Development 140, 216-225 (2013) doi:10.1242/dev.086975

A continuum of transcriptional identities visualized by combinatorial fluorescent *in situ* hybridization

Lars Martin Jakt*, Satoko Moriwaki and Shinichi Nishikawa

SUMMARY

Oligonucleotide-based fluorescent *in situ* hybridization (FISH) coupled with high-resolution high-sensitivity microscopy allows the visualization of single RNA molecules within fixed cells and tissues as distinct foci. We show here that combinatorial labeling of RNA molecules with several fluorescent dyes extends the number of genes that can be targeted simultaneously beyond the number of fluorophores used. This approach also inherently validates the identification of transcripts reducing false positive counts. We have used combinatorial FISH and image analysis to measure the transcript densities of six genes using three fluorophores. This has allowed us to visualize the endothelial maturation of lateral mesoderm in an *in vitro* ES differentiation assay from a single snapshot of molecular identities. Our observations show that, under these specific conditions, endothelial maturation follows a homogeneous course with a gradual increase in expression of *Cdh5* and a concomitant loss of early transcription factors, arguing that maturation is governed in a generally deterministic manner. This methodology is limited by the number of fluorophores that can be used and by the available microscopic resolution, but currently available equipment should allow the visualization of transcripts from 10 or more genes simultaneously.

KEY WORDS: RNA visualization, Cell state, Differentiation, Fluorescent *in situ* hybridization, Single cell, Single molecule

INTRODUCTION

The primary unit of biological control is the cell. Measurement of cellular properties using bulk populations of cells is thus problematic unless a reasonable extent of homogeneity of identity or behavior exists within the population. As the output of development or cellular differentiation is the generation of cellular heterogeneity, the study of such processes requires simultaneous measurements at the single cell level of a sufficient number of parameters to describe the internal variance. As cell interactions are fundamental to developmental processes, the description of spatial arrangements is also desirable.

We therefore sought to develop methods for *in situ* multiparametric quantification of gene expression at the single cell level in large numbers of cells. Currently, existing methods for single cell multiparametric measurements can be broadly divided into cytometry and RT-PCR-based assays. Modern RT-PCR-based methods allow the simultaneous measurements of large numbers of parameters in intermediate cell numbers (~100); however, measurements rely on exponential amplification and expression estimates can vary by orders of magnitude from linear-based methods (Hayashi et al., 2008; Schwanhäusser et al., 2011) making data interpretation difficult.

Cytometry (e.g. fluorescence activated cell sorting; FACS) allows rapid multiparameter (up to 10) measurements on very large number of cells; recently, the combination of cytometry with mass spectrometry has allowed the large scale characterization of 31 antigens simultaneously (Bendall et al., 2011). Cytometry is, however, limited by the availability of suitable antibodies and its accuracy is difficult to assess.

In situ-based methods allow the localization of expressing cells, but have historically not provided quantitative expression measurements. Although it was shown in 1998 that oligonucleotide-based fluorescent *in situ* hybridization (FISH) combined with high resolution microscopy can be used to visualize individual mRNA molecules (Femino et al., 1998), this route to quantification has only been used to address very specific questions (e.g. Raj et al., 2006). More recently, Raj et al. (Raj et al., 2008) proposed the use of numerous short (20 nucleotide) probes labeled with single fluorophores for the identification of individual transcripts. This method has two primary benefits over previous methods: (1) the use of large numbers of probes targeted to individual transcript regions provides an amplification of the signal that is dependent on the sequence targeted; (2) a lower cost of probe synthesis allowing routine use. Such methods provide unambiguous quantification of transcripts but have been limited in mammalian cells to the measurement of three genes simultaneously (Itzkovitz and van Oudenaarden, 2011).

The use of multiple probes per transcript allows the encoding of transcript identities using a combination of fluorophores and it was demonstrated in 2002 that this can be used to detect the sites of transcription of more than 10 genes simultaneously (Levsky et al., 2002). This and subsequent work (Raj et al., 2006) demonstrated the stochastic nature of transcription, showing that active transcription cannot be used to describe cell state, as gene activation (a change in state) implies not continuous transcription but an increase in the probability of transcriptional events. Conceptually, the methods used by Levsky et al. (Levsky et al., 2002) can be used to detect individual transcripts; however, the use of a small number of probes per transcript (up to five used) do not provide an amplification of the signal in the absence of colocalization of large numbers of transcripts (observed at sites of transcription) resulting in a poor signal. Here, we show that the use of large numbers of short probes allows transcripts to be targeted by several fluorophores, allowing the simultaneous quantification of gene expression from more than six genes in cultures of mammalian cells. We also present a workflow for the development and verification of probe sets, FISH, microscopy and analyses of both raw image data and subsequent transcript counts.

Stem Cell Biology Group, Riken Center for Developmental Biology, 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Japan.

*Author for correspondence (mjakt@cdb.riken.jp)

Accepted 25 September 2012

We have used these methods in combination with *in vitro* differentiation of mouse embryonic stem (ES) cells in order to test whether combinatorial FISH can be used to determine the manner in which cellular identities change. We have previously investigated the transition from primitive mesoderm to the endothelial lineage using both high-throughput and conventional methodologies (Kataoka et al., 2011; Sakurai et al., 2006; Yamashita et al., 2000). This has allowed us to identify genes that change during this transition; however, this has not provided a quantitative understanding of the manner in which gene expression changes during the process. Furthermore, the unknown heterogeneity of the bulk populations analyzed makes it impossible to determine co-expression or order of induction. Using combinatorial FISH, we show that it is possible to determine axes of both lineage and stage differentiation through a single snapshot of cell identities, and thus derive an order of events.

MATERIALS AND METHODS

Probe preparation and fluorescent *in situ* hybridization were carried out mostly as described previously (Raj et al., 2008), with some modifications in sample preparation, buffers and processing details. Reagents were obtained from Wako (Osaka, Japan) unless otherwise mentioned.

Probe design

As ~24 probes (each containing a single fluorophore) are required to detect transcripts reliably, a minimum of around 500 bases of a suitable target sequence is required for each fluorophore. In order to reduce differences in signal intensities owing to alternative splice forms, we first selected target sequences from exons common to most known splice forms. We then masked regions of those that have strong similarities to other transcript sequences using blast at http://www.ensembl.org/. The resulting sequences were used as input to the probe design tool at http://www.singlemoleculefish.com using default settings. We designed 48 or 72 probes per gene, allowing the use of either two or three fluorophores. We attempted to avoid non-coding regions (as these may contain more secondary structure), but given prior constraints (sequence similarity and common exons), we often used extensive non-coding regions as well as GC-rich regions. Probe sequences are in supplementary material Table S1.

Probe synthesis and labeling

Probes were ordered from Biosearch Technologies (http://www.biosearchtech.com) with 3′ amino modifications in 96-well plates at a 5 nmol scale of synthesis. Dyes were purchased from Invitrogen (Alexa 488) or GE Healthcare (Cy3, Cy5) as NHS or 5-SDP esters. Probes were dissolved to a target concentration of 500 ng/μl on the basis of expected yield. For each gene, we created intercalated pools of probes containing 24 or 48 oligos (3 μl of each oligo). These pools were vacuum dried using a DNA110 Speedvac and dissolved in 20 μl 0.1 M carbonate buffer (pH 8.8) containing 140-200 ng of freshly dissolved fluorophore (NHS or 5-SDP esters). The mixtures were incubated overnight at 25°C with shaking. Unincorporated fluorophore was removed by gel filtration using NAP5 Sephadex G25 columns (GE Healthcare) preceded by either ethanol precipitation or serial extraction with 1-butanol. Eluates were concentrated to a few μl by rotary evaporation and resuspended in 50 μl 5% acetonitrile (AcN) in 0.1 M triethylammonium acetate (TEAA) (buffer A). Unlabeled and labeled oligonucleotides were separated using reverse-phase high-performance liquid chromatography (HPLC) using a gradient of 18 to 30% buffer B (65% AcN in 0.1 M TEAA) on a 0.5 ml C18 4 μm Inertsil HPLC column (GL Sciences, Tokyo, Japan). Fractions containing labeled probes were pooled and subjected to one more round of gel-filtration and the concentration adjusted to 50 ng/μl.

Sample preparation

Combinatorial FISH was performed on cells grown on glass slides coated with E-Cadherin-IgG-Fc fusion protein (E-Cadherin-Fc: R&D Systems, 648-EC) (Nagaoka et al., 2006) (undifferentiated), or Collagen type IV (differentiation). E-Cadherin-Fc (10 μg/ml in PBS) and Collagen type IV (10 μg/ml in 0.1 M HCl) were adhered to poly-L-lysine (PLL) coated slides (Matsunami Glass, Osaka, Japan) for a period of 4 hours to overnight at 37°C. Slides were washed briefly with PBS before seeding cells. We used Grace Bio-Labs (http://www.gracebio.com) silicon gaskets (CW-8R-1.0 CultureWell Gasket) to create eight wells with a diameter of 6 mm, into which 5000-20,000 cells were cultured for up to 2 days in 50 μl medium.

Embryonic stem (ES) cell maintenance and differentiation

CCE ES cells were maintained and differentiated as described previously (Yamashita et al., 2000). Mesodermal differentiation was initiated by plating 20,000 ES cells in ColIV-coated six-well plates (Becton Dickinson) with αMEM (Gibco, Life Technologies) containing 10% fetal calf serum. These cells were split on day four and 10,000-20,000 cells were transferred to ColIV-coated glass slides and cultured for up to 48 hours in SFO3 (Sanko Junyaku) medium supplemented with 30 ng/ml VEGF. EB3 ES cells were kindly provided by Hitoshi Niwa (Riken, Japan).

FISH

Cells grown on slides were briefly washed with PBS and fixed with 4% PFA at room temperature for 10 minutes. Following fixation, cells were washed three times (5 minutes each) in an excess of ice-cold PBS. Slides were then either dehydrated to 70% ethanol and stored at 4°C (for up to 2 weeks), or directly incubated with prehybridization buffer [PH; 10% formamide, 2×SSC, 0.1% Triton X-100, 0.02% BSA, 2 mM Ribonucleoside Vanadyl Complex (New England Biolabs, http://www.neb.com/)] at 30°C. Stored slides were rehydrated prior to incubation with prehybridization buffer as above before FISH. Probe sets were combined to give a final concentration of 1-2 ng/μl per probe set. If used, unlabeled competitors were added at this stage and the mixtures concentrated by rotary evaporation to a few microlitres and resuspended directly in hybridization buffer (HB; PH + 10% dextran sulfate, usually 20-30 μl/well). Hybridization was carried out in a humid chamber (moistened by 10% formamide) at 30°C overnight. The hybridization solution was applied directly to the wells formed by the silicon gasket and left uncovered. The following day, slides were washed twice with 10% formamide, 2×SSC, 0.1% Triton X-100 at 30°C for 30 minutes each. Slides were then transferred to PBS, stained with DAPI (0.5 μg/ml in PBS for 1 minute), briefly washed with PBS, dehydrated and mounted in Prolong Gold or freshly prepared Prolong antifade (Invitrogen) mounting medium. Slides were viewed on the following day, and, if sealed (with nail polish) and stored dry at −20°C, could be kept for up to 6 months without major degradation in the signal.

Probe competition

FISH was carried out as described, but including an excess (5 ng/μl per probe) of unlabelled oligonucleotides in the hybridization mixture.

Imaging

Slides were imaged using a Deltavision Core (Applied Precision, http://www.appliedprecision.com/) system equipped with a 60× 1.42 NA lens (Olympus), using standard filters for DAPI, FITC, TRITC and Cy5 illuminated either by a mercury or xenon light source. Image stacks were created using 200 nm step sizes with a 1024×1024 camera giving final voxel dimensions of 107×107×200 nm (*xyz*). The resulting image stacks were deconvolved using the Deltavision deconvolution program (decon3d) using 10 iterations with default settings.

Image analysis

Image stacks were viewed and analyzed by functions implemented in a custom application (dvreader) written in C++ using the Qt4 graphical user interface toolkit. All source has been released under the GPL (http://www.gnu.org/copyleft/gpl.html) and is available at http://www.gitorious.org/dvreader.

We used a thresholded three-dimensional watershed algorithm to initially define local maxima (blobs); blobs were initiated when intensity values exceeded a specified minimum (optionally after applying a local background subtraction). Only blobs with peak values above a user-specified threshold were included in further analyses. Transcript identities were assigned by identifying blob peaks from additional fluorescent

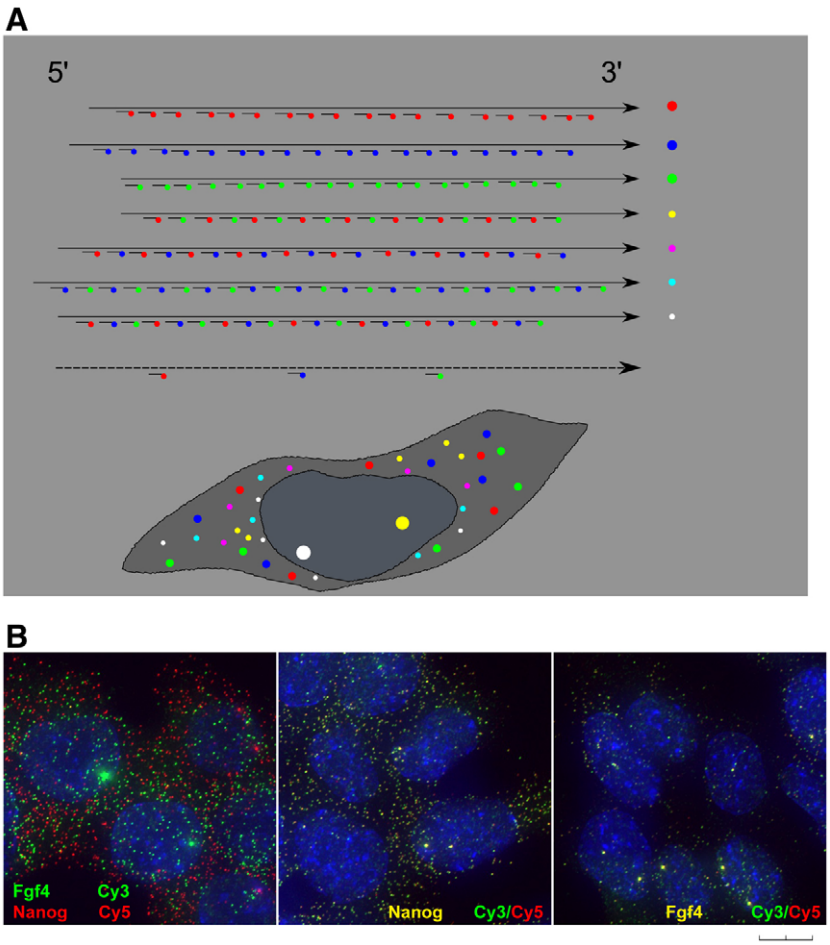


Fig. 1. Combinatorial fluorescent in situ hybridization (FISH). (**A**) Transcripts from different genes are hybridized with a set of short probes (20 nucleotides long) that have been covalently linked to a single fluorophore each. Probe sets contain either probes labeled with one kind of fluorophore producing bright singly fluorescent signals, or with probe sets containing more than one type that produce less bright signals in multiple fluorescent channels. Promiscuous hybridization to non-target transcripts (broken line) is likely to be restricted to a small number of probes and to be indistinguishable from the general background. Sites of transcription give rise to strong signals present within the nucleus. (**B**) EB3 ES cells grown in 2i medium (Ying et al., 2008) on E-cadherin-fc-coated (Nagaoka et al., 2006) glass coverslips were hybridized against probe sets targeted to *Nanog* and *Fgf4*. Left panel, FISH against *Nanog* and *Fgf4* transcripts using probe sets containing either Cy3- (Fgf4, green) or Cy5- (Nanog, red) labeled probes; middle and right panels, FISH against *Nanog* and *Fgf4* transcripts, respectively, using probe sets containing both Cy3- and Cy5-labeled probes. All probe sets contained a total of 48 probes. Scale bar: 10 μm.

channels with a peak-peak distance of less than 1-2 voxels after correcting for apo-chromatic- and emission filter-induced shifts. The resulting sets of overlapping blobs (blob sets) were filtered using parameters describing the constituent blobs (e.g. mean, max voxel intensity, blob volume, etc.). Allowable ranges for the parameters were set manually for each individual fluorophore and transcript id, as signals using different fluorophores and probe-sets have different expected properties (e.g. larger probe numbers give stronger signals). In the experiments described here, we have simply dropped blob sets that do not conform appropriately, but one can envision an iterative classification and filtering approach.

Nuclear segmentation

Nuclei were segmented by a simple tracing algorithm that is initiated upon encountering a pixel above a given threshold. Once such a pixel is encountered the algorithm searches the immediate space around that pixel for a maximal neighboring pixel above the threshold and iterates until it encompasses a region. The resulting perimeters were filtered by length followed by manual inspection and editing where necessary.

Cell segmentation

Blob sets were assigned to nuclei using a nearest-neighbor method. Clusters of blob sets were seeded from unassigned blob sets and expanded iteratively by identifying the blob set lying closest to any member of the cluster. The expansion was stopped, either when within a minimum distance of a nucleus, or when incorporating a member of an already assigned cluster. The resulting cell perimeters were manually checked and amended when necessary using functions built into the dvreader application. Although blob sets were mapped in three dimensions, both nuclear segmentation and manual correction of cell boundaries was performed in two dimensions, resulting in inevitable overlaps. Transcripts falling within such overlapping regions were omitted from further analyses.

Data were analyzed using both R functions and a custom-written application for summarizing n-dimensional relationships in two-dimensional space. The source code to this can be obtained from the sod/ directory of the dvreader source.

RESULTS

Combinatorial FISH tested in ES cells

To test hybridization conditions for combinatorial detection of individual transcripts, we first designed probes against *Fgf4* and *Nanog*. These genes are both expressed at high levels in undifferentiated ES cells and their expression is lost upon differentiation. Forty-eight 20-nucleotide probes were designed against both genes. The probes were combined into pools of odd- and even-numbered probes, and each pool labeled separately with either Cy3 or Cy5. Transcripts from both genes could thus be probed either with single colors (*Fgf4*, Cy3; *Nanog*; Cy5; 48 probes each) or with both fluorophores in combination (24 probes each). ES cells hybridized with singly labeled probe sets against both genes contained large numbers of bright singly fluorescent foci (Fig. 1B, left), demonstrating a general lack of colocalization of transcripts from the two genes. By contrast, cells hybridized with probe sets targeted against a single gene (*Fgf4* or *Nanog*), but labeled with Cy3 and Cy5, contained, almost exclusively, fluorescent foci labeled by both fluorophores (Fig. 1B, middle and right).

The almost complete overlap between channels suggested that the vast majority of signals are due to the presence of the targeted transcript. However, the presence of a small number of singly

labeled foci of similar size and intensity to the major signal indicates the need for extensive validation of probe sets if used to detect genes expressed at very low levels.

Expanding the probe set repertoire

We next designed probes against a number of angiogenic and hematopoietic genes (*Etv2*, *Fli1*, *Cdh5*, *Cldn5*, *Gata2*, *Gata1*, *Tal1* and *Runx1*) to facilitate the investigation of the differentiation of blood and endothelial lineages. Each probe set contained either 48 or 72 individual probes to allow single, double and triple labeling of individual transcripts. Probes were combined into either odd and even, or into sets containing every third position for probe sets containing 48 or 72 probes, respectively. Individual pools of probes were labeled with Alexa 488 (a488), Cy3 or Cy5 and tested on both ES cells (which should not express any of the genes) and on ES-derived ECs.

Approximately half of the probe sets tested gave rise to expected signals; however, the remaining probe sets caused massive background signals when used in FISH (Fig. 2). In order to identify background-producing probes we attempted to compete out the background using unlabeled probes. In most cases, background signals could be removed by the inclusion of an excess of one or a small number of unlabeled probes. In a few cases, we were unable to pinpoint all problematic probes and in one case (*Gata1*) we have as yet been unable to reduce the background to satisfactory levels.

We observed a range of different categories of background ranging from smooth cytoplasmic staining, globular nuclear and speckled to one or two extremely bright diffraction limited foci per nucleus (Fig. 2). Most of these probes cause similar signals in both ES and differentiated cells (mesoderm and mesoderm derivatives), suggesting that they interact with parts of the basic cellular apparatus.

We then used competitors to allow FISH with complete probe sets on cultures containing ECs to evaluate the performance of combinatorial FISH. FISH using singly labeled probe sets resulted in almost exclusively non-overlapping fluorescent dots (Fig. 3A). By contrast, using three fluorophores to target four genes with combinatorial labeling (*Tal1* a488/Cy5, *Gata2* a488/Cy3, *Etv2* Cy3/Cy5 and *Fli1* a488/Cy3/Cy5) resulted almost exclusively in fluorescent foci that overlapped between channels (Fig. 3B). This lack of non-overlapping signals prompted us to use a combination of both singly and combinatorially labeled probe sets. This resulted in the appearance of six classes of fluorescent foci with the expected combinations (Fig. 3C; Fig. 4A). As the resulting colors reminded us of those in a bowl of candy, we propose the term candyFISH to refer to this type of analysis.

To increase the number of usable fluorophores, we also used mega-stokes dyes (Dyomics, http://www.dyomics.com/) that can be distinguished from more standard dyes by using excitation at shorter wavelengths (owing to their larger Stokes shift). These dyes are not very bright, and are not suitable for use in combinatorial detection; however, they can be used in singly labeled probe sets allowing us to expand the number of genes targeted simultaneously (Fig. 3D, seven genes: *Gata2* a488, *Tal1* Cy3/Cy5, *Runx1* Cy3, *Etv2* Cy5, *Kdr* a488/Cy3, *Pdgfra* a488/Cy5 and *Fli1* Dy520). This allows the extension of the number of distinct type of foci that can be distinguished; but the analysis of the resulting images is complicated by the appearance of some degree of fluorescent resonant energy transfer (FRET) at combinatorially labeled transcripts. The use of more tightly spaced conventional fluorophores (e.g. Levsky et al., 2002) should more easily extend the number of usable dyes, but we have been limited by the filter sets available to us.

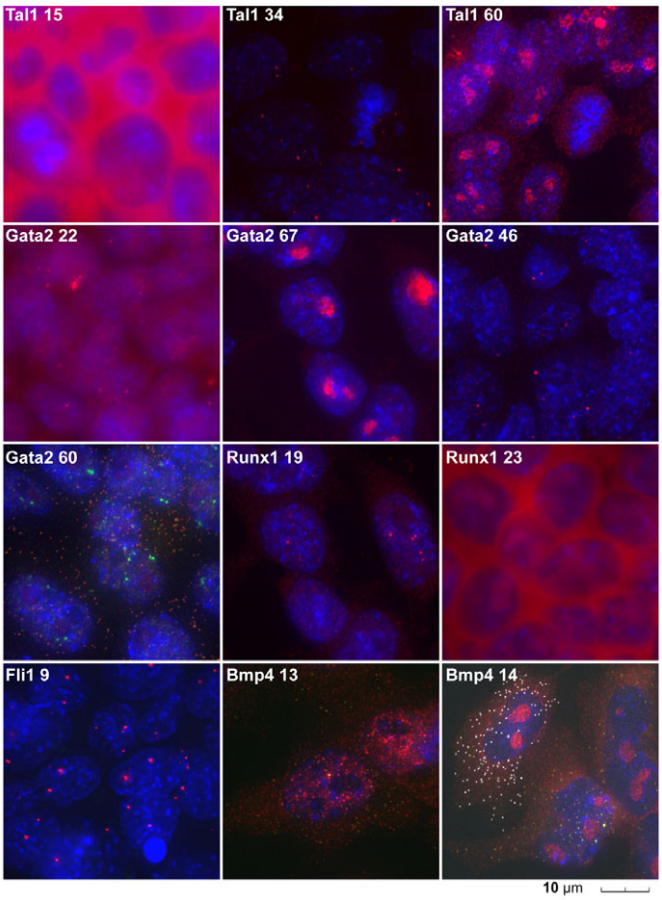


Fig. 2. Background screening. ES cells or ES-derived endothelial cultures containing ~50% ECs were subjected to FISH against partial probe sets (6-24 probes) in the absence of the indicated competitors. The specific probes giving rise to background signals were identified by competition with an excess of unlabeled competitor (not shown). In all cases, competition with a single 20-mer nucleotide was able to completely remove the background shown. Probes causing background signals are indicated. For an example of FISH in the presence of competitors see Fig. 3B,C. FISH for *Gata2* 60, *Bmp4* 13 and *Bmp4* 14 probes were performed against differentiated cells expressing the transcripts in order to distinguish appropriate (those showing overlap with alternate probes against the same transcript but labeled with a different fluorophore) and false signals. For these panels, yellow dots indicate signals that overlap with alternate probes against the same gene; for *Gata2* 60 false signals are visible in green; for all others false signals are in red. The *Bmp4* 14 panel also shows *Fli1* signals in white. Different exposure conditions have been used owing to the large range of intensity of backgrounds; in particular, *Tal1* 15 and *Fli1* 9 cause very high intensity background.

Converting image to expression data

In order to obtain image data from more (complete) cells, we collected image data from overlapping three-dimensional image stacks giving total dimensions of 400-500 μm in the *xy* plane. We then used a thresholded watershed algorithm to detect three-dimensional local peaks (blobs) for each fluorescent channel and assigned transcript identity on the basis of overlap between fluorescent channels.

Although it is conceptually easy to enumerate transcripts, it is not clear how to assign transcripts to individual cells. This is especially true in cultures of differentiating cells in which

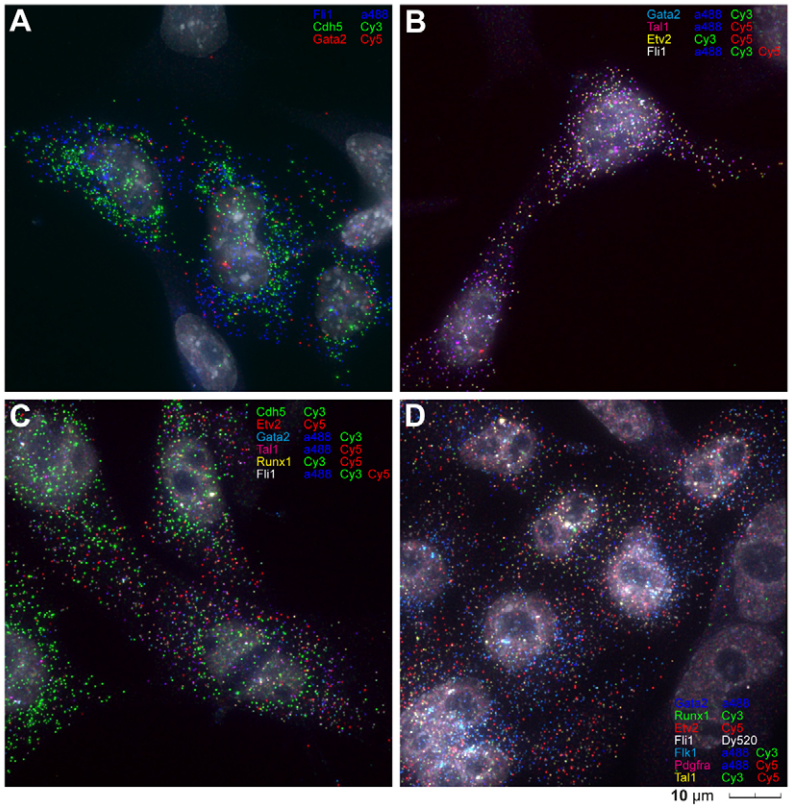


Fig. 3. Combinatorial FISH. (A-D) FISH was performed using either non-combinatorial (A) or combinatorial (B-D) probe sets. (**A**) FISH against *Fli1*, *Cdh5* and *Gata2* using probe sets containing 48 probes all labeled with a single fluorescent dye. (**B**) FISH against *Gata2*, *Tal1*, *Etv2* and *Fli1* using combinatorial probe sets. (**C**) FISH against six genes using probe sets containing one (*Cdh5*, *Etv2*), two (*Gata2*, *Tal1*, *Runx1*) and three (*Fli1*) different dyes. (**D**) FISH against seven genes using either double combinations or single combinations and including one mega-stokes dye (Dy520) that can be distinguished from the other fluorophores by using TRITC excitation and Cy5 emission filters. The combinations of probe sets used are indicated in the figures. *Fli1* transcripts in C and D appear more yellowish than white and are difficult to distinguish by eye from more orange *Runx1* transcripts (of which there are very few). Probe sets in B,C contain problematic probes that are indicated in Fig. 2; the background from these was suppressed by the use of competitors.

assumptions regarding cell size and shape that have been used previously (Allalou and Wählby, 2009) do not apply. Although it may be possible to delineate cell boundaries using membrane markers, in practice these often do not work well in mesodermal cells and doing so would require an additional fluorophore channel. Instead, we used the observation that transcript density tends to increase with proximity to the nucleus by implementing a nearest-neighbor method that extends clusters of blobs to their nearest neighbor until merging with a nucleus (Fig. 4B, top left). Hence, we first defined nuclear borders and then assigned transcripts to individual nuclei.

The assigned transcripts were used to create two-dimensional cell borders to assess the accuracy of transcript assignment. Although the method produced a plausible transcript assignment (Fig. 4B, top right), an inspection of transcript composition indicated that errors do occur (Fig. 4B, bottom left), and we manually edited borders where necessary. The resulting cell outlines are restricted to two dimensions (though transcripts are mapped in three dimension) and this limits us to analyzing monolayers of cells. However, these observations suggest the use of transcript composition as a means to define cellular units (this being implicitly used in manual editing), though this is complicated by intracellular transcript localization and is by no means straightforward to implement.

Tracking the mesoderm to endothelial transition

To demonstrate the utility of this technology, we addressed a set of questions related to the differentiation of mesoderm to ECs. The phenotype of the *Etv2*-null mutant indicates that ECs are derived from *Etv2*-expressing mesoderm (Lee et al., 2008) and it has been shown that *Etv2* can drive the expression of endothelial genes (De Val et al., 2008; Hayashi et al., 2012). *Etv2* has also been shown to be a direct inducer of *Tal1*, *Fli1* and *Gata2* (Kataoka et al., 2011), which are thought to stabilize the hematopoietic state (De Val and Black, 2009; Pimanda et al., 2007). Hence, the course of gene induction caused by *Etv2* during endothelial differentiation remains unclear.

We derived a mixture of endothelial and non-endothelial cells by differentiating ES cells on collagen type IV (ColIV) as described previously (Yamashita et al., 2000). This differentiation procedure generally results in the appearance of about 50% ECs (as indicated by Cdh5 expression) and 50% smooth muscle actin (Acta2)-expressing cells. These cells were fixed and hybridized with probes against six genes that mark the endothelial and hemangiogenic lineages (*Tal1*, *Runx1*, *Fli1*, *Gata2*, *Cdh5* and *Etv2*). We collected data from three image sets (supplementary material Fig. S1) and quantified transcript numbers (supplementary material Table S2).

Within the 124 cells assayed, total transcript counts varied from 0 to 831 (Fig. 5A). The total number of transcripts varied both with lineage and cell size, with large transcript counts only being observed in large cells. Slightly more than two-thirds of the cells expressed substantial numbers of endothelial transcripts; those that did showed co-expression of all genes, but with low levels of *Runx1* and variable levels of *Etv2*. Somewhat surprisingly, a proportion of non-endothelial cells contained low levels of *Runx1*. No cells contained *Runx1* transcripts at levels comparable with those observed in embryonic cells (data not shown), suggesting that *Runx1* expression in these cells is below functional levels.

Although we observed co-expression of *Gata2*, *Fli1* and *Tal1* with *Cdh5*, the expression levels within cells did not correlate with *Cdh5* expression but rather showed anti-correlation, with high levels of *Cdh5* being associated with lower levels of these transcription factors (Fig. 5B), suggesting that the endothelial state is stabilized by other factors or mechanisms. This anti-correlation was particularly strong for *Etv2*, consistent with its role in the initiation of lateral mesoderm differentiation. By contrast, we observed a strong correlation in expression levels between *Gata2*, *Fli1* and *Tal1*.

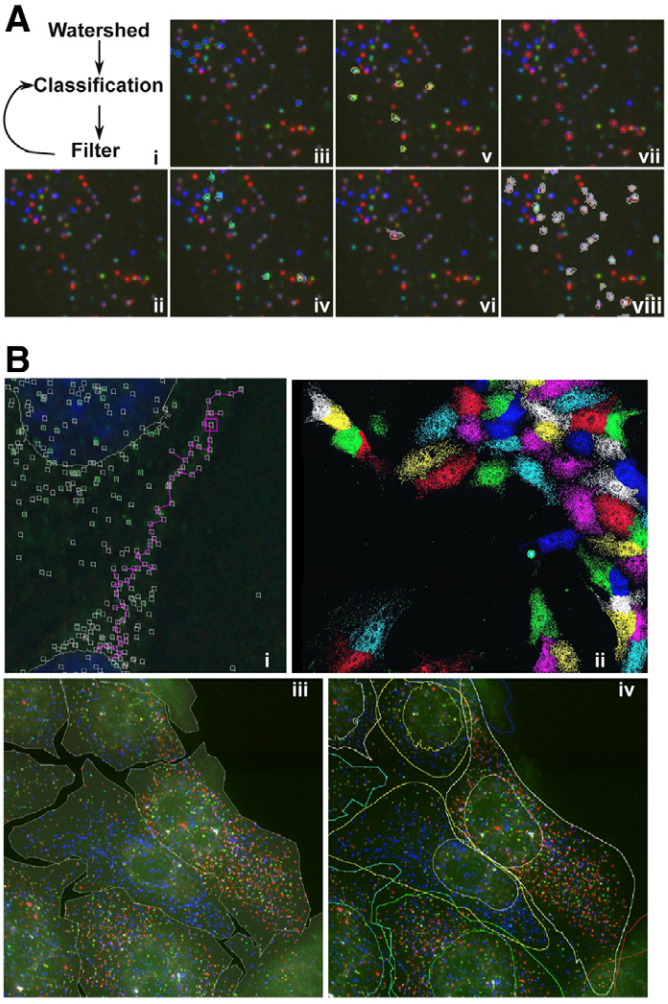


Fig. 4. Transcript and cell segmentation. (**A**) Transcript segmentation. (i) Local three-dimensional maxima (blobs) were first identified using a thresholded watershed algorithm for each fluorescent channel. Individual blobs were then classified into the seven possible identities, depending on the presence of overlapping blobs in additional fluorescent channels. The resulting blob sets were filtered using a set of descriptive parameters (e.g. mean, volume, maximum) using fluorophore and blob set type-specific ranges of values. Only blob sets where all constituent blobs passed the criteria were included in transcript counts. (ii-viii) Maximum intensity projections of a selected area showing a range of different types of transcripts. (ii) Raw image. (iii-viii) Images with the peak level *xy* border area of blobs superimposed: (iii) *Cdh5*, (iv) *Tal1*, (v) *Gata2*, (vi) *Runx1*, (vii) *Etv2* and (viii) *Fli1*. Alexa 488, green; Cy3, blue; Cy5, red). (**B**) Cell segmentation. Cells were segmented by mapping transcripts to nuclei using a nearest neighbor method. (i) Nearest-neighbor cluster growth. A cluster was initiated at the blob indicated by the purple square. This cluster was then grown by including the blob lying closest to a cluster member until a blob overlapping with a nucleus was included. This allows for clusters to grow along gradients of high intensity and is somewhat analogous to a watershed algorithm. (ii) Blobs were assigned to nuclei as described in i, except that cluster growth was also terminated when a previously classified cluster member was included into the growing cluster. Image shows blobs colored by nuclear identity. This method provides a superficially reasonable cell segmentation, but an inspection of transcript composition indicated the need for manual correction (iii). (iv) Cell and nuclear borders after manual correction. Note the presence of overlapping regions; transcripts lying within these regions were excluded from further analysis.

With the exception of *Etv2*, and to some extent *Gata2*, gene expression levels formed a continuum of intensities between the gene maximum and minimum. By contrast, the maximum level of *Etv2* expression was separated from lower levels, with only one or two cells expressing intermediate levels (Fig. 5B; Fig.6C,D). To quantify this, we calculated the distribution of transcript numbers expressed as a proportion of the maximal count of each gene (Fig. 5C). This confirms that the distribution of *Etv2* expression levels is distinct from that of the other genes.

We used an error minimizing algorithm to display the relationships between cells observed in this experiment (Fig. 6A). In order to minimize the impact of cell size, we used transcript densities estimated by dividing transcript counts by cell areas. Given that *Etv2* is required for the appearance of endothelial or hematopoietic cells, we can assume that cells expressing markers for these lineages have expressed *Etv2*; similarly, we can to some extent use *Etv2* expression as an indicator of differentiation stage, as our observations suggest that it is generally expressed as a pulse. This analysis indicates that the population of cells containing appreciable quantities of transcripts can be divided into three or four main components: (1) a single cell expressing high level of *Etv2*; (2) a number of cells expressing low levels of *Runx1*; (3) a main body of cells lying on an apparent axis of endothelial maturation; and (4) a handful of outlier cells characterized by high levels of *Tal1* (Fig. 6A).

It follows from this that we can divide the heterogeneity within the population into lineage and chronological (or extent of differentiation) heterogeneity, with the majority of the population proceeding along an axis of endothelial differentiation starting from *Etv2* high/[*Tal1*, *Fli1*, *Gata2* (TF)] low/*Cdh5* low and proceeding to an intermediate *Etv2* low/TF high/*Cdh5* intermediate and finally to an $Etv2^{-}$/TF low/*Cdh5* high identity.

To provide a quantitative view of the change in expression along this axis, we defined a vector of differentiation running from cell 20 (*Etv2* high) to cell 45 (*Cdh5* high) within the reduced dimensionality space (Fig. 6A). Cells were mapped onto this vector and filtered by their distance from it to include the main endothelial population (within parallel lines in Fig. 6A). Means and standard deviations of a sliding window (width, 0.2; slide increments, 0.005; where 1 is the full length of the vector) were then calculated (Fig. 6C). Expression of *Cdh5* increased concomitantly with a decrease in *Etv2*, *Tal1*, *Fli1* and *Gata2*. To visualize the variance in expression along the vector, we plotted individual transcript densities versus position (Fig. 6D). This shows a homogeneous linear increase in Cdh5 expression along the vector, with the decrease in expression of Fli1 and Tal1 being characterized by a larger variance.

In addition to the main population, there are at least two cells that seem to mature (i.e. low *Etv2*, but high *Tal1* and *Fli1*) without activation of *Cdh5*. However, as these cells are rare, it is difficult to determine their relationship to the main endothelial population. However, we can refer back to the original image data sets to confirm that these data points are not due to measurement artefacts. Indeed, these cells not only show a difference in their transcriptional identity, but also in their shape: having a less flattened morphology than typical (Fig. 6B).

In the above analysis, we have inferred a chronological heterogeneity on the basis of *Etv2* and *Cdh5* expression. In order to confirm this, we carried out FISH at 6, 15, 24 and 48 hours after replating in VEGF containing SFO3 medium. In order to exclude the possibility that the observed anti-correlation between *Fli1* and *Cdh5* was affected by misclassification of *Fli1* transcripts due to high concentrations of *Cdh5* transcripts (which share one

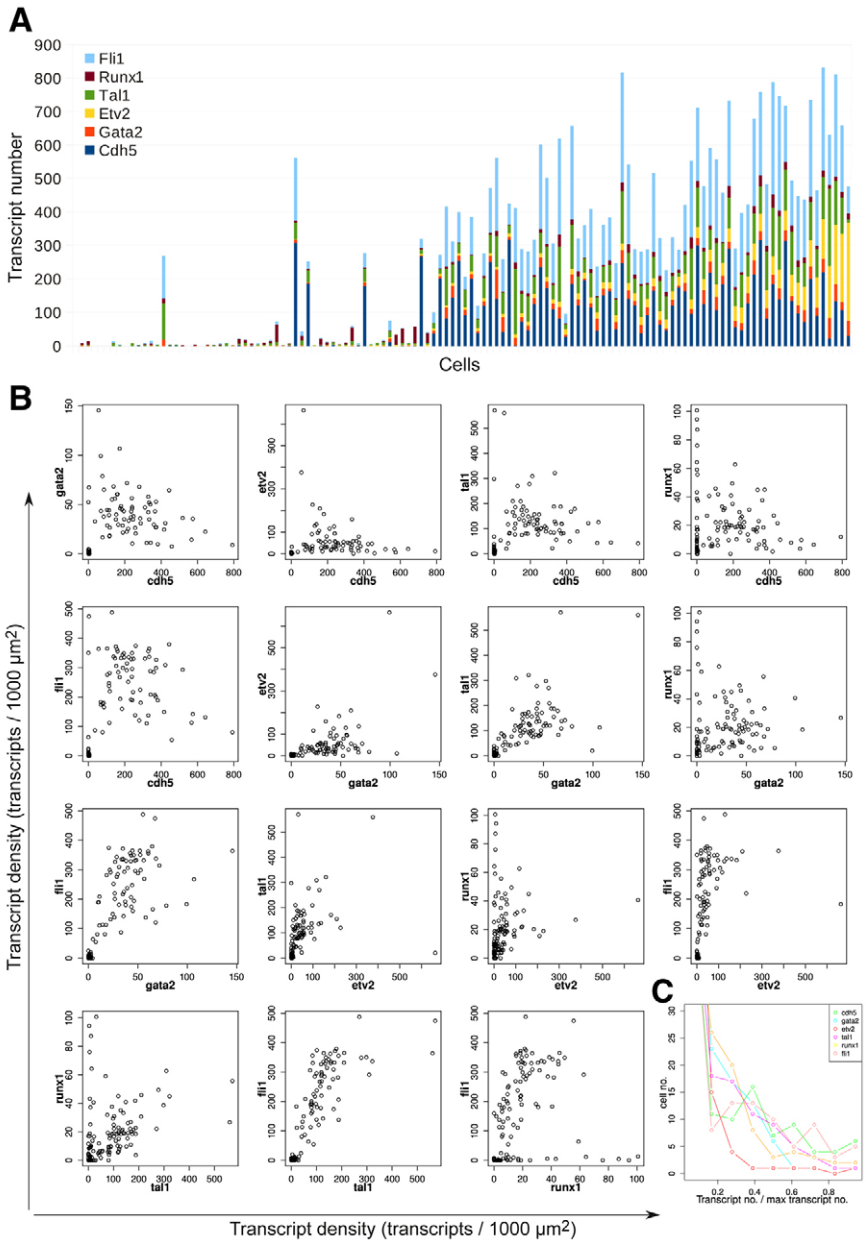


Fig. 5. Transcript counts. FISH against *Cdh5*, *Gata2*, *Etv2*, *Tal1*, *Runx1* and *Fli1* was performed on ES-derived mesoderm cells differentiated for 24 hours in the presence of VEGF. Transcripts were then enumerated and associated with cells as described in Fig. 4. (**A**) Counts of transcripts per cell: *x*-axis indicates cell IDs; *y*-axis indicates transcript counts for the indicated genes. Total transcript numbers ranged from 0 to 831, with *Cdh5* and *Fli1* being the most highly expressed genes. (**B**) Pairwise plots of transcript densities for individual genes. Transcript counts were normalized by cell area and plotted against each other. *Cdh5* shows an anti-correlation with most of the genes in endothelial cells, whereas a clear correlation of expression was observed for *Gata2*, *Fli1* and *Tal1*. (**C**) Distribution of transcript counts expressed as a proportion of individual maxima; *y*-axis indicates the number of cells containing the number of transcripts indicated on the *x*-axis. The plot has been truncated along the *y*-axis to highlight differences at intermediate levels.

fluorophore), we used non-combinatorial FISH targeted against *Etv2*, *Fli1* and *Cdh5* (Fig. 6E).

Six hours after VEGF exposure, we observed a small number of cells containing a range of *Etv2* transcript numbers, some of which contained very low levels of *Fli1* or *Cdh5*. At 16 hours, *Etv2*-positive cells were common, and most of those also contained *Fli1* and *Cdh5* transcripts. At 24 hours, we observed a similar situation to that described above: most cells contained a mixture of *Fli1*, *Etv2* and *Cdh5*; there was a small number of *Etv2* high cells; and *Cdh5* high cells contained lower levels of *Fli1*. At 48 hours, most cells in the endothelial lineage contained very high densities of *Cdh5* transcripts but much lower levels of *Fli1*. Very few *Etv2* transcripts were observed at this time.

DISCUSSION

Here, we demonstrate the feasibility of using combinatorial FISH to simultaneously measure RNA abundance from more than six genes in single cells. We have used this to analyze how cellular identities change during the differentiation of primitive mesoderm to an endothelial identity. This is a well-characterized process that has been studied in detail, resulting in the identification of a small number of essential genes [*Vegfa* (Carmeliet et al., 1996), *Kdr* (Shalaby et al., 1995) and *Etv2* (Kataoka et al., 2011; Lee et al., 2008)] and a large number of genes that are activated and or repressed during the process. However, our previous attempts to describe the process based on microarray data of sorted populations were frustrated by the inability to determine whether observed co-induction of genes was occurring within individual cells. To test the possibility of using combinatorial FISH to clarify this situation, we concentrated on six genes known to be involved in the development of the endothelial and hematopoietic lineages. *Etv2* is absolutely required for the appearance of both lineages (Lee et al., 2008), though its expression both in embryos and *in vitro* seems to be limited to a brief window associated with the transition from a primitive to lateral mesoderm identity. *Tal1*, *Fli1* and *Gata2* appear to be crucial targets of *Etv2*, and their

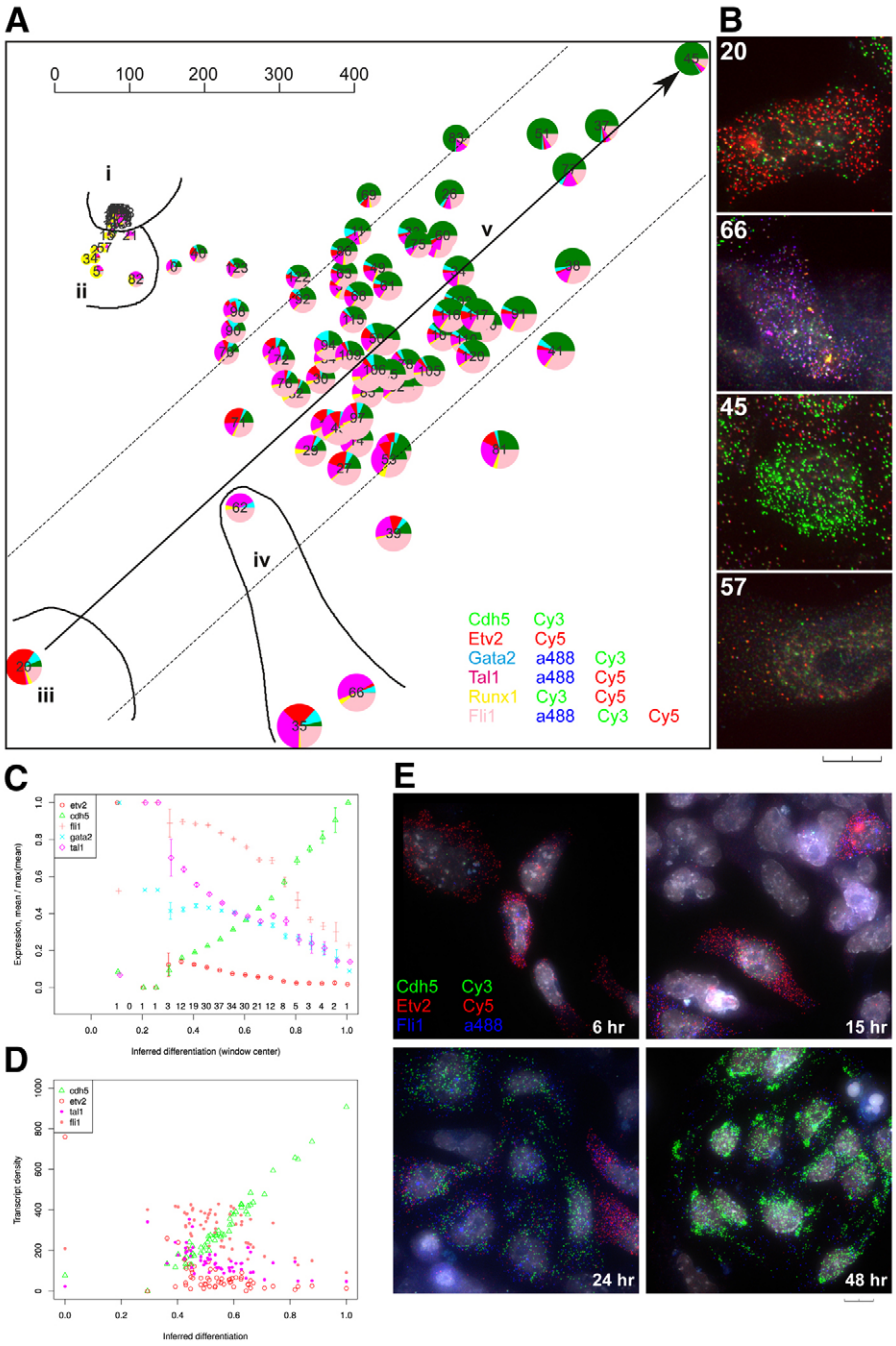


Fig. 6. Cell identities. (**A**) Nodes representing individual cells were arranged in a two-dimensional space using a heuristic that attempts to maintain cell-cell relationships based on cell expression profiles. Each pie-chart represents a single cell. Distances between charts indicate euclidean distances between cells calculated using transcript densities; scale bar indicates distance in units of transcripts per 1000 μm^2, as in Fig. 5B. Pie segment areas are related to the total transcript density (transcripts/unit area) and the segment sizes are proportional to the individual gene transcript densities. Five main cell identities can be discerned and are indicated in the figure as: (i) cells expressing very low or no transcripts, (ii) cells expressing primarily low levels of *Runx1*, (iii) a single *Etv2* high cell, (iv) a small number of cells containing high levels of *Tal1* but low levels of *Cdh5* and (v) a main field of cells expressing variable levels of all the genes assayed. The primary axis of differentiation within this population (arrow) of cells seems to extend from the lower-left to the upper-right corner, with a graded loss of *Etv2* being accompanied first by an increase in *Cdh5* and endothelial transcription factor expression, followed by a further increase in *Cdh5* and loss of *Gata2*, *Tal1* and *Fli1* expression. Arrow connecting cells 20 and 45, and lines parallel to it, indicate vector and filtering used in C. (**B**) Images of cells taking extreme positions in A. Numbers indicate cell IDs as shown in A. Alexa 488, blue; Cy3, green; Cy5, red. (**C**) Change in expression along the inferred route of endothelial maturation. Cells were projected onto a vector running between cells 20 and 45. Mean transcript densities for cells lying between the parallel lines in A were calculated for a sliding window running along the vector. Error bars indicate standard errors; numbers along the *x*-axis indicate the number of cells in each window (each cell is represented in five window positions). (**D**) Plot of individual transcript densities versus inferred differentiation for *Cdh5*, *Etv2*, *Tal1* and *Fli1* (each vertical stack of points represent a single cell). Transcript densities are given as transcripts per 1000 μm^2. (**E**) ES cells were differentiated on collagen in the presence of serum and then replated in serum-free medium with VEGF for the indicated times. FISH was performed in a non-combinatorial manner against *Cdh5*, *Etv2* and *Fli1*.

expression overlaps with that of *Etv2* in early embryos (Kataoka et al., 2011). *Cdh5* encodes a cell-adhesion molecule that is crucial for endothelial function though dispensable for endothelial differentiation (Gory-Fauré et al., 1999). *Runx1* is crucial for definitive hematopoiesis and is expressed in lateral mesoderm (Sakurai et al., 2006). Expression in endothelial or endothelial descendants seems to be necessary for the appearance of hematopoietic stem cells (Chen et al., 2009). However, data from *in vitro* differentiation suggests that *Runx1* is not directly regulated by *Etv2* (Kataoka et al., 2011).

Hence, we would expect that *Etv2* activation is followed by expression of *Tal1*, *Fli1* and *Gata2*, followed by *Cdh5*, with a gradual reduction of *Etv2* coinciding with an increase in *Cdh5*. Indeed, this is almost exactly what we observe. Furthermore, as we can observe a continuum (albeit interrupted) of cell identities from a presumably primitive (*Etv2* high, others low) to a more differentiated state (*Cdh5* high), we can infer this route of differentiation from a single timepoint.

Since expression of *Cdh5* was observed in the least differentiated cells (i.e. high *Etv2*) that only contained low levels of *Fli1*, *Tal1* and *Gata2*, we can surmise that *Cdh5* is also a direct target of *Etv2*; although as its expression is increased during the process, control must pass to other factors. Under these conditions *Tal1*, *Fli1* and *Gata2* seem to behave in a similar manner to *Etv2* in that their expression seems to be shortlived and lost upon endothelial maturation. Somewhat pleasingly, we see a good correlation in expression between *Tal1*, *Fli1* and *Gata2*. These three genes are thought to form a positively reinforcing triad motif that stabilizes their expression in hematopoietic precursors (Pimanda et al., 2007). This is consistent with our observations; however, as their expression is lost during endothelial maturation it seems that the correlation may be due not to internal feedback within the loop but due to common external regulation.

Although we observed a gradual reduction in levels of *Fli1, Tal1* and *Gata2* with an increase in *Cdh5* expression (Fig. 5C; Fig. 6A,C,D), the reduction in *Etv2* levels drops rapidly, suggesting that the high *Etv2* state is unstable and rapidly resolves itself to an *Etv2* low state. Hence, although these transcription factors (*Etv2*, *Tal1*, *Fli1* and *Gata2*) are all expressed as a pulse during the differentiation process, the manner in which expression is lost differs, suggesting different modes of regulation.

In contrast to previous expectation (Levsky and Singer, 2003), our data suggest that under these conditions, endothelial differentiation is remarkably homogeneous. All cells expressing *Cdh5* (marking this lineage) also contain appreciable levels of all the other factors, and their transcript densities seem to be related to the extent of differentiation. Outside of this main population we also observe around 20 cells that contain no, or very low levels of, these transcripts; and we also observe two or three cells with high levels of *Tal1* and *Fli1*. This low level of hetereogeneity is probably, partly at least, caused by the use of defined conditions (no serum) and a two-dimensional culture system that can provide cells with a homogeneous environment (Nishikawa et al., 2007).

These data suggest that the process of endothelial maturation follows a consistent course that does not appear to be under much stochastic influence; variance within the endothelial population appears to be caused by differences in the timing of lineage entry. By contrast, lineage entry appears to be stochastic, suggesting that the processes of lineage choice and maturation are governed by different regulatory mechanisms. However, this cannot be concluded from our present data as we have little information regarding the heterogeneity of the initiating population.

Taken together, the data presented here show not only that combinatorial FISH can be used to quantitate transcription from multiple genes at the single cell level, but also that this can be sufficient to track the change of cellular identity during differentiation processes. However, we have only been able to implement these methods in monolayer cultures owing to the difficulty of accurately segmenting cells in three dimensions and the increased autofluorescence associated with thicker tissues. Although we have been able to visualize individual transcripts in whole-mount embryos (mouse E7.5-8.0) using longer wavelength emitting dyes (Cy3/Cy5), the signal-to-noise ratio has not been sufficient to allow for combinatorial detection. This problem can be overcome by disassociating cells followed by immobilization on slides for microscopy. Though this will result in the loss of spatial information, this can to some extent be abrogated by performing FISH on cells sorted by fluorescence cytometry. Recent developments (Cella Zanacchi et al., 2011) in light-sheet microscopy (Greger et al., 2007) may be sufficient to make the methods outlined here applicable to thicker tissues. Such developments should also abrogate the problem of 3-D cell segmentation due to the associated increase in resolution in the Z-axis. Our approach is limited by the number of fluorophores and the imaging resolution; however, Lubeck and Cai (Lubeck and Cai, 2012) have recently demonstrated the detection of transcripts from 32 genes in yeast using switchable fluorophores and super-resolution microscopy. Although the imaging used by Lubeck and Cai (Lubeck and Cai, 2012) is currently only suitable for very thin (~100 nm) samples due to relying on near-field microscopy we believe this will change in the future and that the number of transcripts that can be targeted will drastically increase.

Acknowledgements
We thank Paul O'Neill for help with revising this manuscript and Yuko Kiyosue for microscopy support.

Funding
The authors and the majority of the work were supported by Riken CDB intramural funding. Part of this work was supported by Grants-in-Aid for Scientific Research (S) (20229005) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests statement
The authors declare no competing financial interests.

Supplementary material
Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.086975/-/DC1

References

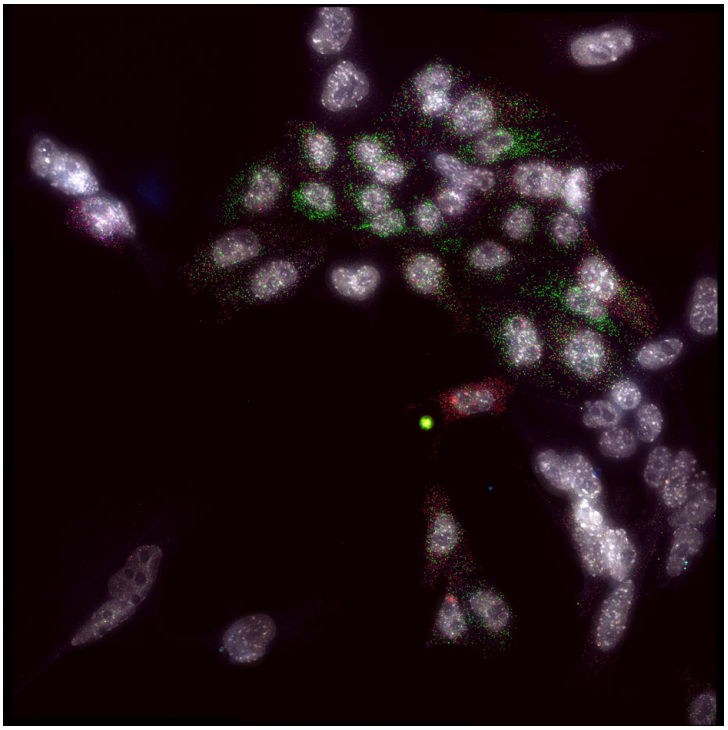
Allalou, A. and Wählby, C. (2009). BlobFinder, a tool for fluorescence microscopy image cytometry. *Comput. Methods Programs Biomed.* **94**, 58-65.
Bendall, S. C., Simonds, E. F., Qiu, P., Amir, A. D., Krutzik, P. O., Finck, R., Bruggner, R. V., Melamed, R., Trejo, A., Ornatsky, O. I. et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687-696.
Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C. et al. (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435-439.
Cella Zanacchi, F., Lavagnino, Z., Perrone Donnorso, M., Del Bue, A., Furia, L., Faretta, M. and Diaspro, A. (2011). Live-cell 3D super-resolution imaging in thick biological samples. *Nat. Methods* **8**, 1047-1049.
Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. and Speck, N. A. (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887-891.
De Val, S. and Black, B. L. (2009). Transcriptional control of endothelial cell development. *Dev. Cell* **16**, 180-195.
De Val, S., Chi, N. C., Meadows, S. M., Minovitsky, S., Anderson, J. P., Harris, I. S., Ehlers, M. L., Agarwal, P., Visel, A., Xu, S. M. et al. (2008). Combinatorial

regulation of endothelial gene expression by ets and forkhead transcription factors. *Cell* **135**, 1053-1064.
Femino, A. M., Fay, F. S., Fogarty, K. and Singer, R. H. (1998). Visualization of single RNA transcripts in situ. *Science* **280**, 585-590.
Gory-Fauré, S., Prandini, M. H., Pointu, H., Roullot, V., Pignot-Paintrand, I., Vernet, M. and Huber, P. (1999). Role of vascular endothelial-cadherin in vascular morphogenesis. *Development* **126**, 2093-2102.
Greger, K., Swoger, J. and Stelzer, E. H. K. (2007). Basic building units and properties of a fluorescence single plane illumination microscope. *Rev. Sci. Instrum.* **78**, 023705.
Hayashi, K., Lopes, S. M., Tang, F. and Surani, M. A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391-401.
Hayashi, M., Pluchinotta, M., Momiyama, A., Tanaka, Y., Nishikawa, S. and Kataoka, H. (2012). Endothelialization and altered hematopoiesis by persistent etv2 expression in mice. *Exp. Hematol.* **40**, 738-750.e11.
Itzkovitz, S. and van Oudenaarden, A. (2011). Validating transcripts with probes and imaging technology. *Nat. Methods* **8**, S12-S19.
Kataoka, H., Hayashi, M., Nakagawa, R., Tanaka, Y., Izumi, N., Nishikawa, S., Jakt, M. L., Tarui, H. and Nishikawa, S. (2011). Etv2/ER71 induces vascular mesoderm from Flk1+PDGFRα+ primitive mesoderm. *Blood* **118**, 6975-6986.
Lee, D., Park, C., Lee, H., Lugus, J. J., Kim, S. H., Arentson, E., Chung, Y. S., Gomez, G., Kyba, M., Lin, S. et al. (2008). ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. *Cell Stem Cell* **2**, 497-507.
Levsky, J. M. and Singer, R. H. (2003). Gene expression and the myth of the average cell. *Trends Cell Biol.* **13**, 4-6.
Levsky, J. M., Shenoy, S. M., Pezo, R. C. and Singer, R. H. (2002). Single-cell gene expression profiling. *Science* **297**, 836-840.
Lubeck, E. and Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743-748.
Nagaoka, M., Koshimizu, U., Yuasa, S., Hattori, F., Chen, H., Tanaka, T., Okabe, M., Fukuda, K. and Akaike, T. (2006). E-cadherin-coated plates maintain pluripotent ES cells without colony formation. *PLoS ONE* **1**, e15.
Nishikawa, S., Jakt, L. M. and Era, T. (2007). Embryonic stem-cell culture as a tool for developmental cell biology. *Nat. Rev. Mol. Cell Biol.* **8**, 502-507.
Pimanda, J. E., Ottersbach, K., Knezevic, K., Kinston, S., Chan, W. Y. I., Wilson, N. K., Landry, J. R., Wood, A. D., Kolb-Kokocinski, A., Green, A. R. et al. (2007). Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development. *Proc. Natl. Acad. Sci. USA* **104**, 17692-17697.
Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309.
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.
Sakurai, H., Era, T., Jakt, L. M., Okada, M., Nakai, S., Nishikawa, S. and Nishikawa, S. (2006). In vitro modeling of paraxial and lateral mesoderm differentiation reveals early reversibility. *Stem Cells* **24**, 575-586.
Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* **473**, 337-342.
Shalaby, F., Rossant, J., Yamaguchi, T. P., Gertsenstein, M., Wu, X. F., Breitman, M. L. and Schuh, A. C. (1995). Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62-66.
Yamashita, J., Itoh, H., Hirashima, M., Ogawa, M., Nishikawa, S., Yurugi, T., Naito, M., Nakao, K. and Nishikawa, S. (2000). Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature* **408**, 92-96.
Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523.

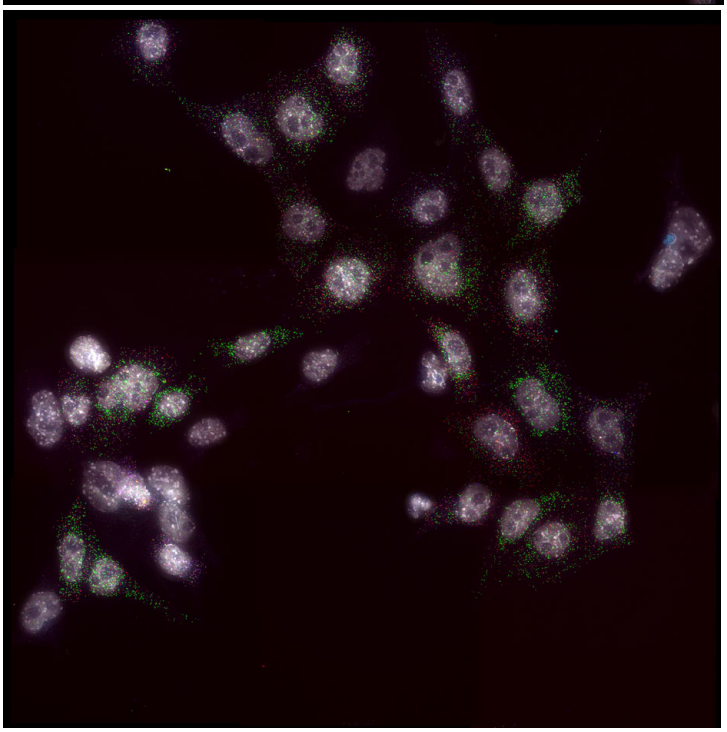
Supplementary Images

Maximum intensity projections of the three image sets (panels 1-3) used to derive transcript counts summarised in main figures 5 and 6, and in supplementary table 2. Image set ids are denoted in supplementary table 2. CCE ES cells were differentiated for four days in aMEM 10% serum on Collagen type IV (Col IV). At day four, 20,000 cells were re-plated in a 6mm well on a Col IV coated glass slide and re-cultured for a further 24 hours in serum free medium (SFO3) and VEGF (30 ng/ml). The cells were then hybridised with probes against Cdh5 (Cy3), Gata2 (a488*, Cy3), Etv2 (Cy5), Tal1 (a488, Cy5), Runx1 (Cy3, Cy5) and Fli1 (a488, Cy3, Cy5) as described in the main text. Colours are: DAPI white, a488 blue, Cy3 green, Cy5 red, resulting in the following labelling: Cdh5 green, Gata2 cyan, Etv2 red, Tal1 magenta, Runx1 orange (~ish) and Fli1 whitish. Note that single coloured labelled transcripts give much stronger signals (eg, Cdh5 and Etv2) which tend to obscure the lower intensity combinatorially labelled transcripts when viewed by eye. This is especially true for Gata2 since the number of Gata2 transcripts is generally low. The images were derived from overlapping z-stacks merged using the dvreader application (http://www.gitorious.org/dvreader). Pixel dimensions are 107 x 107 nm. All three images were viewed with the same viewing parameters (i.e. brightness / contrast). *a488 = Alexa fluor 488. Original data and non-compressed tiff files of the projections are available from the authors upon request.

1



2



3