respond to similar regulatory cues, and allow the discovery of unstudied and unusual gene expression patterns.

**Identification of potential developmental regulators**

We generated vISHs for >150 DNA-binding transcription factors that are detectably expressed in the stage 6 embryo (fig. S6), many of which are un- or understudied with respect to early development. Additionally, we predicted patterned expression of 16 genes that contain DNA binding domains (36). These are likely transcription factors (fig. S6A) and we experimentally validated the vISH predictions for two out of two patterned candidates, CG34224 and CG10553 (Fig. 5B). This comprehensive overview of transcription factor expression allows spatial assessment of regulator combinations that may activate or restrict target genes locally (fig. S6B).

Several long noncoding RNAs (lncRNAs) have also been shown to be potent regulators [e.g., (37)], but have not been assayed globally and systematically in early *D. melanogaster* development. By screening DVEX, we identified dozens of expressed and patterned lncRNAs (fig. S6C). The lncRNAs CR44317, CR44691, and CR45693 for example are weakly expressed, rendering them barely detectable in whole embryo sequencing data (23); however, RNA in situ hybridization showed reliable transcript signals in the predicted spatial domains (Fig. 5C and fig. S4). Additionally, vISH predictions for CR45559 and CR44917 were partially confirmed (Fig. 5C and fig. S4). The expression patterns of these lncRNAs range from dorsoventral modulation to gap-, terminal- and pair-rule patterns.

CR43432 expression prediction was particularly unusual, as it combines ventral, posterior, and dorsal aspects. CR43432 appears to “wrap around” lateral regions of the...
embryo and it appeared specifically excluded from the neu-rectoderm by vISH (Fig. 5D). In fact, expression is strongly anti-correlated with the neurectoderm marker SoxN at the single-cell transcriptome level and double ISH confirms mutu-ally exclusive expression (Fig. 5D). Additionally, CR43432 is highly expressed in yolk nuclei (Fig. 5D, right). The com-plementary non-neurogenic expression of CR43432 suggests it might act to delimit neurogenic genes or to promote non-neurogenic fates. In total, we discovered ~40 lncRNAs pre-dicted in a multitude of patterns (fig. S6C). Taken together, vISH is a powerful tool to discover novel putative regulators of embryonic patterning.

Cell communication by spatially regulated signaling

The Hippo signaling pathway is a major regulator of organ size, cell cycle, and proliferation (38, 39), but has to our knowledge not been implicated in the early embryo. By que-rying where transcripts of ligands, ligand modulators, recep-tors, and signal transducers are expressed, we identified patterned expression of major Hippo signaling components along the anteroposterior axis, with overlapping expression primarily in an anterior domain (Fig. 6A). Coexpression of these molecules may promote Hippo signaling, which culmi-nates in the phosphorylation of the transcription factor Yorkie, thereby diminishing Yorkie’s nuclear localization (38, 39). After mitotic arrest at stage 5, cell cycle re-entry is delayed in an anterior region (40) and it is conceivable that active Hippo signaling in that domain delays mitotic onset. Using antibodies against Yorkie and a mitosis marker (phosphorylated histone H3), we detected higher nuclear-cytoplasmic ratios of Yorkie in cells undergoing mitosis in anterior patches at ~stage 7 (Fig. 6B), suggesting active Hippo signaling in intervening regions. To our knowledge Hippo signaling has previously not been implicated in cell-cycle regulation in early Drosophila development.

Additionally, we predict that components of other signal-ing pathways are expressed in a spatially restricted fashion, including alternate ligands, receptors, and antagonists of Dpp/TGFβ (fig. S7A). Our experimental data suggests ante-rior repression of the TGFβ signaling cascade (fig. S7B) (see Supplementary Note 4: “TGFβ signaling modulation”). Hence, by analyzing the expression of signaling molecules in the embryo with spatial resolution, we are able to predict where signals originate and where they can be transduced.

Detection of evolutionary gene expression changes

Several cis- and trans-regulatory circuits have diverged be-tween D. melanogaster and D. virilis [e.g., (41, 42)]. We asked whether changes in gene expression patterns over the course of speciation might be detectable using DVEX. Cluster-ing obtained from 673 stage 6 D. virilis high quality cells (fig. S8A) bore a striking similarity D. melanogaster with respect to cluster number, proportional cluster size, and cluster mapping by vISH (compare Fig. 3 and fig. S8A). Gene expression correlation between merged transcriptome data of the two species was high (R = 0.77). We used the virtual embryos of D. melanogaster and virilis to systemati-cally compute vISHs, compare orthologs and identify diver-gences.

The genes CG6660 and GJ14350 are homologous by pro-tein conservation and genomic synteny (table S8), with CG6660 predicted not to be expressed in D. melanogaster (fig. S8B, left), whereas GJ14350 was predicted to be ex-pressed in an anteroposterior stripe-modulated pattern in D. virilis (fig. S8C, left). By RNA in situ hybridization CG6660 was not detectable, whereas GJ14350 was expressed in stripes similar to the prediction (fig. S8C). For the ho-mologous pair fok/GJ17890 (table S8), fok was predicted and verified by in situ hybridization to be expressed in an ante-rior ventral patch in D. melanogaster (fig. S8D), but vISH predicted absence of the anterior patch and weak posterior expression of GJ17880 in D. virilis. In situ hybridization in D. virilis showed that although there is a tendency for low posterior expression of GJ17880 at early stage 6/7 by RNA in situ hybridization, robust posterior staining was not seen until stage 8; however, the absence of anterior expression in D. virilis was confirmed (fig. S8E). These examples illustrate that DVEX can serve as a sensitive tool for the identification of gene expression changes.

Discussion

Here, we resolved a metazoan embryo composed of ~6000 (or ~3000 when considering bilateral symmetry) individual cells. Although the Drosophila embryo may be an extreme example where each cell has a unique transcriptional pro-file, the transcriptomes of neighboring cells can be very simi-lar to each other. To successfully map dissociated and sequenced cells to their correct spatial position based on combinatorial expression of marker genes, we required a suitable set of marker genes, deep capture of gene expres-sion in each cell, and powerful computational mapping to be able to confidently score differences between an enor-mous number of possible mapping possibilities. To illustrate the latter point, if considering only 1000 sequenced cells across only 1000 locations, one already would have to cal-culate one million possibilities.

We were able to overcome these challenges and to pro-duce a “virtual embryo” with ~8000 genes per cell due to three main reasons. First, the 84 in situ markers captured sufficient spatial transcriptional complexity to allow us to guide mapping of each sequenced cell to its positions. Sec-ondly, we optimized our Drop-seq approach to reliably cap-ture thousands of genes per cell. Thirdly, and perhaps most importantly, we devised a novel mapping strategy,
“DistMap”, which reliably maps single-cell transcriptomes back to their origin. DistMap is scalable and extendable to other 3D tissues at single-cell resolution. This is because DistMap uses measured gene expression and does not require transcript-level imputation and its scoring scheme is suitable for sparse data sets. Additionally, distributed mapping limits the effect of outliers and populates positions with transcript information beyond the base sequencing level; in this way, from an original depth of ~3500 genes captured/cell, we were able to assign on average ~8000 genes/cell. Nevertheless, DistMap clearly can be improved in several respects; for example, it currently uses binarized rather than continuous data and maps each cell independently, rather than allowing mapped cells to improve subsequent scoring.

Once a virtual embryo has been produced, what kind of biology can be learned? We first built a computational platform (www.dvex.org) that allows interactive interrogation of single-cell transcriptome data in spatial context, including the computation of gradients. We then leveraged DVEX to compute thousands of virtual in situ and to select genes that had “interesting” expression patterns. For example, we identified patterned transcription factors never implicated in early development before, as well as dozens of IncRNAs with intriguing and sometimes novel expression patterns. Since we used a second fly species to control for cell doublet frequency, we incidentally acquired a virtual embryo for Drosophila virilis. Even though these species are separated by at least 40 million years of evolution and have clearly diverged cis-regulatory DNA sequences [e.g., (43, 44)], we found only a few cases with clear expression divergence, which highlights strong selection pressure on maintaining gene expression patterns at this early stage. It also suggests a large extent of gene regulatory plasticity where cis-regulatory sequences may diverge, whereas the overall expression patterns remain largely unchanged.

We uncovered a substantial amount of transcriptional modulation of components of major signaling pathways. Local expression of ligands sets up “signal sources”, but the ability to respond to these signals appears to be heavily regulated at the transcriptional level, even early in development – from patterned expression of specific receptor molecules, to modulators of signal transduction. One such case is Hippo signaling, which has not been described to play a role in early Drosophila development. Active Hippo signaling has been connected to cell cycle delay and diminished proliferation (39). Thus, the prediction of expression of major Hippo pathway components in an anterior subdomain (Fig. 6A) was of interest. Indeed, we detected evidence of productive Hippo signaling by showing that the transcriptional effector Yki is diminished in anterior nuclei that do not undergo mitosis (Fig. 6B). More than 30 years ago, Volker Hartenstein and Jose Campos-Ortega employed fuchsin staining to show that mitotic reentry after stage 6 occurs asynchronously (40). Our data shows that localized Hippo signaling constitutes a mechanism that breaks synchronicity of cell cycle re-entry in early fly embryogenesis.

In general, how many guide in situ are needed to reconstruct tissues after dissociation and single-cell sequencing? The answer depends, apart from sequencing depth, clearly on the transcriptional complexity and developmental stage of the tissue. In early metazoan development, most decisions about spatial identity are carried out by a temporal cascade of combinations of transcription factors (45). In our case, 84 in situ (mostly transcription factors) sufficed to uniquely and individually label most of the ~6000 cells. However, it may be possible to assemble complex tissues from sequenced cells without using in situ markers as guides, somewhat akin to solving a puzzle. Clearly, we need a better understanding of the “design principles” of gene regulation to achieve this or to test ideas about these principles. For example, in early development, the expression of most genes generally does not change in a discontinuous fashion from cell to cell. This feature could be implemented in future versions of DistMap to reduce the number of guide expression patterns needed.

REFERENCES AND NOTES


ACKNOWLEDGMENTS

We thank M. Biggin and S. Keranen (LBNL) for discussions and unpublished BDTNP data, R. Satija for initial DropSeq help, S. Ugowski (MDC) for experimental assistance, S. Small (NYU) for a transgenic Drosophila line, A. Stathopoulos (Caltech, NIH R35GM118146) for sharing unpublished results, J. Zeitlinger (Stowers) for Yki antibody, E. Lauter (Columbia) for pMad antibody, N. Friedman (Hebrew University) and members of the Rajewsky and Zinzen labs for constructive discussions, D. Munteanu (BIMSB/MDC) for IT support, and the DFG (SPP 1738, RA 838/8-1, RA 838/5-1) for funding. Raw and processed data sets are available from the GEO repository (GSE95025). The DistMap R-package is available at https://github.com/rajewsky-lab/distmap. N.R. and R.Z. defined strategy, supervised, and procured funding; N.K., P.W., J.A., C.Ko., N.R., and R.Z. designed experimental strategy; P.W. did fly genetics and embryo collections; J.A. set up; C.Ko. supervised; J.A., S.A., A.B., and C.Ko. performed Drop-seq; A.B., S.A., and P.W. prepared sequencing libraries; N.K. developed and implemented computational analyses/tools, including DistMap and DVEX; P.W. and C.Ki. validated predictions experimentally; N.K., P.W., C.Ko., N.R., and R.Z. analyzed data and wrote the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aan3235/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S8

Tables S1 to S8

References (46–57)

29 March 2017; accepted 24 August 2017

Published online 31 August 2017

10.1126/science.aan3235

First release: 31 August 2017 www.sciencemag.org  (Page numbers not final at time of first release) 7
Fig. 1. De- and reconstructing the embryo by single-cell transcriptomics combined with spatial mapping. (A) Single-cell sequencing of the Drosophila embryo: ~1000 hand-picked stage 6 fly embryos are dissociated per Drop-seq replicate, cells are fixed and counted, single cells are combined with barcoded capture beads, libraries are prepared and sequenced. Finally, single-cell transcriptomes are deconvolved, resulting in a digital gene expression matrix for further analysis. (B) Mapping cells back to the embryo: Single-cell transcriptomes are correlated with high-resolution gene expression patterns across 84 marker genes, cells are mapped to positions within a virtual embryo, expression patterns are computed by combining the mapping probabilities with the expression levels (virtual in situ hybridization).
Fig. 2. Reconstructing the embryo by spatial mapping based on distributed scores. (A) DistMap. The 84 BDTNP gene expression patterns (I) and the single-cell expression profiles (II) were binarized. (III) Confusion matrices are calculated scoring expression (dis-)agreement between the transcriptomes and the ~3000 positional bins of the reference atlas. Matthews’s correlation coefficients (MCCs) are calculated for every cell/bin combination. (IV) Positional assignment for each cell is distributed based on MCCs across all bins. (B) Density plot showing mapping confidence (mean Euclidean distance) between a cell’s highest scoring location and the following six. Single-cell transcriptomes (red) map to embryo positions with similar confidence as cells of the reference atlas (green). (C) Bin coverage across the embryo. >87% of all locations in the embryo are confidently covered ($P < 0.05$; see Methods for details). (D) The virtual fly embryo has a resolution of 6000-8000 genes/cell.
Fig. 3. Sequenced cells cluster by spatial identity. (A) 2D t-SNE representation of the high quality cells shows 9 major clusters grouped by transcriptome similarity. (B) Mapping of clusters reveals that cells within each cluster share a contiguous spatial domain.
Fig. 4. DVEX accurately predicts spatial gene expression patterns. DVEX (www.dvex.org) is the online resource for the virtual embryo. (A) Virtual in situ hybridization (vISH) for, the pair rule gene ftz (red) and the mesodermal gene sna (green) in five orientations. Stippled box indicates cells analyzed in (B). EL, egg length; DV, dorsoventral; AP, anteroposterior. (B) Quantification of relative expression per cell mapped along an axis (here dorsoventral) for stumps (expressed in the ventral mesoderm, left) and the vnd::DsRed reporter (primarily expressed in the ventral neurectoderm, right). Relative expression in log space, thresholds were 0.85, embryos are oriented anterior left. (C) Examples of marker genes and their expression in t-SNE clustered cells. Expression indicated, grey (low) to red (high).
Fig. 5. Prediction accuracy and detection of new regulators. (A) vISH predictions are accurate across a wide variety of expression patterns. Expression of CGs had not been reported previously. (B) Patterned expression of putative transcription factors. (C) Patterned expression of lncRNAs. (D) CR43432 and pan-neurogenic genes are expressed in complimentary patterns. Dual vISH of SoxN and CR43432 (top left), double in situ hybridization validates the predicted expression. CR43432 is additionally expressed in yolk nuclei (not shown in vISH).
Fig. 6. Spatial regulation of Hippo signaling in the embryo. (A) vISHs predict patterned expression of Hippo signaling components in stage 6 embryos. Shown are components involved in receiving the signal (receptors, ligands), transducing it through the cytoplasm (transducers) and inhibition of the transcriptional cofactor Yorkie (Yki). Ubiquitous pathway components are indicated, vISHs for patterned components are shown. Most patterned positive Hippo components are expressed in anterior regions. Active signaling culminates in nuclear exclusion of Yki. (B) Shown is the anterior of a stage 7 embryo, cephalic furrow indicated by stippled white line, anterior left. Staining with antibodies against phosphorylated histone H3 (H3S10-P) marks cells undergoing mitosis, nuclear Yorkie (Yki) is depleted in cells not marked by H3S10-P.