nature cell biology

# Early lineage restriction in temporally distinct populations of *Mesp1* progenitors during mammalian heart development

**Fabienne Lescroart[1,8], Samira Chabab[1,8], Xionghui Lin[1], Steffen Rulands[2,3], Catherine Paulissen[1], Annie Rodolosse[4], Herbert Auer[4], Younes Achouri[5], Christine Dubois[1], Antoine Bondue[1,6], Benjamin D. Simons[2,3] and Cédric Blanpain[1,7,9]**

**Cardiac development arises from two sources of mesoderm progenitors, the first heart field (FHF) and the second (SHF). *Mesp1* has been proposed to mark the most primitive multipotent cardiac progenitors common for both heart fields. Here, using clonal analysis of the earliest prospective cardiovascular progenitors in a temporally controlled manner during early gastrulation, we found that *Mesp1* progenitors consist of two temporally distinct pools of progenitors restricted to either the FHF or the SHF. FHF progenitors were unipotent, whereas SHF progenitors were either unipotent or bipotent. Microarray and single-cell PCR with reverse transcription analysis of *Mesp1* progenitors revealed the existence of molecularly distinct populations of *Mesp1* progenitors, consistent with their lineage and regional contribution. Together, these results provide evidence that heart development arises from distinct populations of unipotent and bipotent cardiac progenitors that independently express *Mesp1* at different time points during their specification, revealing that the regional segregation and lineage restriction of cardiac progenitors occur very early during gastrulation.**

The mammalian heart is the first functional organ that forms during embryonic development and is composed of cardiomyocytes (CMs), endothelial cells (ECs), epicardial derived cells (EPDCs) and smooth muscle cells[1] (SMCs). Cardiac development arises from two sources of mesoderm progenitors, the first heart field (FHF) and the second heart field[2,3] (SHF). Retrospective clonal analysis suggests the existence of a common progenitor for both heart fields, although the timing of the lineage segregation remains unclear[3]. *Mesp1* is the earliest known marker of cardiac progenitors[4,5]. Overexpression of *Mesp1* in embryonic stem cells[6–9] (ESCs) suggests that Mesp1 promotes the specification of the most primitive multipotent cardiac progenitors[7]. Lineage tracing using *Mesp1–Cre* knock-in mice showed also that almost all myocardial cells, including derivatives of the FHF and SHF, derive from *Mesp1*-expressing progenitors[4]. However, lineage tracing using *Mesp1–Cre* at the population level does not allow the assessment of whether FHF and SHF progenitors arise from a common progenitor or whether *Mesp1* is expressed independently in distinct cardiac progenitors. To identify the developmental origin of organ regionalization and the timing of lineage segregation, it is essential to perform temporal clonal labelling in prospective progenitors[10].

One of the key questions in mammalian development is the timing with which the progenitor becomes specified to differentiate into their different lineages. During chick heart development, it has been initially proposed that cardiac and vascular lineage could be already pre-specified at the early stage of gastrulation[11,12]. In contrast, subsequent genetic lineage tracing *in vivo* and clonal differentiation of cardiovascular progenitors *in vitro* support the notion that, during mouse embryonic development, cardiovascular progenitors remain multipotent until the later stages of cardiogenesis at the time where they begin to express transcription factors such as Nkx2-5 and Isl1 (refs 6,7,13–15). So far, no study has assessed the fate of prospective mouse cardiovascular progenitors into the different cardiovascular lineages using single-cell marking *in vivo*.

## RESULTS

### Doxycycline-inducible *Mesp1* reporter and Cre -mediated recombination

To assess the contribution of single *Mesp1*-expressing progenitors at different time points during embryonic development, we generated a tetracycline-inducible *Mesp1–rtTA* transgenic mouse, in which

[1]Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium. [2]Cavendish Laboratory, Department of Physics, J. J. Thomson Avenue, Cambridge CB3 0HE, UK. [3]The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. [4]Functional Genomics Core, Institute for Research in Biomedicine, Barcelona 08028, Spain. [5]Université Catholique de Louvain, de Duve Institute, Brussels B-1200, Belgium. [6]Department of Cardiology, Hopital Erasme, Brussels B-1070, Belgium. [7]WELBIO, Université Libre de Bruxelles, Brussels B-1070, Belgium. [8]These authors contributed equally to this work. [9]Correspondence should be addressed to C.B. (e-mail: Cedric.Blanpain@ulb.ac.be)

Received 31 March 2014; accepted 10 July 2014; published online 24 August 2014; DOI: 10.1038/ncb3024

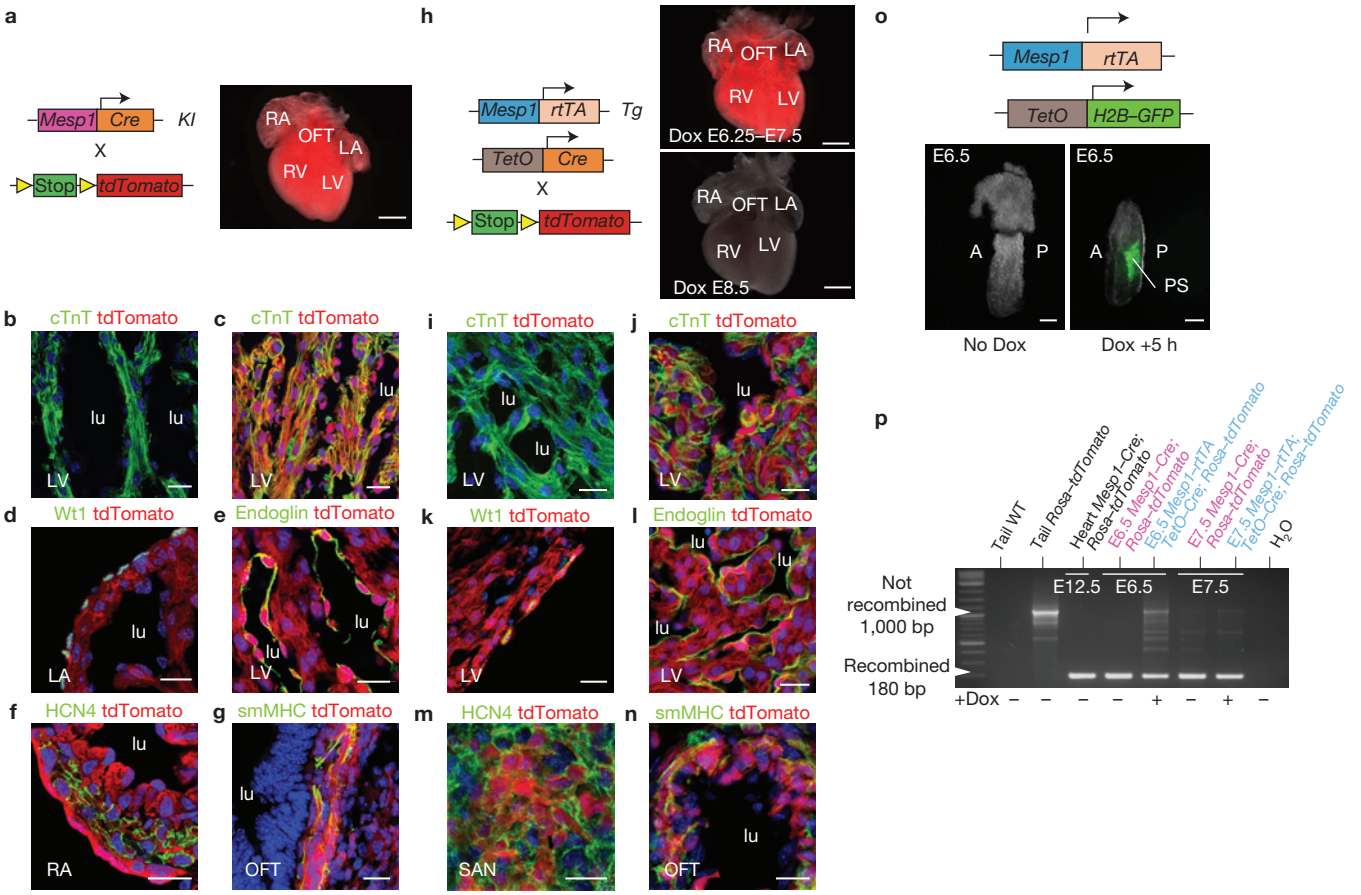



**Figure 1** *Mesp1–rtTA* transgenic mice faithfully recapitulate *Mesp1* endogenous expression. (**a**) Macroscopic analysis of a *Mesp1–Cre/Rosa–tdTomato* embryo at E14.5. LV: left ventricle, RV: right ventricle, LA: left atrium, RA: right atrium, OFT: outflow tract. Scale bar, 500 μm. (**b**,**c**) Confocal analysis of *Rosa–tdTomato* (**b**) and *Mesp1–Cre/Rosa–tdTomato* heart sections (**c**) at E14.5 co-stained with anti-cardiac troponin T (cTnT) antibody. (**d**–**g**) Confocal analysis of *Mesp1–Cre/Rosa–tdTomato* heart sections at E14.5 co-stained with epicardial (Wt1; **d**), EC (endoglin; **e**), pace-maker (Hcn4; **f**) and SMC (smMHC; **g**) markers. lu: lumen. Scale bars, 20 μm. (**h**) Scheme of the genetic strategy used for the characterization of the *Mesp1–rtTA* transgenic mice. Dox administration leads to the activation of the Cre recombinase between E6.25 and E7.5 in *Mesp1–rtTA/TetO–Cre/Rosa–tdTomato* embryos but no activation of the Cre recombinase was detected when Dox was administrated later (E8.5). Scale bars, 500 μm. (**i**,**j**) Confocal analysis of *Rosa–tdTomato* (**i**) and *Mesp1-rtTA/TetO-Cre/Rosa-tdTomato* heart sections (**j**) at E14.5 co-stained with anti-cardiac troponin T (cTnT). (**k**–**n**) Confocal analysis of *Mesp1–rtTA/TetO–Cre/Rosa–tdTomato* heart sections at E14.5 co-stained with epicardial (Wt1; **k**), EC (endoglin; **l**), pace-maker (Hcn4; **m**) and SMC (smMHC; **n**) markers. SAN: sino-atrial node. Scale bars, 20 μm. (**o**) Temporal analysis of the activation of the *Mesp1–rtTA* transgene. Whereas GFP expression was not induced in embryos in the absence of Dox, GFP$^+$ cells could be detected only 5 h after Dox injection in the primitive streak (PS) and nascent mesoderm. Scale bars, 100 μm. A: anterior, P: posterior. (**p**) Temporal analysis of the recombination of the *Rosa–tdTomato* locus investigated by PCR following Dox administration. The *Rosa–tdTomato* locus was recombined as soon as 6 h following Dox administration in *Mesp1–rtTA/TetO–Cre/Rosa–tdTomato* embryos at E6.25 and E7.25, as found with *Mesp1–Cre/Rosa–tdTomato* embryos at the same time points. Negative controls including WT tail and *Rosa–tdTomato* tail show PCR amplification corresponding to the unrecombined *Rosa–tdTomato* locus (around 1,000 bp) and *Mesp1–Cre/Rosa–tdTomato* hearts at E12.5 (positive control) show the recombined *Rosa–tdTomato* locus (about 180 bp).

the doxycycline (Dox)-dependent transactivator (*Mesp1–rtTA*) is expressed under the control of a fragment of the *Mesp1* promoter expressed in cardiac progenitors during mouse embryonic development and ESC differentiation[7,16] (Fig. 1). We identified 6 *Mesp1–rtTA* founders that produce embryos with faithful expression of tdTomato in the heart when Dox was administrated to *Mesp1-rtTA/TetO-Cre/Rosa-tdTomato* embryos between embryonic day (E)6.25 and E7.5, corresponding to the timing of endogenous *Mesp1* expression[4,17]. The expression of the tdTomato was similar to that found in the *Mesp1–Cre/Rosa–tdTomato* embryos (Fig. 1a,h), indicating that the *Mesp1–rtTA* transgene targets the same cells as in *Mesp1–Cre* knock-in. Dox administration during the later stage of cardiac development in *Mesp1-rtTA/TetO-Cre/Rosa-tdTomato* embryos after E8.0 did not induce *Mesp1-rtTA/TetO-Cre/Rosa-tdTomato* expression, consistent with the transient expression of *Mesp1* during the early step of cardiovascular progenitor specification[4] (Fig. 1h). Finally, Dox administration to *Mesp1-rtTA/TetO-Cre/Rosa-tdTomato* embryos leads to the same labelling of all cardiovascular cell types of the FHF and SHF such as CMs, conduction cells, endocardial cells and EPDCs (Fig. 1a–n), with the exception of some unlabelled SMCs in the SHF deriving from the neural crest[18] (Supplementary Fig. 1a,b).

To assess the temporal activation of the *Mesp1–rtTA* transgene on Dox administration, we administrated Dox to

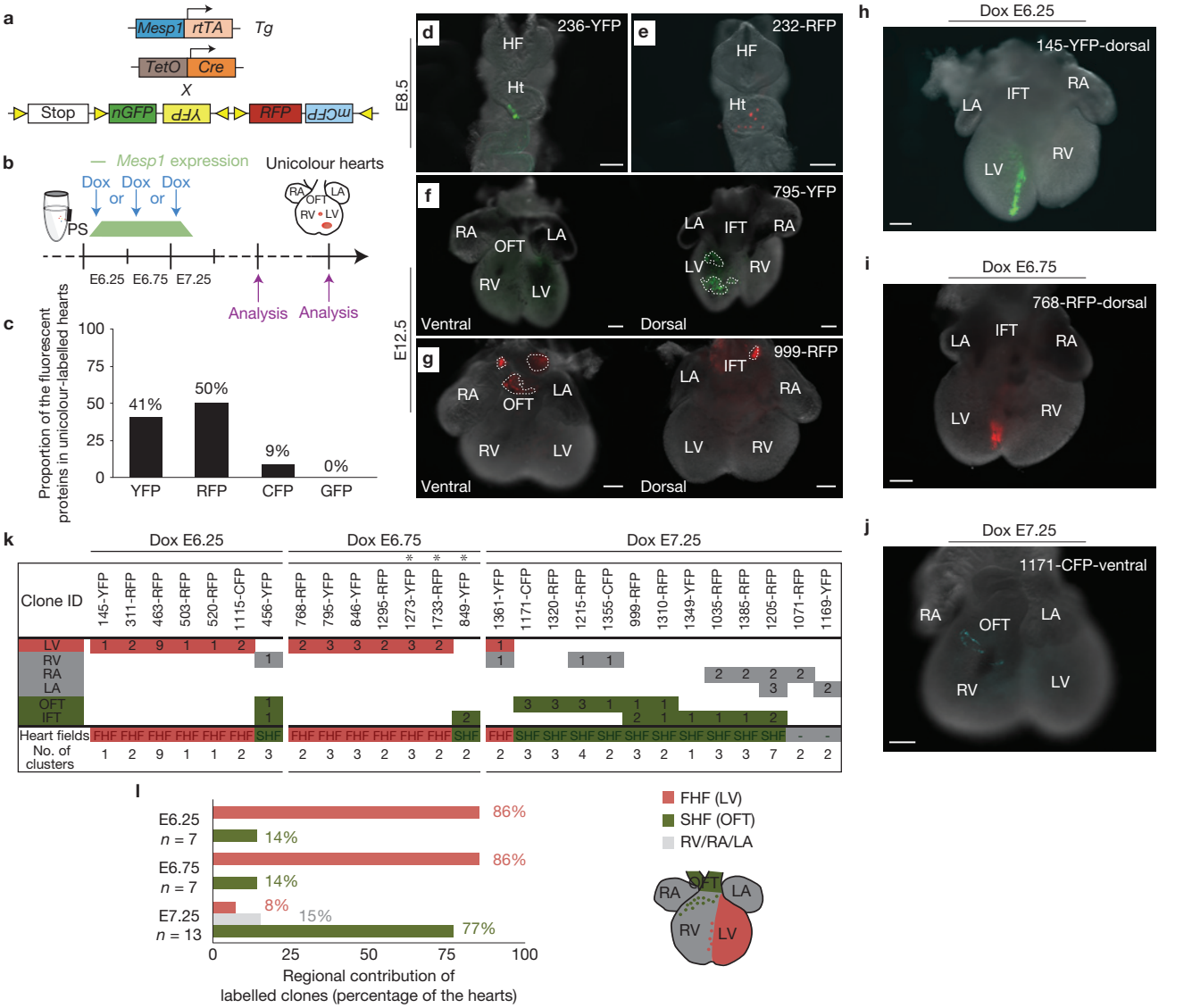



**Figure 2** Two temporally distinct populations of *Mesp1* progenitors contribute to the development of the FHF and SHF. (**a**) Scheme of the genetic strategy used for the clonal tracing of *Mesp1*-expressing progenitors with different fluorescent proteins to assess their regional contribution. (**b**) A low dose of Dox was injected between E6.25 and E7.25. Induced *Mesp1–rtTA/TetO–Cre/Rosa–Confetti* unicolour embryos were analysed at E8.5 and E12.5. PS: primitive streak, RA: right atrium, LA: left atrium, OFT: outflow tract, RV: right ventricle, LV: left ventricle. (**c**) Proportion of the fluorescent proteins in unicolour-labelled hearts. ($n=7$ unicolour hearts at E8.5 and $n=37$ unicolour hearts at E12.5). (**d**,**e**) Examples of IFT: inflow tract unicolour-labelled hearts at E8.5. HF: head folds, Ht: heart. (**f**,**g**) Examples of *Mesp1–rtTA/TetO–Cre/Rosa–Confetti* unicolour-labelled hearts at E12.5. Note that each patch is localized within either the FHF or the SHF but no unicolour patches that encompassed derivatives of the FHF and the SHF were observed. IFT: inflow tract. Scale bars, 200 μm. (**h–j**) Examples of E12.5 unicolour hearts induced at E6.25 (**h**) and E6.75 (**i**) showing the labelling of FHF-derived progenitors, whereas Dox administration at E7.25 shows preferential labelling of SHF progenitors (**j**). Scale bars, 200 μm. (**k**) Graph depicting in all unicolour hearts the regional contribution of the labelled cells and the number of clusters of labelled cells per chamber according to the developmental time of Dox administration. Asterisks indicate that labelling was also detected in the epicardial layer. (**l**) Quantification of the regional (FHF and SHF) contribution of patches of *Mesp1*$^{+}$ labelled cells in unicolour hearts shows preferential labelling of the FHF (red) during Dox administration at early time points (E6.25 and E6.75), whereas Dox administration in the late stage of cardiac progenitor specification (E7.25) shows preferential labelling of *Mesp1* progenitors that contribute to the SHF (green) derivatives. The number on the upper right in each panel refers to the ID of the labelled heart.

*Mesp1-rtTA/TetO-H2B-GFP* mice at E6.25, when *Mesp1* begins to be expressed[4,17]. Already at 5 h following Dox administration, H2B–GFP was detectable in the primitive streak and the nascent cardiac mesoderm (Fig. 1o), in a similar pattern to that previously reported for *Mesp1–LacZ* knock-in mice[4,17]. *In situ* hybridization revealed that *Mesp1* and *Cre* were expressed at the same location in *Mesp1–Cre* knock-in and *Mesp1-rtTA/TetO-Cre* embryos treated with Dox (Supplementary Fig. 1f–h). PCR analysis showed that the Rosa–*tdTomato* locus was recombined, as early as 6 h following Dox administration at E6.25 and E7.25, similar to the case for *Mesp1–Cre* knock-in embryos (Fig. 1p). All of these experiments indicate that Dox administration to *Mesp1-rtTA/TetO-Cre* embryos targets cardiovascular progenitors of both heart fields and faithfully recapitulates *Mesp1–Cre* knock-in mice.

### Two temporally distinct populations of *Mesp1* progenitors contribute to the FHF and SHF development

To investigate the contribution of a single *Mesp1*-expressing cell, we titrated the dose of Dox required to label *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* hearts at clonal density, as defined by the dose of Dox allowing the recombination of a single fluorescent protein per heart and found that $0.575\,\mu g\,g^{-1}$ of Dox was the lowest dose that could be used to induce the labelling of cardiac progenitors from E6.25 to E7.25 (Supplementary Fig. 2a).

To assess whether a single *Mesp1* cell could mark a common progenitor of both heart fields, we administrated this lowest dose of Dox between E6.25 and E7.25, and analysed the contribution of labelled clones to heart morphogenesis at E12.5 (Fig. 2a,b), when the segregation between the FHF and SHF derivatives is clearly established[3,19]. From the ensemble of labelled hearts, 22% (37 out of 161) were unicolour, possibly arising from a single recombination event. However, in these unicolour hearts resulting from very low Cre activity, the frequencies of different colours were not equal: YFP and RFP were over-represented as compared with the CFP and nuclear GFP (Fig. 2c), with the latter almost not expressed at all, as previously reported[20].

Unicolour hearts collected at E8.5 contained no more than 12 labelled cells, identifiable as a cluster of unicolour-labelled cells in the heart tube (Fig. 2d,e), which were not always cohesive (Fig. 2e). These data support the idea that *Mesp1*-derived progenitors minimally expand from their specification in the primitive streak to the initial stage of heart tube development and may undergo a certain degree of cellular dispersion or fragmentation. Interestingly, by E12.5, most of the single-colour hearts contained more than one cluster of labelled cells with a mean of about 3 clusters per heart (2.5 clusters $\pm$ 0.37) suggesting that clones derived from *Mesp1*-derived progenitors may become separated into more than one fragment (Fig. 2f,g), so that the total number of labelled patches represents the combined result of multiple cell induction and clonal fragmentation (Supplementary Note).

To functionally categorize with high fidelity the relative contribution of *Mesp1*-expressing cells to the FHF and SHF lineages, we defined as FHF derivatives embryos in which the left ventricle was labelled, and as SHF derivatives hearts in which the outflow tract and inflow tract were labelled[3,21]. Out of 27 unicolour hearts analysed at E12.5, no unicolour clones were found to be present in both heart fields (Fig. 2f–k). Only 2 out of 27 unicolour hearts could not be classified into FHF or SHF, as they presented clones located only in the atria or the right ventricle, which are believed to derive from both heart fields[3,19] (Fig. 2k).

As the clonal dose of Dox did not induce heart labelling when administrated at E5.75 (Supplementary Fig. 2), we administrated a dose of Dox 40 times higher to investigate whether Dox administration before E6.25 can target early multipotent *Mesp1*-expressing cells that would escape our clonal analysis. This early induction marked cells that were exclusively distributed in the FHF (Supplementary Fig. 2), ruling out the possibility that early *Mesp1*-expressing cells common for both heart fields were missed in our clonal tracing.

Dox administration at the earliest time point of cardiac progenitor specification resulted in the preferential labelling of the left ventricle (6 out of 7 hearts at E6.25 and 6 out of 7 hearts at E6.75; Fig. 2h,i,k,l), consistent with the initial emergence of *Mesp1*-derived FHF progenitors. In contrast, Dox administration at a later time point (E7.25) induced a preferential labelling of SHF derivatives (10 out of 13 hearts; Fig. 2j–l), indicating that these two pools of cardiac progenitors are specified at different time points during development.

### Bio-statistical modelling of the multicolour-labelled hearts to infer clonal fragmentation and multi-regional contribution of single *Mesp1*-expressing cells

Although this observation strongly suggests that *Mesp1* progenitors are already restricted to the FHF or SHF, to define the degree of clonal fragmentation, the regional contribution of the distinct progenitor pools, and the timing of their specification, we turned to a more rigorous statistical analysis based on the full range of clonal data including multicolour hearts (Fig. 3a,b). Although cell labelling and clonal fragmentation occur in a stochastic manner (Fig. 3c), the relative induction frequency, $pN$ (the probability of induction of an individual *Mesp1*-expressing cell times the total number of cardiac precursors), and the clonal fragmentation rate, $f$, could be inferred from the total ensemble of labelled hearts (161 labelled hearts translating to $n = 263$ independent hearts by colour) using statistical inference (Fig. 3d and Supplementary Note). By comparing the relative frequency of bicolour and tricolour hearts, we could infer the induction frequency, $pN = 1.3 \pm 0.05$, independent of the clone fragmentation rate. Then, by fitting the distribution of fragment numbers to a model based on stochastic fragmentation (Fig. 3e and Supplementary Fig. 3a,b), we found a fragmentation rate of $f = 1.6 \pm 0.2$.

With the known fragmentation rate $f$ and induction frequency $pN$, we could then assess with a defined level of confidence which of the labelled hearts of any given colour are likely to derive from a single induced cell. In particular, we found that hearts with 3 fragments or less of a given colour were likely to be monoclonal (Fig. 3f, examples in Fig. 3g,h, Supplementary Table 1 and Supplementary Note). Following this classification, we identified 89 clones in our collection of multicolour hearts that were likely to be of monoclonal origin. Remarkably, we found that all of the clones that contained fragments in the FHF or SHF were restricted to one or the other heart field, confirming that the FHF and SHF progenitors arise from distinct *Mesp1* progenitors. In contrast, of the 69 clones that had fragments in the FHF, 15% also have fragments in the other heart compartments. Similarly, of the 20 clones that have fragments in the SHF, 55% have fragments in other heart compartments (Fig. 3i and Supplementary Table 1), demonstrating that once heart progenitors have been specified, they are likely to undergo clonal fragmentation that will contribute to the morphogenesis of distinct heart regions, consistent with the regions associated with the FHF and the SHF obtained by retrospective clonal analysis[3].

By assessing the proportion of FHF and SHF precursors that are labelled at each induction time, we found that most FHF derivatives were induced from E6.25 to E6.75 whereas most SHF derivatives were labelled between E6.75 and E7.25 (Fig. 3j). Finally, by computing $pN$ and $f$ for each heart field separately, we found that $f = 1.4 \pm 0.2$ for the FHF whereas $f = 1.9 \pm 0.3$ for the SHF, showing that the latter undergoes a slightly higher rate of fragmentation (Supplementary Fig. 3c). Together, these results indicate that *Mesp1*-expressing cardiac

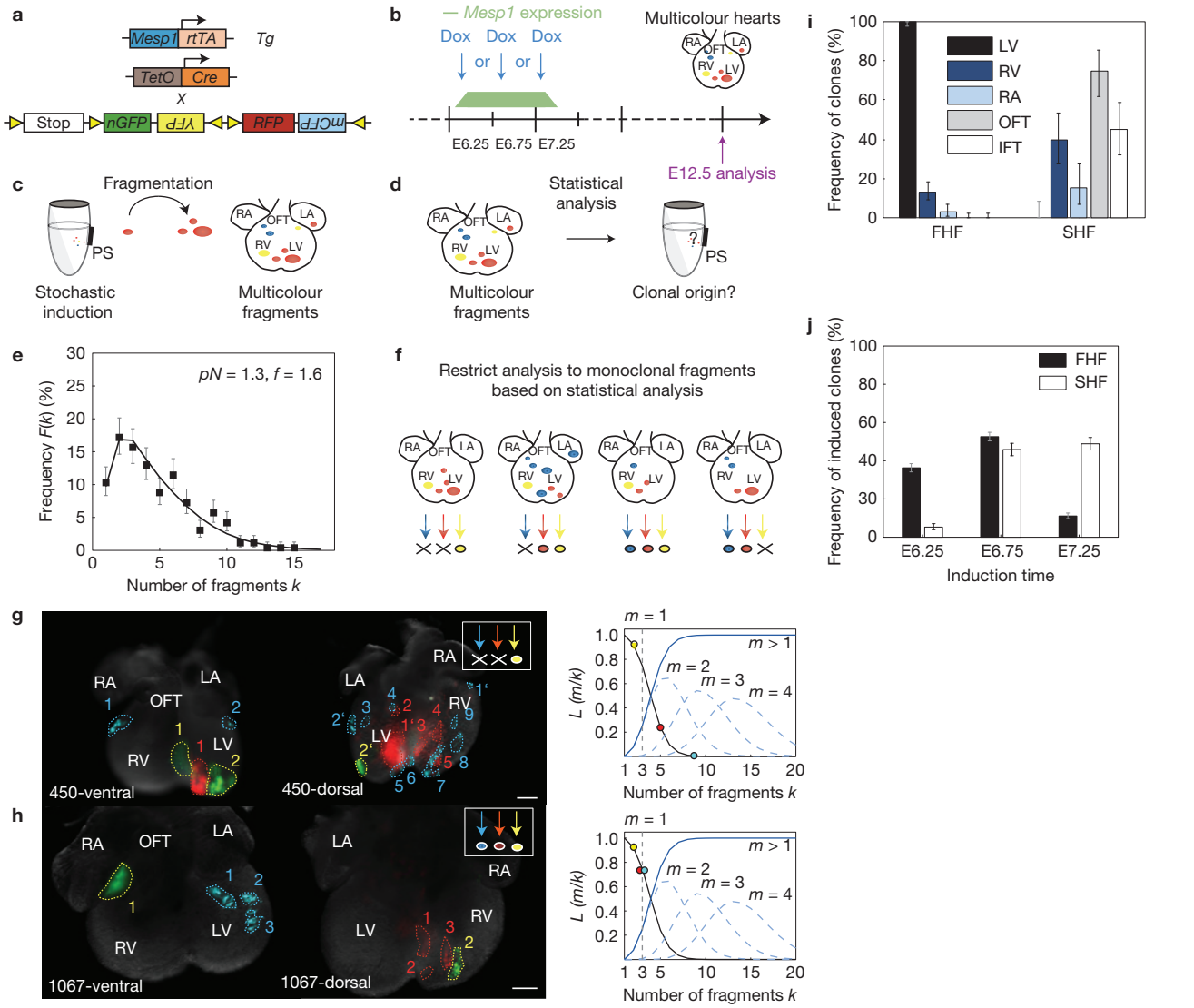



**Figure 3** Bio-statistical modelling of the the multicolour-labelled hearts. (**a**) Scheme of the genetic strategy used for the clonal tracing of *Mesp1*-expressing progenitors with different fluorescent proteins. (**b**) A low dose of Dox was injected at E6.25, E6.75 or E7.25. Multicolour induced hearts were analysed at E12.5 and classified according to their regional contribution. LV, left ventricle, RV: right ventricle, LA: left atrium, RA: right atrium, OFT: outflow tract. (**c**) On Dox administration, *Mesp1*-expressing cells are stochastically labelled in different colours. During early development, cells migrate and are rearranged such that growing clones may fragment into disconnected clusters. PS, primitive streak. (**d**) Statistical analysis of uni- and multicolour hearts was performed to infer induction frequency ($pN$) and the fragmentation rate ($f$). (**e**) The stochastic nature of the lineage labelling and fragmentation results in a broad distribution of fragment numbers (squares). With an induction frequency $pN = 1.3$, and the fragmentation rate $f = 1.6$, the statistical model (solid line) is in excellent agreement with the experimental data. $n = 263$ hearts by colour. (**f**) Statistical analysis allows restriction of the analysis to fragments that are likely to be monoclonal with a known error rate of 12% (Supplementary Fig. 5c and Supplementary Note). (**g**,**h**) Examples of E12.5 multicolour hearts induced at E6.25 (**g**) or E7.25 (**h**). Scale bars, 200 μm. In the right corner is indicated which colour is considered as clonal, based on the statistical analysis. We compare the probability $L(m=1|k)$ that $k$ fragments stem from a single clone (black line) with the probability $L(m>1|k)$ that these fragments stem from more than one cell (solid blue line). The latter is given by the sum contributions of clones with multiple cell origins (dashed blue lines). We consider $k$ fragments as monoclonal if $L(m=1|k) > L(m>1|k)$, which leaves us with a threshold value of $k=3$ (dashed grey line). The circles denote fragment numbers of the three fluorescent markers in the examples shown. (**i**) Regional contribution of FHF and SHF progenitors in monoclonal data sets ($n=89$), showing the contribution of the FHF and SHF progenitors to other cardiac regions. IFT, inflow tract. (**j**) Temporal appearance of FHF and SHF progenitors inferred from all data sets at each induction time ($n=263$ hearts by colour). The number on the bottom left of each panel refers to the ID of the labelled heart. Error bars indicate one sigma Poisson confidence intervals.

progenitors consist of two temporally distinct populations that sequentially contribute to FHF and SHF development.

*Mesp1* lineage is not exclusive to the heart but also marks other mesodermal lineages such as head muscles[22,23]. Retrospective clonal analysis has suggested a common origin for the head muscles and myocardium derived from the SHF (ref. 24). Interestingly, 11% of the embryos analysed showed co-labelling of the head muscles and the heart with the same colour (Supplementary Fig. 4a,b). The labelling of the head muscles was preferentially observed at the late induction time and was associated with the labelling of SHF derivatives including

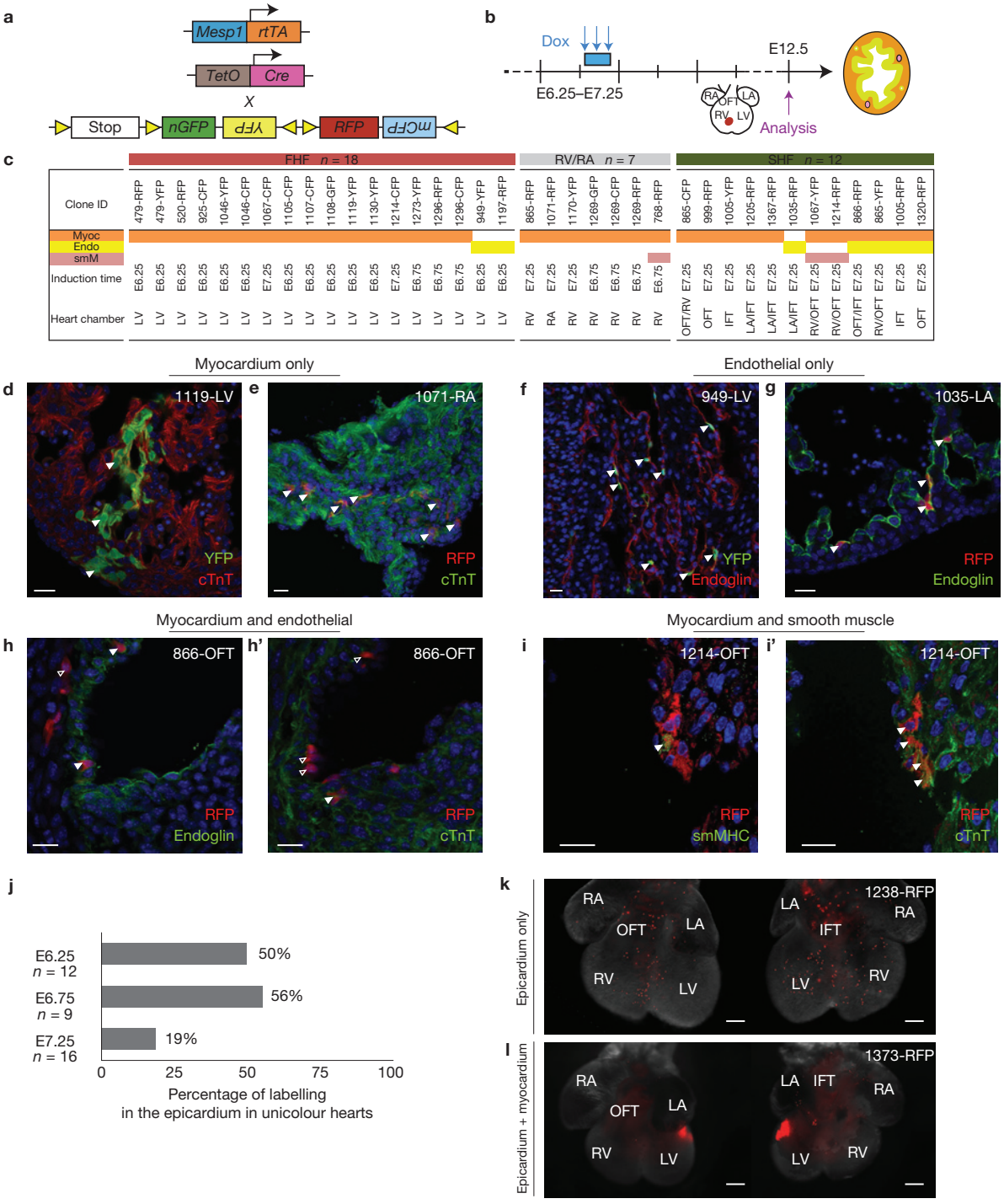



**Figure 4** Clonal analysis of lineage differentiation of *Mesp1*-derived progenitors *in vivo*. (**a**) Scheme of the genetic strategy used for the clonal tracing of *Mesp1*-expressing progenitors with different fluorescent proteins to assess their fate. (**b**) A low dose of Dox was injected into the pregnant female between E6.25 and E7.25 and induced *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* embryos were analysed at E12.5 for the expression of markers specific for the different cardiovascular lineages of the heart: CMs (cTnT), ECs (endoglin) and SMCs (smMHC). (**c**) Fate of the labelled cells in the different sectioned hearts is assessed by confocal analysis of co-immunostaining of the three markers in a given cluster. The localization of the patches within the different heart chambers and their FHF and SHF origins are indicated below: Myoc: myocardium, Endo: endothelial, smM: smooth muscle. OFT: outflow tract, IFT: inflow tract, RV: right ventricle, LV: left ventricle, RA: right atrium, LA: left atrium. (**d**–**i**) Confocal analyses of serial sections of fluorescently labelled hearts co-stained for CM and EC markers show that clones in the left ventricle differentiated into either only CM (**d**) or EC fate (**f**) and no FHF progenitors show clones positive for CM and EC markers. (**h**,**i**) In contrast, bipotent clones presenting the ability to differentiate at the clonal level into either CMs (**h**) and ECs (**h'**) or CMs (**i**) and SMCs (**i'**) can be observed in the SHF. Arrowheads point to double marked cells. Scale bars, 20 μm. (**j**) Percentage of labelling in the epicardium in unicolour hearts depending on the time of induction. (**k**,**l**) Examples of E12.5 unicolour hearts showing labelling only in the epicardial layer (**k**) or in the epicardium and myocardium (**l**). Scale bars, 200 μm. The number on the upper right in each panel refers to the ID of the labelled heart.

the RV (Supplementary Fig. 4c,d). These results indicate that common progenitors for head muscles and heart myocardium encompass the pool of *Mesp1* progenitors contributing to the SHF, consistent with previous retrospective clonal analysis[24].

### *Mesp1* progenitors consist of unipotent and bipotent progenitors

Until now, most studies assessing the differentiation potential of cardiac progenitor cells at the clonal level have been performed *in vitro*, and therefore may lack some important extrinsic cues that cardiac progenitors encounter during their *in vivo* specification. *In vitro* differentiation of single fluorescence-activated cell sorting (FACS)-isolated early cardiac progenitors (Mesp1–GFP or Brachyury–GFP/Flk1) from mouse embryos and during ESC differentiation shows that these early cardiac progenitors differentiate into CMs, ECs and SMCs, a fraction of which are multipotent at the clonal level[7,15]. Likewise, later born Nkx2-5/cKit$^+$ cardiac progenitors cells, which are preferentially enriched for FHF progenitors, differentiate into CMs, SMCs or both[13], whereas Isl1/Flk1$^+$ cells, which are preferentially enriched for SHF progenitors, give rise to colonies that differentiate into CMs, SMCs and ECs at the clonal level *in vitro*[14]. Conflicting results have been obtained concerning the fate of cardiac progenitors *in vivo* during vertebrate development[25]. Dye- and retroviral-based tracing analyses during chick heart morphogenesis suggest that CMs and ECs arise from distinct pools of progenitors[11,12], whereas lineage tracing in mouse embryos showed that these progenitors can differentiate into myocardial cells, SMCs and ECs at the population level[14,26,27], supporting the notion that during mouse development, cardiac progenitors are multipotent[25]. However, the constitutive activity of the Cre expressed in the cardiac cells precludes assessment at the clonal level as to whether the different cell types (CMs, SMCs and ECs) arise from multipotent or distinct unipotent progenitors.

To assess the fate of single *Mesp1*-expressing progenitors during cardiovascular development *in vivo*, we assessed the coexpression of fluorescent proteins with specific markers of the different cardiovascular cell types in clonally induced *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* embryos. We analysed hearts expressing fluorescently labelled patches at E12.5 and assessed the fate of the *Mesp1*$^+$ labelled cells on serial sections in a given unicolour patch (Fig. 4a–i). Surprisingly, all *Mesp1*-derived clones found in the left ventricle and in the atria were differentiated into either CMs or ECs (Fig. 4c–g). The unipotent *Mesp1*-derived CM progenitors are likely to give rise to the recently identified HCN4$^+$ unipotent FHF CM progenitors that are identified later during cardiac development[28,29]. Whereas the clones of CMs in the ventricles remain relatively cohesive, the clones of ECs composing the endocardium were not cohesive and were intermingled with many unlabelled ECs (Supplementary Fig. 5). In contrast, although some of the *Mesp1* progenitors of the SHF were also unipotent, differentiating into either CM or ECs, as previously reported during avian heart development[11,12,30], *Mesp1* progenitors of the SHF can also be bipotent, especially in the outflow or inflow tract regions (85% of the bipotent clones), differentiating into CMs and ECs (Fig. 4c,h–h'), or CMs and SMCs (Fig. 4c,i–i') at the clonal level.

Finally, we assessed the developmental origin and fate of the progenitors of the epicardium, the envelope that surrounds the heart, which give rise to the cardiac fibroblasts and SMCs of the coronary arteries[31]. The developmental origin of the epicardium in respect to the other cardiovascular progenitors remains unclear[32–34]. Our *Mesp1* clonal analysis revealed that 13 out of 37 unicolour induced hearts showed labelling in the epicardium (Fig. 4j–l), mostly arising following Dox administration at the earliest time of *Mesp1* progenitor specification (Fig. 4j). Ten of the thirteen epicardium unicolour-labelled hearts (77%) showed only contribution to the epicardium (Fig. 4k), and 3 out of 13 hearts (23%) were also associated with labelled CMs (Fig. 4l), suggesting that most epicardial cells arise from an independent population of unipotent *Mesp1* progenitors that will give rise to the epicardium lineage, and a small fraction of *Mesp1* progenitors may be bipotent, giving rise to CMs and EPDCs.

### The molecular heterogeneity of *Mesp1* progenitors reflects their regional and lineage-restricted contribution

To gain further insights into the molecular mechanisms that control *Mesp1* progenitor specification and lineage segregation during the early stage of cardiac mesoderm formation, we performed transcriptional profiling of *Mesp1*-expressing cells during the early and late stage of *Mesp1* progenitors. To this end, we administrated Dox to *Mesp1-rtTA/TetO-H2B-GFP* embryos at E6.25, or E7.25, isolated *Mesp1* H2B–GFP$^+$ and H2B–GFP$^-$ cells by FACS 6 h later, and performed microarray analysis in two independent biological experiments (Fig. 5a). At E6.5, *Mesp1* was the sixth most upregulated probe out of 46,000 probes, further demonstrating that our transgenic approach faithfully marked *Mesp1*-expressing cells. Interestingly, the comparison of these Mesp1 *in vivo* arrays to previous published arrays performed following Mesp1 overexpression or Mesp1–GFP$^+$ cells during ESC differentiation[6,7] (Fig. 5b) showed an important overlap between the genes differentially regulated in the Mesp1 GFP$^+$ cells at E6.5 and the genes regulated by Mesp1 gain of function in ESCs or associated with Mesp1–GFP at day 3 of ESC differentiation (Supplementary Table 2 and 3). Gene Ontology analysis revealed that *Mesp1* progenitors at E6.5 are statistically highly enriched in genes regulating embryonic patterning and regionalization, heart and blood vessel morphogenesis, and transcriptional regulation (Fig. 5c). These genes comprised many key transcriptional factors known to act upstream of Mesp1 (for example, *Eomes, T*; refs 35,36), downstream of Mesp1 or co-regulated with Mesp1 and regulating EMT (for example, *Snail1*) or controlling cardiovascular development (for example, *Gata4, Gata6, Hand1, Meis2*; refs 6,8,9; Fig. 5d and Supplementary Table 2). Many genes controlling key developmental signalling pathways, such as Wnt, Notch, BMP, TGF-b, FGF pathways that are regulated by Mesp1 *in vitro*[6–8], were also preferentially expressed in *Mesp1*-expressing cells *in vivo* (Supplementary Table 2). Also *Mesp1*-expressing cells preferentially expressed genes associated with cell polarity and migration (for example, *Fn*, *Cdh11*, *N-cadh, Wnt5a*, *Vangl1*, *Ninein*; Supplementary Table 2), consistent with the role of Mesp1 in regulating cardiac progenitor migration[4,37]. *Flk1* and *Pdgfra*, two genes encoding cell surface markers previously shown to mark *Mesp1*-expressing cardiovascular progenitors during mouse and human ESC and induced pluripotent stem cell differentiation[6,15], were also upregulated in Mesp1–GFP *in vivo* (Fig. 5e–i), and the same combination of cell surface markers (Flk1, Pdgfra and CXCR4) could be used to greatly enrich early *Mesp1* progenitors during embryonic development *in vivo* (Fig. 5j).

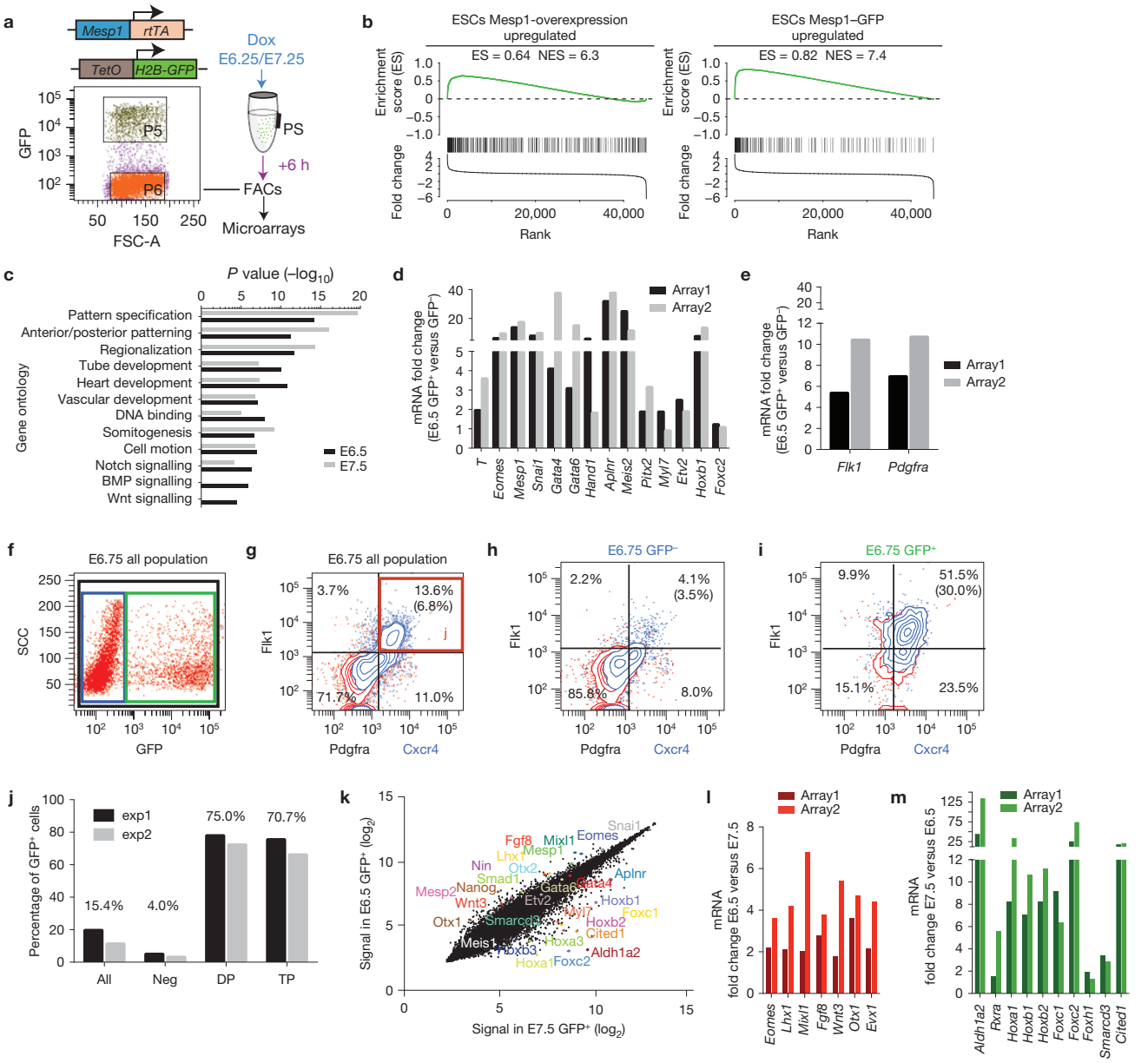



**Figure 5** Molecular signature of early and late *Mesp1*-expressing cells *in vivo*. (**a**) Genetic and cell-sorting strategy used to assess the molecular signature of early and late *Mesp1*-expressing cells *in vivo*. Induced *Mesp1–rtTA/TetO–H2B–GFP* embryos at E6.25 or E7.25 were dissected 6 h after Dox administration. GFP$^+$ and GFP$^-$ cells were isolated by FACS and microarray analyses were performed in two independent biological experiments. PS: primitive streak. (**b**) Gene set enrichment analysis of Mesp1–GFP signature at E6.5 showing the distribution of genes upregulated by Mesp1 overexpression in ESCs (ref. 6; left) or the genes upregulated in ESCs *Mesp1–GFP* (ref. 7; right). Genes are shown within the rank order list of all the microarray probe sets of E6.5 GFP$^+$ cells. The highly significant enrichment score (ES) and normalized enrichment score (NES) are shown for each analysis. (**c**) Gene ontology enrichment in *Mesp1–GFP*-expressing cells at E6.5 (black) or E7.5 (grey) ($n=2$). (**d**) Expression of early mesodermal markers, *Mesp1*, EMT markers such as *Snai1* and cardiac progenitor markers in E6.5 *Mesp1* GFP$^+$ cells as measured by microarrays. The fold change is presented over the GFP$^-$ population ($n=2$). (**e**) Surface marker expression in E6.5 *Mesp1* GFP$^+$ cells as measured by microarrays ($n=2$). (**f**) FACS analysis showing GFP expression in E6.75 *Mesp1–rtTA/TetO–H2B–GFP* embryos 6 h following Dox administration. (**g–i**) FACS analysis of the combined expression of Cxcr4 (blue), Pdgfra and Flk1 expression in all of the living cells (**g**), in GFP$^-$ (**h**) and *Mesp1* GFP$^+$ (**i**) populations shows that the GFP$^+$ population is enriched in triple-positive (TP) cells ($n=2$). The percentage of cells in each quadrant is shown and the percentage of Pdgfra$^+$/Flk1$^+$/Cxcr4$^+$ cells is shown in brackets. (**j**) FACs analysis of E6.75 *Mesp1–rtTA/TetO–H2B–GFP* embryonic cells showing that the Flk1$^+$/Pdgfra$^+$ double-positive (DP) cells and Flk1$^+$/Pdgfra$^+$/Cxcr4$^+$ (TP) triple-positive cells are highly enriched in *Mesp1*–GFP-expressing cells ($n=2$). (**k**) Comparison of *Mesp1*-expressing cells at E6.5 and E7.5. Dot plot representing the signal of each probe (merge of the two independent biological samples) showing that some key developmental genes are differentially expressed between E6.5 and E7.5. (**l**,**m**) mRNA expression at E6.5 and E7.5, as defined by microarray analysis. Genes upregulated at E6.5 (**l**) and at E7.5 (**m**) ($n=2$).

Comparison between Mesp1–GFP$^+$ cells at E6.5 and E7.5 revealed that *Mesp1* progenitors share very similar expression profiles with several Mesp1 direct target genes, such that *Gata4, Gata6* and *Aplnr* were upregulated in *Mesp1*$^+$ cells at the early and late

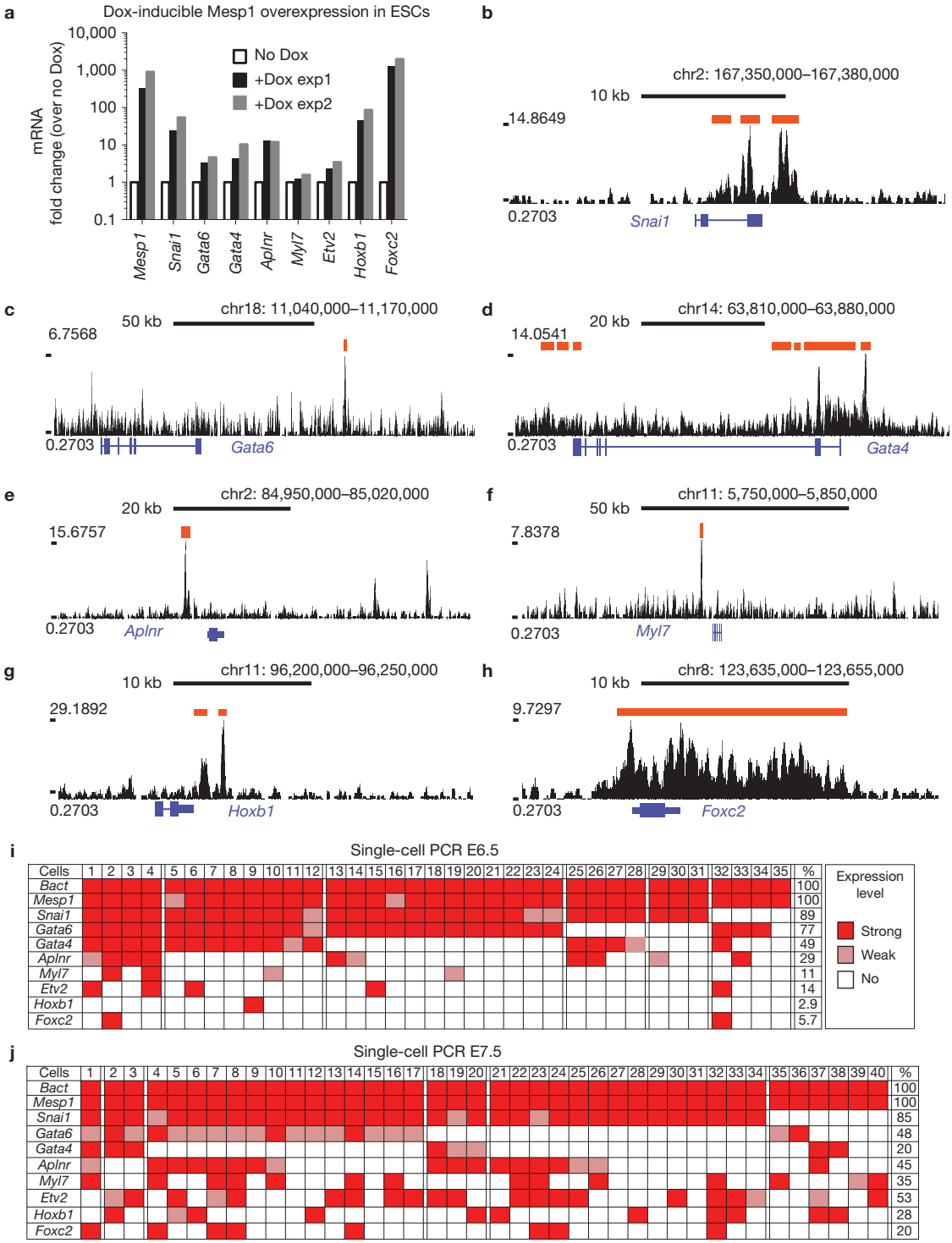



**Figure 6** Different temporal expression of Mesp1 direct target genes. (**a**) Quantitative RT-PCR analysis of Mesp1 target genes 24 h after Dox administration in Dox-inducible Mesp1-expression cells at day 2 of ESC differentiation. The fold change is presented over the unstimulated cells ($n=2$). (**b**–**h**) Mesp1 ChIP-Seq for *Snai1* (**b**), *Gata6* (**c**), *Gata4* (**d**), *Aplnr* (**e**), *Myl7* (**f**), *Hoxb1* (**g**) and *Foxc2* (**h**) showing that these genes are direct target genes of Mesp1 in ESCs. Red bars indicate significant peaks. (**i**,**j**) Single-cell RT-PCR analysis of *Snai1*, *Gata6*, *Gata4*, *Aplnr*, *Myl7*, *Hoxb1* and *Foxc2* as well as *Etv2* in *Mesp1* GFP$^+$ cells at E6.5 (**i**) and E7.25 (**j**). β-*actin* and *Mesp1* were used as internal positive controls. A dark colour indicates strong expression whereas a light colour indicates weak expression (Supplementary Fig. 6). Blank cells indicate that no PCR amplification of the genes was detected. Percentages of cells expressing the markers are indicated on the right.

time points (Fig. 5k). Despite these similarities, early and late *Mesp1*-expressing cells present also important molecular differences including the differential expression of transcription factors and Hox-related genes, previously identified in controlling pattern and

regionalization in other tissues[38–40], suggesting that these genes may regulate the patterning of the primitive streak (Supplementary Table 2). *Mixl1* (ref. 41), *Otx1* (ref. 42), *Evx1* (ref. 43) and *Lhx1* (ref. 44) were preferentially expressed in the early *Mesp1* cells (Fig. 5k,l), whereas many genes known to be associated with or controlling the morphogenesis of the SHF, such as *Aldh1a2* (ref. 45), *RXRa* (ref. 46), *Foxh1* (ref. 47), *Hoxa1*, *Hoxb1* and *Hoxb2* (ref. 48), *Smarcd3* (ref. 49), *Foxc1/Foxc2* (ref. 50) and *Cited1* (ref. 51), were more highly expressed in *Mesp1* progenitors at E7.5 (Fig. 5k,m). In addition, late *Mesp1* progenitors also preferentially express genes controlling somitogenesis (for example, *Notch1*, *Dll1*, *Lnfg* and *EphA4*; Supplementary Table 2), consistent with the well-known expression of *Mesp1* and its target genes in the first somites[52]. Together, the transcriptional profiling analyses of *Mesp1* progenitors during the early and late stages of *Mesp1* expression identify known as well as putative markers distinguishing FHF and SHF progenitors.

To further explore the molecular heterogeneity of *Mesp1* progenitors during embryonic development, we performed single-cell PCR with reverse transcription (RT-PCR) analysis to analyse the expression of several direct Mesp1 target genes, such as *Snail1*, *Gata4*, *Gata6*, *Aplnr*, *Hoxb1*, *Myl7* and *Foxc2* (Fig. 6a–h), on single FACS-isolated *Mesp1* H2B–GFP$^+$ cells at E6.5 and E7.25 (Fig. 6i,j and Supplementary Fig. 6). Interestingly, not all direct Mesp1 target genes are expressed in every *Mesp1*$^+$ cells at the same time. *Snail1* is the most commonly *Mesp1* co-expressed gene irrespective of the embryonic stages ($n = 75$), followed by *Gata6*, *Gata4* and *Aplnr* (Fig. 6i,j). Interestingly, at E6.5, less than 10% of *Mesp1* cells expressed Mesp1 target genes associated with SHF (*Hoxb1* and *Foxc2*; refs 48,53; Fig. 6i). However, at E7.5, the number of *Mesp1* cells expressing SHF markers increased by tenfold, with 20–30% of cells expressing either *Hoxb1* or *Foxc2* (Fig. 6j). The analysis of the expression of *Myl7*, a marker of CMs (ref. 54), and *Etv2*, a transcription factor associated with endothelial and endocardial cell fate[55–58], revealed that at E6.5, *Mesp1* cells usually expressed either *Myl7* or *Etv2*, whereas at later stages more *Mesp1*-expressing cells co-expressed these 2 markers (Fig. 6j), consistent with the early unipotent FHF and the late bipotent SHF progenitors found in our clonal analysis. These single-cell transcriptional profiling analyses of *Mesp1* progenitors support the existence of molecularly distinct populations of *Mesp1* progenitors, reflecting their different regional and lineage contribution.

## DISCUSSION

In contrast to the current model of cardiovascular development, in which Mesp1 is thought to mark the most primitive multipotent cardiovascular progenitors common to the FHF and SHF, our temporal clonal analysis of *Mesp1*-expressing cells provides compelling evidence that *Mesp1* marks distinct classes of cardiovascular progenitors with restricted lineage differentiation at different time points during gastrulation (Fig. 7). The absence of common FHF and SHF progenitors among *Mesp1*-expressing cells suggests that the common progenitor identified in retrospective clonal analysis[3] exists before gastrulation in the epiblast cells expressing Eomes, a transcription factor that directly controls *Mesp1* expression[35,36].

Our prospective clonal analysis of heart development reveals that, unexpectedly, most *Mesp1*-derived cardiovascular progenitors of the

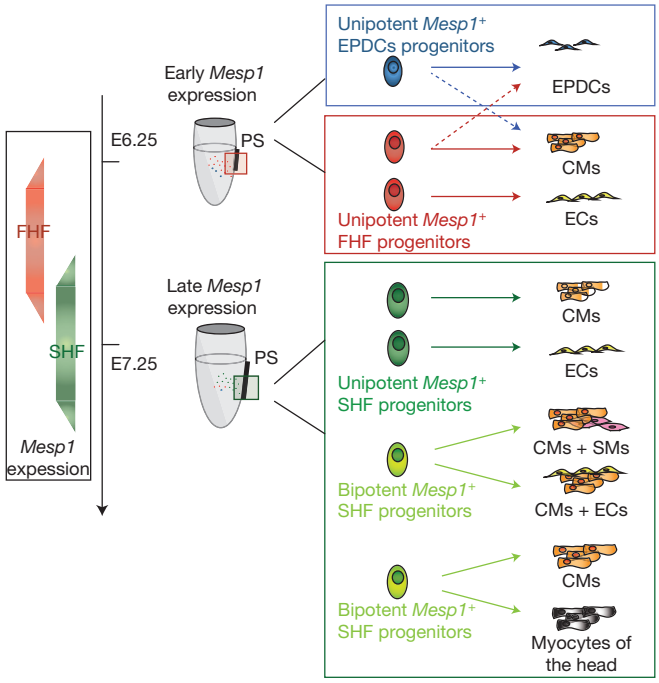



**Figure 7** Revised model of the early step of cardiovascular progenitor specification and lineage commitment during mouse development. Clonal and molecular analysis of *Mesp1* progenitors shows the existence of temporally distinct *Mesp1* progenitors that contribute to the heart development. *Mesp1* progenitors first give rise to the FHF (in red) and then to the SHF (in green) progenitors with an overlapping expression of *Mesp1* in the two populations at E6.75. FHF progenitors are unipotent and give rise to either CMs or ECs. SHF progenitors are either unipotent or bipotent. Epicardial and EPDCs arise as an independent *Mesp1*-derived lineage at the early time points. PS: primitive streak.

FHF are restricted to either CM or EC cell fates at the time of their specification. In contrast, *Mesp1*-derived SHF progenitors can be unipotent or bipotent (Fig. 7). The main difference between the multilineage differentiation potential of cardiovascular progenitors *in vitro*[6,13–15,59] and their more restricted fate *in vivo* suggests that the ultimate fate of the progenitors can be regulated by the environmental cues that the different progenitors encounter during cardiac morphogenesis.

Our molecular analysis of *Mesp1* progenitors provides the first transcriptional profiling of the early cardiac progenitors *in vivo* and uncovered that the two populations of *Mesp1* progenitors, although very similar molecularly, present also notable differences, consistent with their lineage and regional contribution. This analysis identified several key markers, such as *Mixl1*, *Otx1* and *Evx1*, that are preferentially expressed in the early *Mesp1* cells, whereas *Aldh1a2*, *RXRa*, *Foxh1*, *Foxc1/Foxc2*, *Hoxa1*, *Hoxb1*, and *Hoxb2*, *Smarcd3*, all genes known to be expressed during or controlling SHF morphogenesis[45–49,53], are preferentially expressed in the late *Mesp1* progenitors. Further studies will be required to define which of these differentially regulated genes temporally and spatially control the emergence of the distinct populations of *Mesp1* progenitors during gastrulation. Whereas previous studies proposed that Mesp1 acts as a master regulator of cardiovascular development[6,8,9], our single-cell RT-PCR analysis demonstrates that Mesp1 only induces

the expression of a combination of different direct target genes in different cells. Understanding how this specificity is achieved will be important to instruct and/or restrict the fate of multipotent cardiovascular progenitors into a particular cell lineage *in vivo*. The answers to these questions will be important both to design strategies to direct the differentiation of cardiovascular progenitors derived from ESCs and induced pluripotent stem cells specifically into pure populations of CMs, and for improving cellular therapy in cardiac diseases. □

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note: Supplementary Information is available in the online version of the paper*


### ACKNOWLEDGEMENTS

We thank F. Bollet-Quivogne and J-M. Vanderwinden for their help with confocal imaging and M. Tarabichi for his help with gene set enrichment analysis. We thank M. Buckingham and A. Joyner for kindly providing the probes for *in situ* hybridization. We also thank the Genomic core facility of the EMBL, Heidelberg for their help with the ChIP-Seq. F.L. has been sequentially supported by the FNRS and the EMBO long-term fellowship. S.C. is supported by a fellowship of the FRS/FRIA. X.L. is supported by the FNRS. A.B. is supported by the FRS/FNRS. S.R. and B.D.S. are supported by the Wellcome Trust (grant number 098357/Z/12/Z). C.B. is an investigator of WELBIO. This work was supported by the FNRS, the ULB foundation, the Fondation contre le Cancer, the European Research Council (ERC), and the foundation Bettencourt Schueller (C.B. and F.L.).


### AUTHOR CONTRIBUTIONS

C.B., F.L., S.C. and X.L. designed the experiments and performed data analysis. F.L. and S.C. performed most of the experiments. X.L. performed the single-cell PCR analysis. Y.A. generated the *Mesp1–rtTA* transgenic mice. A.R. and H. A. performed microarrays. C.P. provided technical assistance. C.D. helped with FACS isolation of *Mesp1*-expressing cells. A.B. helped in the design and initial characterization of the *Mesp1–rtTA* transgene. B.D.S. and S.R. performed the bio-statistical analysis of the clonal fate data. C.B. and F.L. wrote the manuscript.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.


Published online at www.nature.com/doifinder/10.1038/ncb3024
Reprints and permissions information is available online at www.nature.com/reprints



1. Garry, D. J. & Olson, E. N. A common progenitor at the heart of development. *Cell* **127,** 1101–1104 (2006).
2. Kelly, R. G., Brown, N. A. & Buckingham, M. E. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Dev. Cell* **1,** 435–440 (2001).
3. Meilhac, S. M., Esner, M., Kelly, R. G., Nicolas, J. F. & Buckingham, M. E. The clonal origin of myocardial cells in different regions of the embryonic mouse heart. *Dev. Cell* **6,** 685–698 (2004).
4. Saga, Y. *et al.* MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development* **126,** 3437–3447 (1999).
5. Bondue, A. & Blanpain, C. Mesp1: a key regulator of cardiovascular lineage commitment. *Circ. Res.* **107,** 1414–1427 (2010).
6. Bondue, A. *et al.* Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. *Cell Stem Cell* **3,** 69–84 (2008).
7. Bondue, A. *et al.* Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation. *J. Cell Biol.* **192,** 751–765 (2011).
8. Lindsley, R. C. *et al.* Mesp1 coordinately regulates cardiovascular fate restriction and epithelial-mesenchymal transition in differentiating ESCs. *Cell Stem Cell* **3,** 55–68 (2008).
9. David, R. *et al.* MesP1 drives vertebrate cardiovascular differentiation through Dkk-1-mediated blockade of Wnt-signalling. *Nat. Cell Biol.* **10,** 338–345 (2008).
10. Buckingham, M. E. & Meilhac, S. M. Tracing cells for tracking cell lineage and clonal behavior. *Dev. Cell* **21,** 394–409 (2011).
11. Cohen-Gould, L. & Mikawa, T. The fate diversity of mesodermal cells within the heart field during chicken early embryogenesis. *Dev. Biol.* **177,** 265–273 (1996).
12. Wei, Y. & Mikawa, T. Fate diversity of primitive streak cells during heart field formation in ovo. *Dev. Dyn.* **219,** 505–513 (2000).
13. Wu, S. M. *et al.* Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell* **127,** 1137–1150 (2006).
14. Moretti, A. *et al.* Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127,** 1151–1165 (2006).
15. Kattman, S. J., Huber, T. L. & Keller, G. M. Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev. Cell* **11,** 723–732 (2006).
16. Haraguchi, S. *et al.* Transcriptional regulation of Mesp1 and Mesp2 genes: differential usage of enhancers during development. *Mech. Dev.* **108,** 59–69 (2001).
17. Saga, Y. *et al.* MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development* **122,** 2769–2778 (1996).
18. Kitajima, S., Miyagawa-Tomita, S., Inoue, T., Kanno, J. & Saga, Y. Mesp1-nonexpressing cells contribute to the ventricular cardiac conduction system. *Dev. Dyn.* **235,** 395–402 (2006).
19. Buckingham, M., Meilhac, S. & Zaffran, S. Building the mammalian heart from two sources of myocardial cells. *Nat. Rev. Genet.* **6,** 826–835 (2005).
20. Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143,** 134–144 (2010).
21. Lescroart, F., Mohun, T., Meilhac, S. M., Bennett, M. & Buckingham, M. Lineage tree for the venous pole of the heart: clonal analysis clarifies controversial genealogy based on genetic tracing. *Circ. Res.* **111,** 1313–1322 (2012).
22. Harel, I. *et al.* Distinct origins and genetic programs of head muscle satellite cells. *Dev. Cell* **16,** 822–832 (2009).
23. Sambasivan, R. *et al.* Distinct regulatory cascades govern extraocular and pharyngeal arch muscle progenitor cell fates. *Dev. Cell* **16,** 810–821 (2009).
24. Lescroart, F. *et al.* Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. *Development* **137,** 3269–3279 (2010).
25. Harris, I. S. & Black, B. L. Development of the endocardium. *Pediatr. Cardiol.* **31,** 391–399 (2010).
26. Cai, C. L. *et al.* Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* **5,** 877–889 (2003).
27. Stanley, E. G. *et al.* Efficient Cre-mediated deletion in cardiac progenitor cells conferred by a 3'UTR-ires-Cre allele of the homeobox gene Nkx2-5. *Int. J. Dev. Biol.* **46,** 431–439 (2002).
28. Liang, X. *et al.* HCN4 dynamically marks the first heart field and conduction system precursors. *Circ. Res.* **113,** 399–407 (2013).
29. Spater, D. *et al.* A HCN4+ cardiomyogenic progenitor derived from the first heart field and human pluripotent stem cells. *Nat. Cell Biol.* 1098–1106 (2013).
30. Milgrom-Hoffman, M. *et al.* The heart endocardium is derived from vascular endothelial progenitors. *Development* **138,** 4777–4787 (2011).
31. Riley, P. R. & Smart, N. Vascularizing the heart. *Cardiovasc. Res.* **91,** 260–268 (2011).
32. Ma, Q., Zhou, B. & Pu, W. T. Reassessment of Isl1 and Nkx2-5 cardiac fate maps using a Gata4-based reporter of Cre activity. *Dev. Biol.* **323,** 98–104 (2008).
33. Zhou, B. *et al.* Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart. *Nature* **454,** 109–113 (2008).
34. Zhou, B., von Gise, A., Ma, Q., Rivera-Feliciano, J. & Pu, W. T. Nkx2-5- and Isl1-expressing cardiac progenitors contribute to proepicardium. *Biochem. Biophys. Res. Commun.* **375,** 450–453 (2008).
35. Van den Ameele, J. *et al.* Eomesodermin induces Mesp1 expression and cardiac differentiation from embryonic stem cells in the absence of Activin. *EMBO Rep.* **13,** 355–362 (2012).
36. Costello, I. *et al.* The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nat. Cell Biol.* **13,** 1084–1091 (2011).
37. Christiaen, L. *et al.* The transcription/migration interface in heart precursors of *Ciona intestinalis*. *Science* **320,** 1349–1352 (2008).
38. Lossie, A. C., Nakamura, H., Thomas, S. E. & Justice, M. J. Mutation of I7Rn3 shows that Odz4 is required for mouse gastrulation. *Genetics* **169,** 285–299 (2005).
39. Nakamura, H., Cook, R. N. & Justice, M. J. Mouse Tenm4 is required for mesoderm induction. *BMC Dev. Biol.* **13,** 9 (2013).
40. Cruz, C. *et al.* Induction and patterning of trunk and tail neural ectoderm by the homeobox gene eve1 in zebrafish embryos. *Proc. Natl Acad. Sci. USA* **107,** 3564–3569 (2010).
41. Hart, A. H. *et al.* Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. *Development* **129,** 3597–3608 (2002).
42. Acampora, D. *et al.* OTX1 compensates for OTX2 requirement in regionalisation of anterior neuroectoderm. *Gene. Expr. Patterns* **3,** 497–501 (2003).
43. Dush, M. K. & Martin, G. R. Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. *Dev. Biol.* **151,** 273–287 (1992).
44. Tsang, T. E. *et al.* Lim1 activity is required for intermediate mesoderm differentiation in the mouse embryo. *Dev. Biol.* **223,** 77–90 (2000).
45. Niederreither, K. *et al.* Embryonic retinoic acid synthesis is essential for heart morphogenesis in the mouse. *Development* **128,** 1019–1031 (2001).

46. Sucov, H. M. *et al.* RXR alpha mutant mice establish a genetic basis for vitamin A signaling in heart morphogenesis. *Genes Dev.* **8,** 1007–1018 (1994).
47. Von Both, I. *et al.* Foxh1 is essential for development of the anterior heart field. *Dev. Cell* **7,** 331–345 (2004).
48. Bertrand, N. *et al.* Hox genes define distinct progenitor sub-domains within the second heart field. *Dev. Biol.* **353,** 266–274 (2011).
49. Lickert, H. *et al.* Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432,** 107–112 (2004).
50. Kume, T., Jiang, H., Topczewska, J. M. & Hogan, B. L. The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev.* **15,** 2470–2482 (2001).
51. Dunwoodie, S. L., Rodriguez, T. A. & Beddington, R. S. Msg1 and Mrg1, founding members of a gene family, show distinct patterns of gene expression during mouse embryogenesis. *Mech. Dev.* **72,** 27–40 (1998).
52. Nomura-Kitabayashi, A. *et al.* Hypomorphic Mesp allele distinguishes establishment of rostrocaudal polarity and segment border formation in somitogenesis. *Development* **129,** 2473–2481 (2002).
53. Seo, S. & Kume, T. Forkhead transcription factors, Foxc1 and Foxc2, are required for the morphogenesis of the cardiac outflow tract. *Dev. Biol.* **296,** 421–436 (2006).
54. Kubalak, S. W., Miller-Hance, W. C., O'Brien, T. X., Dyson, E. & Chien, K. R. Chamber specification of atrial myosin light chain-2 expression precedes septation during murine cardiogenesis. *J. Biol. Chem.* **269,** 16961–16970 (1994).
55. Ferdous, A. *et al.* Nkx2-5 transactivates the Ets-related protein 71 gene and specifies an endothelial/endocardial fate in the developing embryo. *Proc. Natl Acad. Sci. USA* **106,** 814–819 (2009).
56. Kataoka, H. *et al.* Etv2/ER71 induces vascular mesoderm from Flk1+PDGFR$\alpha$+ primitive mesoderm. *Blood* **118,** 6975–6986 (2011).
57. Rasmussen, T. L. *et al.* ER71 directs mesodermal fate decisions during embryogenesis. *Development* **138,** 4801–4812 (2011).
58. Palencia-Desai, S. *et al.* Vascular endothelial and endocardial progenitors differentiate as cardiomyocytes in the absence of Etsrp/Etv2 function. *Development* **138,** 4721–4732 (2011).
59. Misfeldt, A. M. *et al.* Endocardial cells are a distinct endothelial lineage derived from Flk1+ multipotent cardiovascular progenitors. *Dev. Biol.* **333,** 78–89 (2009).

## METHODS

**Mice.** *Mesp1–Cre*[4] mice were obtained from M. Buckingham, Pasteur Institute, France. *Rosa–Confetti* mice were kindly provided by H. Clevers, Hubrecht Institute, The Netherlands[20]. *TetO–Cre* mice[60] were provided by A. Nagy, Cincinnati Children's Hospital Medical Center, USA. *Rosa–tdTomato*[61] mice were obtained from Jackson laboratory. *TetO–H2B–GFP*[62] mice were provided by E. Fuchs, The Rockefeller University, USA. Mice colonies were maintained in a certified animal facility in accordance with European guidelines. These experiments were approved by the local ethical committee under the reference #LA1230332(CEBEA). The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

**Generation of Mesp1–rtTA mice.** The coding sequence of the *rtTA* (Clonetech) was subcloned under the control of a 5.6 kb fragment of the *Mesp1* promoter previously shown to be active in cardiac progenitors *in vivo*[16] and during ESC differentiation *in vitro*[7]. The *Mesp1–rtTA* fragment was linearized and microinjected into fertilized oocytes by Y. Achouri from the transgenic core facility of the Université Catholique de Louvain (UCL). Transgenic founders were identified by PCR and their functional characterizations were confirmed by lineage tracing experiments using *Mesp1–rtTA/TetO–Cre/Rosa–tdTomato* following doxycycline (Dox) administration (25 μg g$^{-1}$ of Dox injected at embryonic day (E)6.25 by intravenous injection followed by Dox administration in the drinking water (2 mg ml$^{-1}$) until E7.5).

The induction was also assessed in *Mesp1–rtTA/TetO–H2B–GFP* embryos after intravenous injection of 25 μg g$^{-1}$ of Dox at E6.25. After 5 h, GFP was detected and embryos were imaged with a macroscope (Axiozoom V16, with an Axiocam MRN camera, Carl Zeiss) and the Zen Blue software (Carl Zeiss).

**Assessment of the recombination of the Rosa–tdTomato reporter by PCR.** Genomic DNA was extracted by DNA precipitation with ethanol from the tail of an adult wild-type mouse, from an adult *Rosa–tdTomato* (control for no recombination) mouse or from the heart of an E12.5 *Mesp1–Cre/Rosa–tdTomato* (control for recombination) mouse. For E6.5 and E7.5 *Mesp1–Cre/Rosa–tdTomato* and *Mesp1–rtTA/TetO–Cre/Rosa–tdTomato* embryos, the complete litter ($n = 8$ embryos) was lysed in 50 μl of lysis buffer (1× PCR-buffer (Qiagen), proteinase K (0.04 mg ml$^{-1}$)). Proteinase K was inactivated at 95 °C for 30 min. PCR amplification on genomic DNA was performed by nested PCR with primers flanking the LoxP sites. The first amplification was performed with 50 pg of control DNA or 5 μl of the lysed embryos (E6.5 and E7.5) under the following conditions: amplification with the Go-Taq polymerase (Promega) with F1: 5′-ACGTGCTGGTTATTGTGCTGTCTC-3′ and R2: 5′-CTCCTCGCCCTTGCTCACCAT-3′ primers, 94 °C for 25 s, 60 °C for 25 s, and 72 °C for 10 s for 35 cycles. One microlitre of the PCR product was engaged for a second amplification with a Taq-polymerase from Qiagen and F4: 5′-CCGCGGGCCCTAAGAAGTTCC-3′ and R4: 5′-ACCATGGTGGCGGGATCGTG-3′ primers under the following conditions: 94 °C for 25 s, 60 °C for 25 s, and 72 °C for 10 s for 35 cycles.

***In situ* hybridization.** Embryos were extracted at E6.75 and fixed overnight in 4% paraformaldehyde and processed according to published protocols[63] with some modifications: the proteinase K treatment was performed at the concentration of 5 μg ml$^{-1}$ and the hybridization was realized overnight at 68 °C in 5× SSC (pH 5), 50% formamide, 500 μg ml$^{-1}$ yeast tRNA, 100 μg ml$^{-1}$ heparin, 0.5% CHAPS and 0.2% Tween20. The hybridization signal was revealed by using NBT/BCIP (Roche) for Mesp1 antisense riboprobes or BM Purple (Roche) for Cre antisense riboprobes. Chromogenic substrate and embryos were acquired in 75% glycerol in PBST (0.1% Tween20) with a Leica MZ16F stereomicroscope (Leica Microsystems). Acquisition data were treated with LAS V4.2 software (Leica Microsystems, Belgium) and exported in TIF image format. *In situ* hybridization for *Cre* and *Mesp1* was performed on at least 4 different litters for each genotype or condition.

Antisense riboprobes for *Mesp1* (ref. 4) and *Cre* (ref. 64) were synthesized from vectors respectively kindly provided by M. Buckingham (Institut Pasteur, France) and A. Joyner (Sloan Kettering Institute, USA).

**Clonal analysis of *Mesp1*-expressing cells.** *Mesp1–rtTA/TetO–Cre* mice were crossed with the *Rosa–Confetti* reporter mice. The day of plug identification corresponds to embryonic day E0.5. Doxycycline (Dox; Sigma; 0.575 μg g$^{-1}$ or 25 μg g$^{-1}$) was administered by intravenous injection at E5.75, E6.25, E6.75, E7.25 or E8.5. Embryos were collected at E8.5 or E12.5 and fixed in 4% paraformaldehyde for 30 min and 1 h and 30 min respectively. Fluorescent protein expression was analysed with a macroscope (Axiozoom V16, with an Axiocam MRN camera, Carl Zeiss) and the Zen Blue software (Carl Zeiss). Acquisitions were done with a ×1 PLAN APO Z ×1.0/0.25 objective. For each fluorescent protein, a *z*-stack was acquired and the algorithm 'extended depth of focus' was used to produce two-dimensional images and the data were then merged and exported in TIF image format.

**Immunofluorescence analysis.** Immunofluorescence was performed on frozen heart sections (Leica Cryostat—20 μm). Sections were stained in PBS with 1% BSA, 0.5% Triton and 5% horse serum. Sections were stained with the following primary antibodies: anti-cTnT (MS-295-P, mouse, clone 13-11; 1:100; Neomarkers, Fremont), anti-endoglin (AF1320, goat, 1:500, R&D), anti-smMHC (BT562, rabbit, 1:100, Biomedical Technologies), anti-Wt1 (Sc-192, rabbit, 1:100, Santa Cruz), anti-HCN4 (Ab32675, rat, 1/100, Abcam). Counterstaining of nuclei was performed with Hoechst (1/40,000). Acquisitions were acquired with a confocal microscope (LSM780; Carl Zeiss) with a ×20 Plan Neofluar objective (×20; 0.8 numerical aperture). Collection of sequential 0.22 μm–0.60 μm thicknesses, 1,024 × 1,024-pixel optical sections were acquired for each fluorescent protein. The acquisition data were then treated with Zen black software (Carl Zeiss) and exported in TIF image format. All experiments were reproduced in at least 3 biological samples.

**Microarray analysis.** For transcriptome analysis, *Mesp1–rtTA/TetO–H2B–GFP* embryos induced with 25 μg g$^{-1}$ of Dox by intravenous injection at E6.25 or E7.25 were extracted 6 h after Dox administration and dissected in a dissection medium (DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS)). Embryos were selected for their expression of GFP with a macroscope (Axiozoom V16, with an Axiocam MRN camera, Carl Zeiss). Cells were dissociated for 3 min at 37 °C in trypsin/EDTA (tryspin 0.1%, EDTA 1 mM) and resuspended into PBS supplemented with 10% FBS and 1:1,000 of propidium iodide.

FACS analysis was performed using a FACSAria I at high pressure (70 p.s.i.) and FACSDiva software (BD Biosciences). Living (gated with propidium iodide dye exclusion) cells were sorted on the basis of the expression of GFP. We sorted 50 GFP$^+$ or GFP$^-$ cells at E6.5 or E7.5 and cells were collected directly in 45 μl of lysis buffer (20 mM dithiothreitol, 10 mM Tris-HCl pH 7.4, 0.5% SDS, 0.5 μg μl$^{-1}$ proteinase K). Samples were then lysed at 65 °C for 15 min and frozen. RNA isolation, amplification and microarray were performed by A.R. and H.A. in the Functional Genomics Core, Barcelona. RNA was isolated using magnetic beads. cDNA synthesis, library preparation and amplification were performed as described previously[65]. Amplification was performed for 25 cycles. Subsequently, cDNA was purified on PCR GenElute Clean Up Kit (Sigma Aldrich) and eluted in 30 μl water. cDNA concentration was determined using the Nanodrop 1000 spectrophotometer. Microarrays were then performed on a Mouse Genome 430 PM strip Affymetrix array. The data were normalized using the RMA algorithm. The entire procedure was repeated in two biologically independent samples. Genetic signatures were obtained by considering genes presenting a fold change greater or smaller than 1.5 or −1.5, respectively.

The microarray data have been deposited in the Gene Expression Omnibus of NCBI and are accessible through GEO accession number GSE59033 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59033).

**Flow cytometry.** For flow cytometry analysis, *Mesp1–rtTA/TetO–H2B–GFP* embryos were induced with 25 μg g$^{-1}$ of Dox by intravenous injection from E6.25 followed by Dox administration in the drinking water (2 mg ml$^{-1}$) until E7.5. Embryos were extracted at E6.75, E7.25 or E7.75 dissected in a dissection medium (DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS)). Embryos were selected for their expression of GFP with a macroscope (Axiozoom V16, with an Axiocam MRN camera, Carl Zeiss) ($n = 4/5$ embryos per litter). Cells were dissociated for 3 min at 37 °C in trypsin/EDTA (tryspin 0.1%, EDTA 1 mM) and resuspended in PBS supplemented with 1% bovine serum albumin. Staining for Flk1, Pdgfra and Cxcr4 was performed as previously described[6]. Flk1 (VEGFR2) was stained for 30 min at room temperature using a biotinylated antibody at 1:100 (13-5821, clone Avas12a1; eBioscience) revealed by a streptavidin–phosphatidylethanolamine (PE)–Cy7 secondary antibody at 1:400 (557598, BD). PDGFRa was stained using a PE-coupled rat monoclonal antibody at 1:75 (12-1401, clone APA5; eBioscience). CXCR4 was stained using an APC-coupled rat monoclonal antibody at 1:100 (17-9991, clone 2B11; eBioscience). Living cells were gated by Hoechst dye exclusion (1:3,000). FACS analyses were performed on a FACSFortessa device (BD) in duplicate biological samples.

**qPCR on Mesp1-overexpressing ESCs.** ESC culture, RNA extraction and qPCR analysis were performed as previously described[6]. Briefly, Dox-inducible Mesp1–Flag ESCs were cultured on irradiated mouse embryonic fibroblasts in ESC Medium (DMEM (Gibco) supplemented with 15% ESC-qualified FBS (Invitrogen), 0.1 mM nonessential amino acids (Invitrogen), 1 mM sodium pyruvate (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma-Aldrich), 100 U ml$^{-1}$ penicillin (Invitrogen), 100 μg ml$^{-1}$ streptomycin (Invitrogen), and 1,000 U ml$^{-1}$ leukaemia inhibitory factor (ESGRO)). ESC differentiation was performed in hanging drops of 1,000 cells in 25 μl ESC medium without LIF, supplemented with 50 μg ml$^{-1}$ ascorbic acid (Sigma). Doxycycline (Dox, Sigma) was added to hanging drops at day 2 to a final concentration of 1 μg ml$^{-1}$. After 24 h post-induction of Dox, the embryoid bodies with or without Dox were collected for RT-qPCR. Total RNA and DNase

DOI: 10.1038/ncb3024

treatments were performed according to the manufacturer's instruction (Absolutely RNA MiniPrep Kit, Stratagene). One microgram of total RNA was used for reverse transcription by using Superscript II (Invitrogen) with Oligo dT primer, and qPCR was performed by using Power SYBR Green Master Mix, (Invitrogen) on a real-time PCR system (Mx3005P; Agilent Technologies). qPCR primers were described in Supplementary Table 4.

**Mesp1 ChIP-Seq analysis on Mesp1-overexpression ESCs.** Chromatin for chromatin immmunoprecipitation (ChIP) was obtained from Mesp1-overexpressing differentiating ESCs (ref. 6). Embryoid bodies ($10 \times 10^6$ cells) were collected 36 h after the induction with Dox at day 2 of differentiation and fixed with 1% formaldehyde for 7 min. Formaldehyde was quenched with 0.125 M glycine for 5 min at room temperature. Embryoid bodies were lysed according to the manufacturer's instructions (ChIP-IT Express Kit, Active Motif) and crosslinked DNAs were sonicated during 2 cycles of 5 min (30″ON/30″OFF) with a Bioruptor Sonicator (Diagenode). Sheared DNAs have a size range between 100 and 300 bp. ChIP was performed by using anti-Flag-M5 antibody (F4042, Sigma) or the isotype control (M5284, Sigma) according to the manufacturer's instructions (ChIP-IT express kit, Active Motif). Total yields of DNA obtained after ChIP were purified by the Ipure kit (Diagenode). Libraries were prepared, generated and sequenced at the Genomic core facility (Heidelberg) using a 5X Illumina HiSeq sequencer.

ChIP-Seq reads, from Mesp1 or isotype control ChIP, were aligned to the UCSC mm9 version of the mouse genome using Bowtie[66]. Note that when a read could be aligned at more than one position, only one position (the top scoring) was kept.

Peak identification was performed by using MACS software[67] with standard parameters. The cutoff *P* value used is $1 \times 10^{-5}$.

**Single-cell RT-PCR analysis.** For single-cell analysis, *Mesp1–rtTA/TetO–H2B–GFP* embryos induced with 25 μg g$^{-1}$ of Dox by intravenous injection at E6.25 or E7.0 were extracted 6 h after Dox administration and dissected in a dissection medium (DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS)). Cells were dissociated for 3 min at 37 °C in trypsin/EDTA (tryspin 0.1%, EDTA 1 mM) and resuspended in PBS supplemented with 10% FBS and 1:1,000 of propidium iodide.

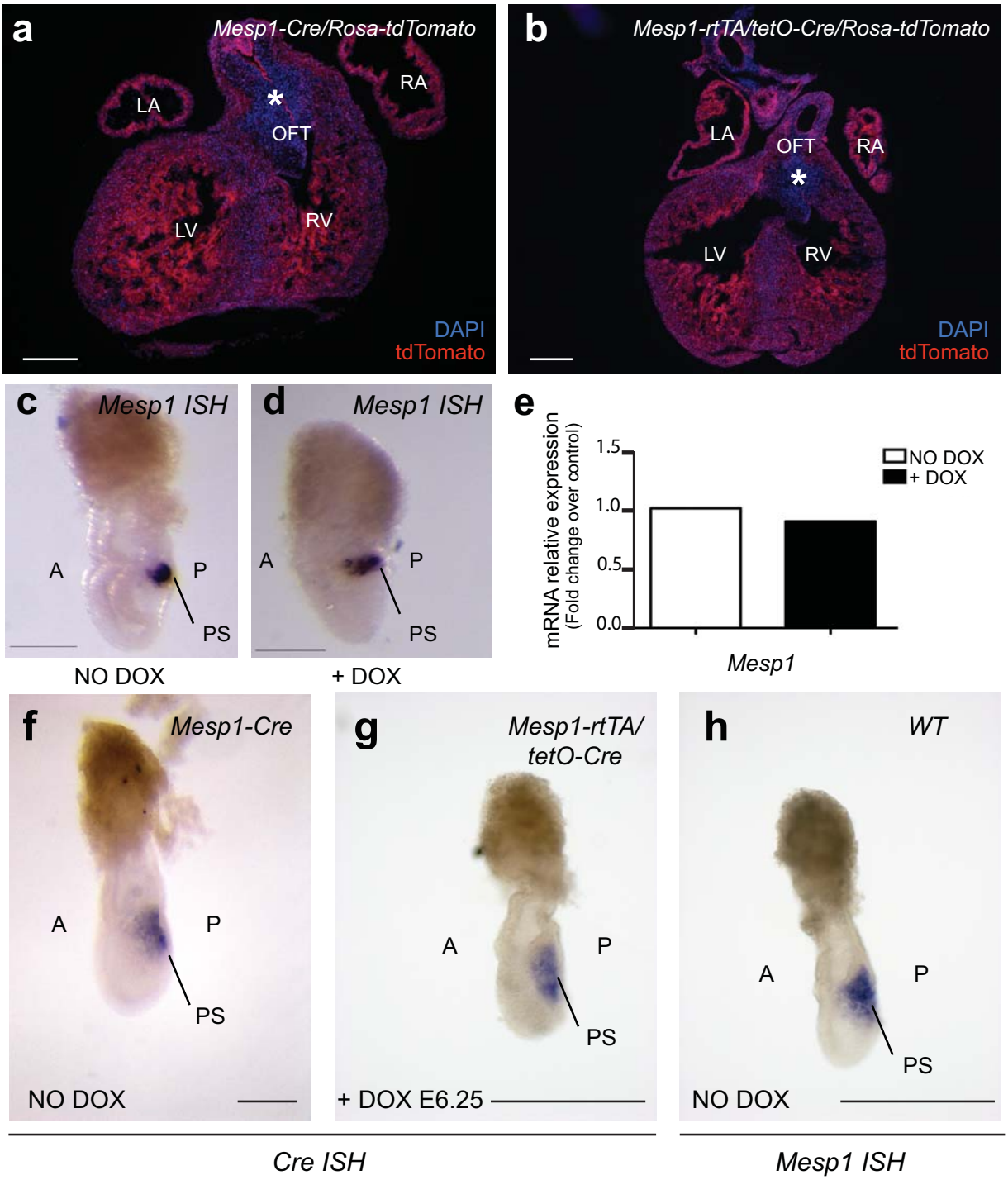
FACS analysis was performed using a FACSAria I at high pressure (70 p.s.i.) and FACSDiva software (BD Biosciences). Living (gated with propidium iodide dye exclusion) cells were sorted in 96-well plates with the FACSDiva software (BD Biosciences) on the basis of the expression of GFP. We have single GFP$^+$ cells at E6.5 or E7.5 and cells were collected directly in 4.5 μl of single-cell first-strand buffer (Superscript III buffer, Invitrogen), 0.5% Nonidet P40 (Pierce), 10 mM dNTP mixture (Invitrogen), 42 pmol l$^{-1}$ of the RT primer[68], 1 mM dithiothreitol (Invitrogen), 10 mM dNTP mixture (Qiagen), SuperRNaseIN (Ambion), and RNAout (Invitrogen). After sorting, single cells were snap-frozen in liquid nitrogen and subsequently lysed at 65 °C for 5 min.

For single-cell PCR, generation of cDNA, and PCR amplification were performed as previously described[69,70] with incorporation of suppression PCR (ref. 68). In brief, after lysis RT primers (Supplementary Table 4) were allowed to anneal at 42 °C before addition of 0.5 μl Superscript III reverse transcriptase and incubation at 45 °C for one hour, and the reaction was inactivated at 70 °C for 15 min. Unannealled RT primer was digested by exonuclease I (NEB) with 6.7 mM $MgCl_2$ at 37 °C for 30 min and inactivated at 80 °C for 25 min. Removal of the RNA template and polyadenylation were performed at the same time by adding RNAseH (Invitrogen), 1.5 mM dATP (Invitrogen) and 30 units of terminal deoxynucleotidyltransferase (TDT, Promega). For secondary strand synthesis, 4 μl of polyadenylated cDNA was mixed with 10 μl of the secondary strand synthesis buffer (7 μl 2× Terra PCR Direct Buffer, 70 pmol l$^{-1}$ tagging primer (Supplementary Table 4), 2.5 μl Milli-Q water, 0.5 μl Terra PCR Direct Polymerase Mix), and performed one round at 98 °C for 130 s, 42 °C for 1 min and 68 °C for 5 min. The reaction tube was immediately placed on ice, and 6 μl PCR buffer (3 μl Terra PCR Direct Buffer, 1.9 μmol l$^{-1}$ suppression PCR primer (Supplementary Table 4), 3 μl ddw) was added, and the first round of PCR enrichment was performed under the following conditions for a total of 25 PCR cycles: 98 °C for 10 s, 65 °C for 30 s, and 68 °C for 4 min. One microlitre of the first-round PCR products was used in the second-round PCR enrichment by using the suppression PCR primer (Supplementary Table 4) for 40 cycles. The second PCR enrichment was performed under the following conditions: 95 °C for 20 s, 65 °C for 30 s and 68 °C for 4 min 0.5 μl of secondary amplification product was used as the template for the PCR by using gene-specific primers (Supplementary Table 4). Primers were designed mainly using Primer Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) or QuantPrimer (http://www.quantprime.de) and are listed in Supplementary Table 4.

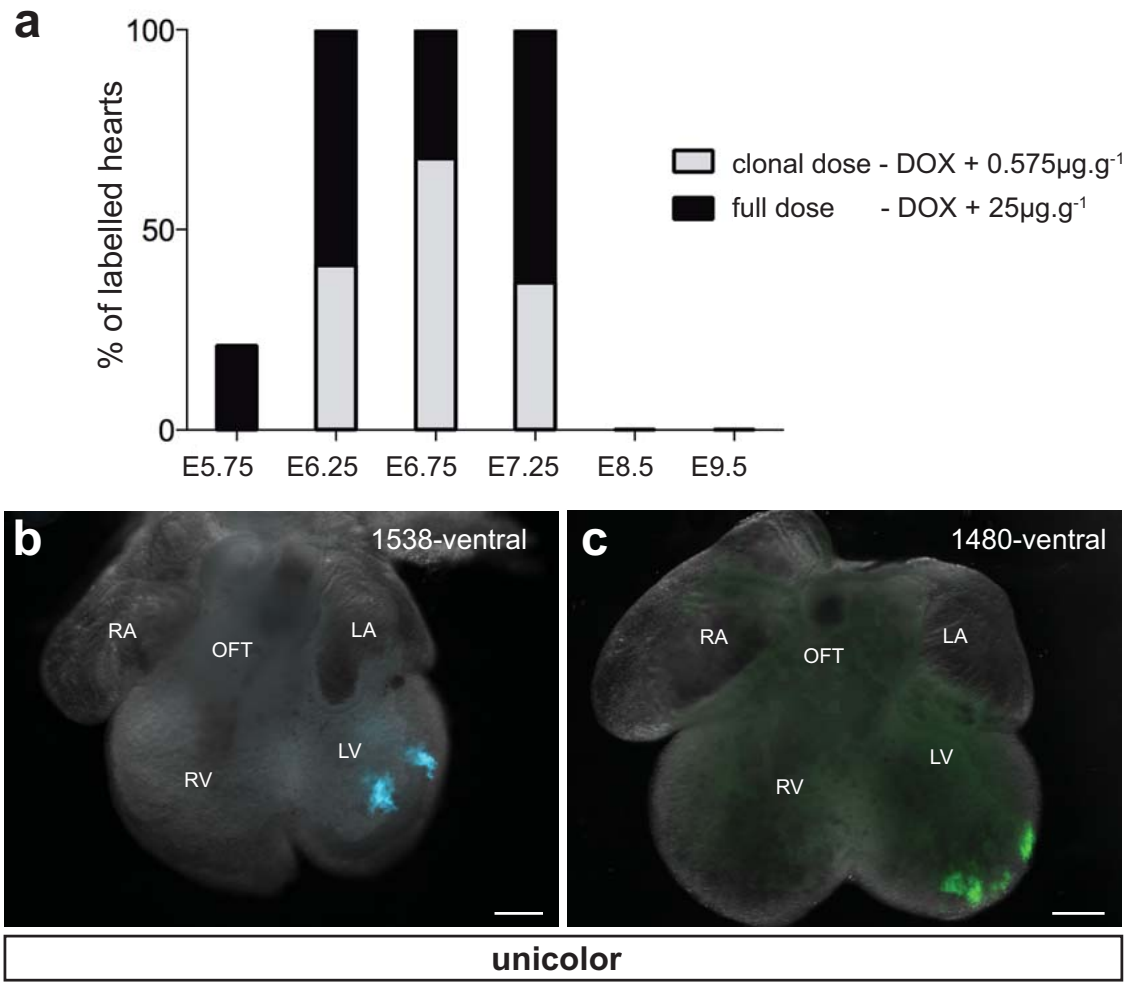

60. Perl, A. K., Wert, S. E., Nagy, A., Lobe, C. G. & Whitsett, J. A. Early restriction of peripheral and proximal cell lineages during formation of the lung. *Proc. Natl Acad. Sci. USA* **99,** 10482–10487 (2002).
61. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13,** 133–140 (2010).
62. Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303,** 359–363 (2004).
63. Piette, D., Hendrickx, M., Willems, E., Kemp, C. R. & Leyns, L. An optimized procedure for whole-mount *in situ* hybridization on mouse embryos and embryoid bodies. *Nat. Protoc.* **3,** 1194–1201 (2008).
64. Zinyk, D. L., Mercer, E. H., Harris, E., Anderson, D. J. & Joyner, A. L. Fate mapping of the mouse midbrain-hindbrain constriction using a site-specific recombination system. *Curr. Biol.* **8,** 665–668 (1998).
65. Gonzalez-Roca, E. *et al.* Accurate expression profiling of very small cell populations. *PLoS ONE* **5,** e14418 (2010).
66. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
67. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
68. Sasagawa, Y. *et al.* Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **14,** R31 (2013).
69. Jensen, K. B. & Watt, F. M. Single-cell expression profiling of human epidermal stem and transit-amplifying cells: Lrig1 is a regulator of stem cell quiescence. *Proc. Natl Acad. Sci. USA* **103,** 11958–11963 (2006).
70. Tan, D. W. *et al.* Single-cell gene expression profiling reveals functional heterogeneity of undifferentiated human epidermal cells. *Development* **140,** 1433–1444 (2013).

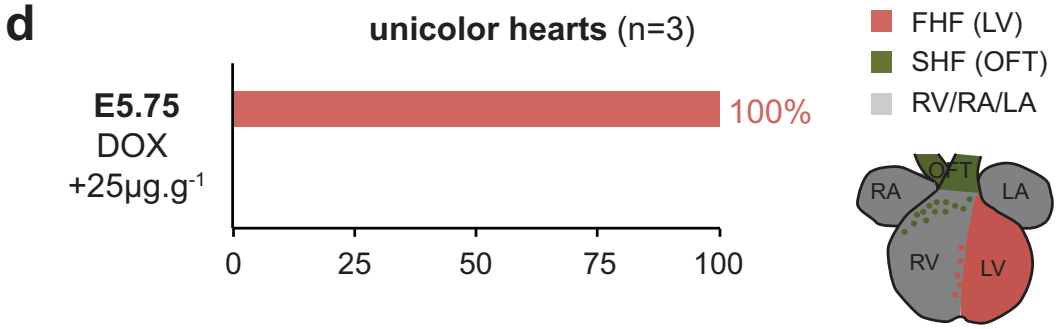
DOI: 10.1038/ncb3024

nature cell biology

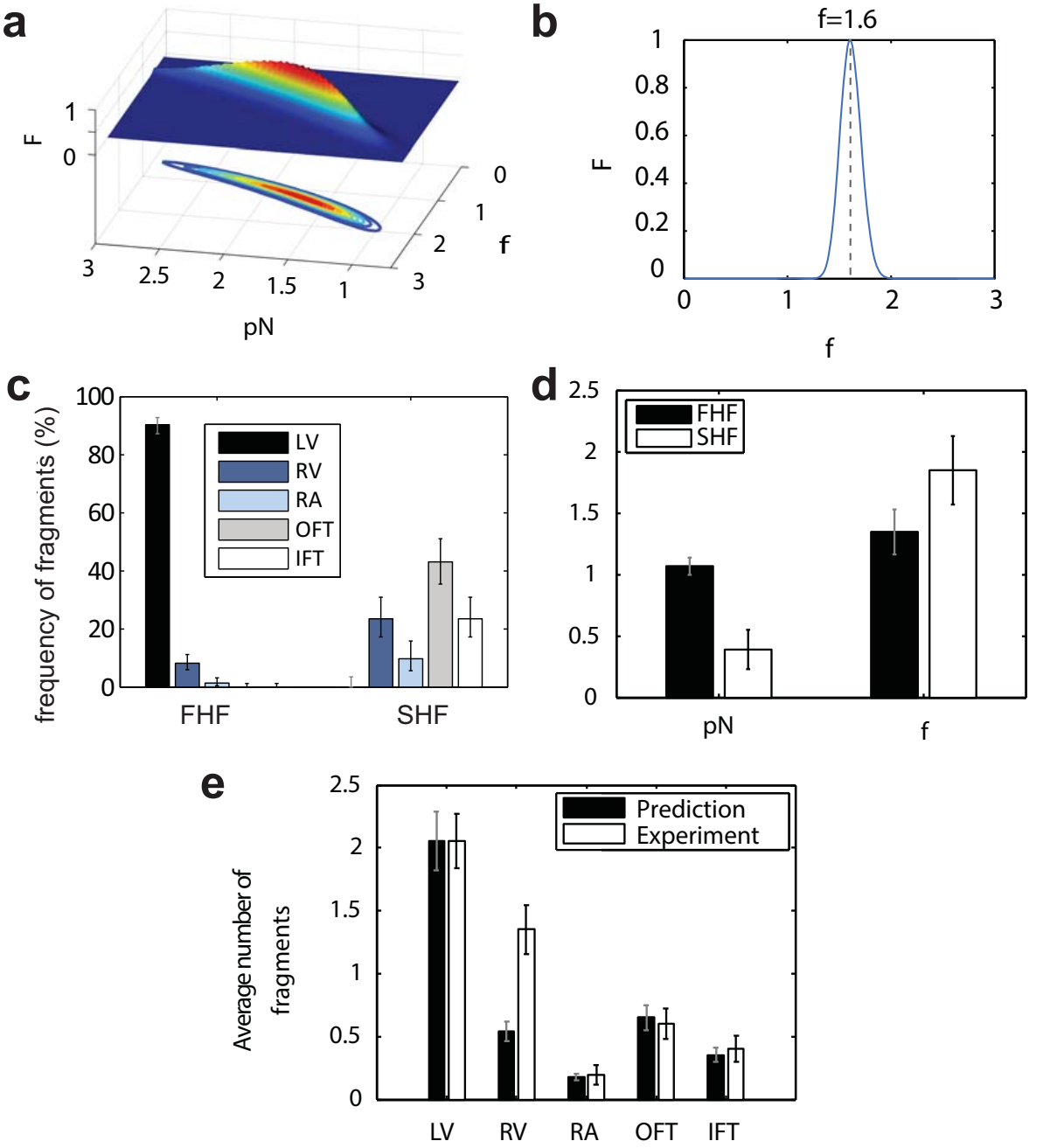



**Supplementary Figure 1** *Mesp1-rtTA/tetO-Cre* transgenic mice induced Cre expression similarly to *Mesp1-Cre* Knock-in mice. **a-b.** Sections of E12.5 *Mesp1-Cre/Rosa-tdTomato* (a) and *Mesp1-rtTA/TetO-Cre/Rosa-tdTomato* hearts (induced by Dox administration between E6.25 and E7.5) (b) and co-stained with DAPI. Both transgenic hearts have a similar expression of the tdTomato with a negative region in the OFT that derive from Mesp1 negative neural crest cells (asterisks). **c-e.** Doxycycline injection has no effect on *Mesp1* expression during early mouse embryonic development. **c-d.** In situ hybridization for *Mesp1* expression in early embryo at E6.5. The detection of *Mesp1* mRNA in the primitive streak (PS) and the nascent lateral mesoderm is similar in embryo that did not receive DOX (c) and in embryos injected with DOX (+ DOX) (d). A, anterior; P, posterior. **e.** Expression of *Mesp1* analyzed by RT-qPCR in early embryos (E6.75) without (n= 9) or after doxycycline injections (n= 6). These data show no difference in *Mesp1* expression after DOX injection. **f-g.** In situ hybridization for *Cre* expression in early embryos at E6.75. The detection of *Cre* mRNA in the primitive streak (PS) and the nascent lateral mesoderm is similar in *Mesp1-Cre* knock-in (f) and in *Mesp1-rtTA/TetO-Cre* transgenic embryos injected with DOX at E6.25 (+ DOX) (g). *Cre* expression is similar to the endogenous *Mesp1* expression in wild type embryos (h). A, anterior; P, posterior. Scale bar: 500μm.

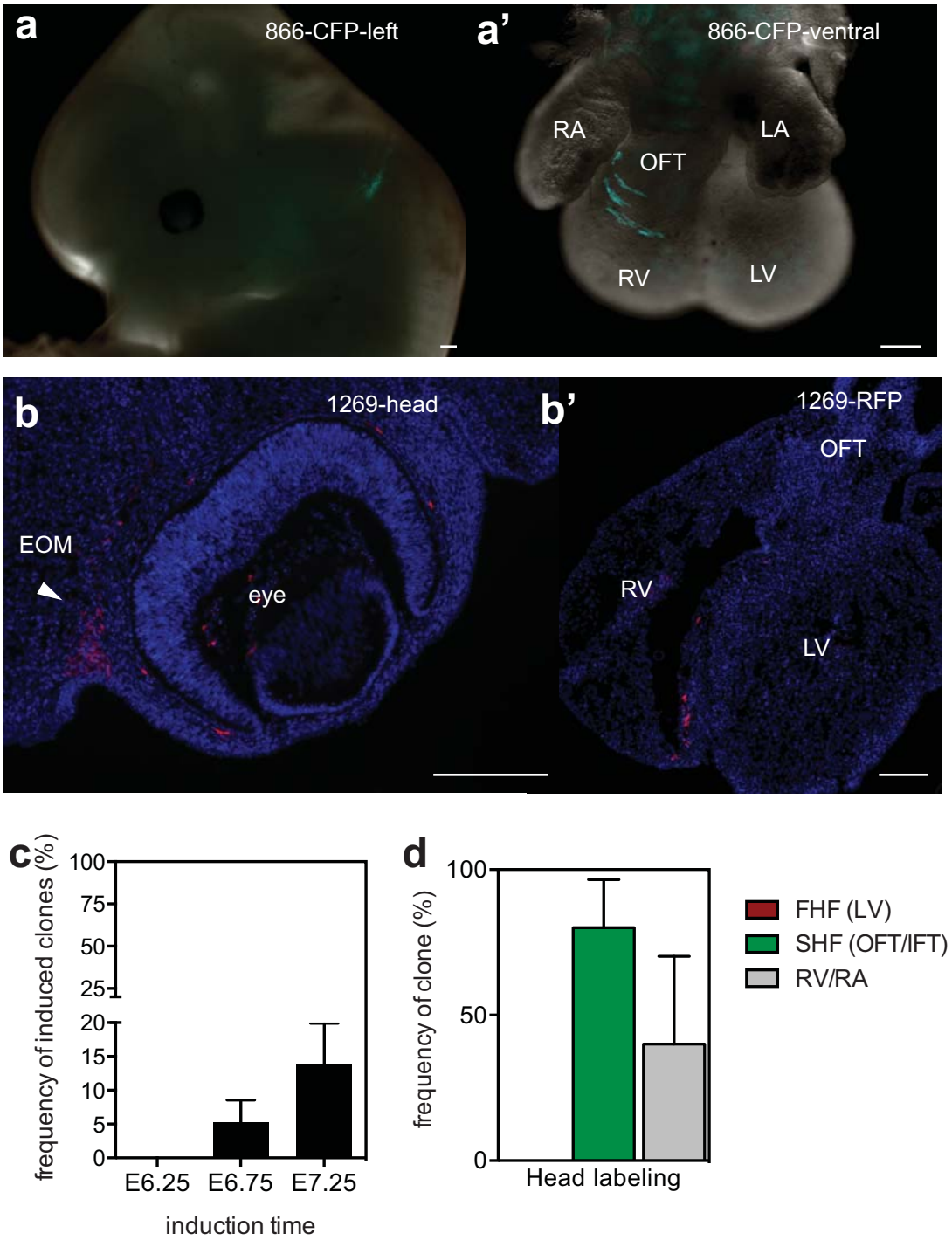


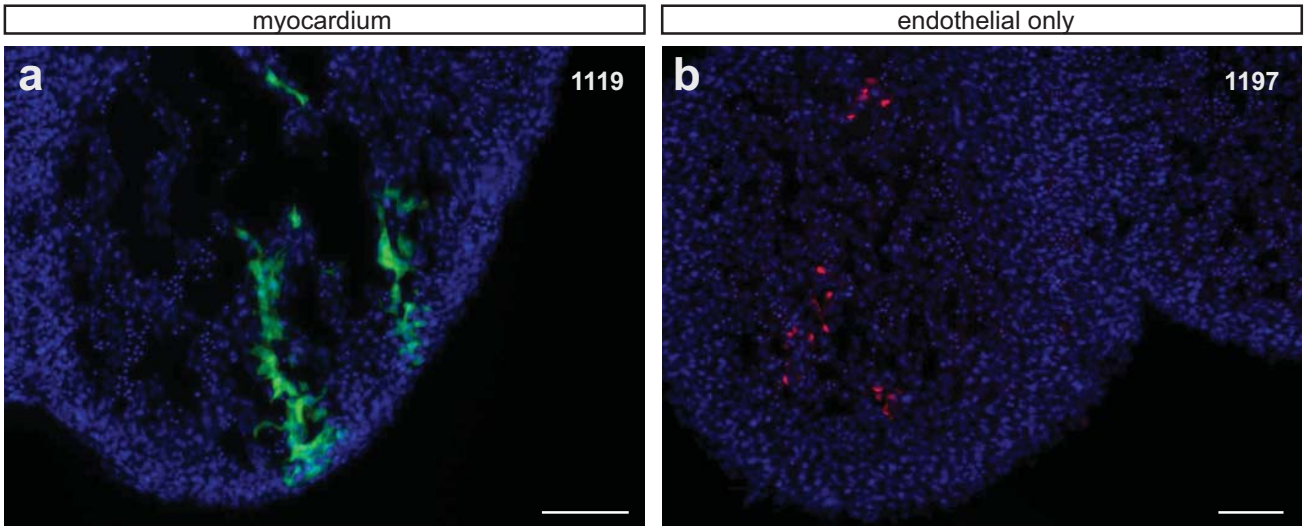

**Supplementary Figure 2** Temporal Dox administration in Mes1-rtTA/TetO-Cre/Rosa-Confetti embryos. **a.** While clonal dose of DOX (0.575µg.g-1) induces labelling in *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* embryos at E6.25 (n=53), at E6.75 (n=118) or at E7.25 (n=65), this dose was not sufficient to induce labelling at E5.75 (n=13). A much higher dose of Dox (25µg.g-1) was required to produce labelling at a clonal density at E5.75 (n=90). This 40 fold increase of DOX is likely to persist at a concentration sufficient to activate the Cre at the time of endogenous *Mesp1* expression. This high dose of DOX never labelled any heart after administration at E8.5 or E9.5 (n=24) supporting the absence of transgene expression after the end of endogenous *Mesp1* expression. **b,c.** Examples of *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* unicolour labelled hearts at E12.5 induced at E5.75 after administration of high dose of Doxycycline (25µg.g-1). Note that each cluster is localized within the LV, FHF derivative and no labelling was detected other compartments. OFT, outflow tract; RV, right ventricle; LV, left ventricle; RA, right atrium; LA, left atrium; IFT, inflow tract. Scale bars: 200 µm. The number on the upper right in each panel refers to the ID of the labelled heart. **d.** Quantification of the regional (FHF and SHF) contribution of patches of Mesp1 labelled cells in unicolour hearts induced at E5.75 with the high dose of Doxycycline (25µg.g-1), shows the exclusive labelling of the FHF (red) similarly to was found at E6.25 (Fig. 2m).



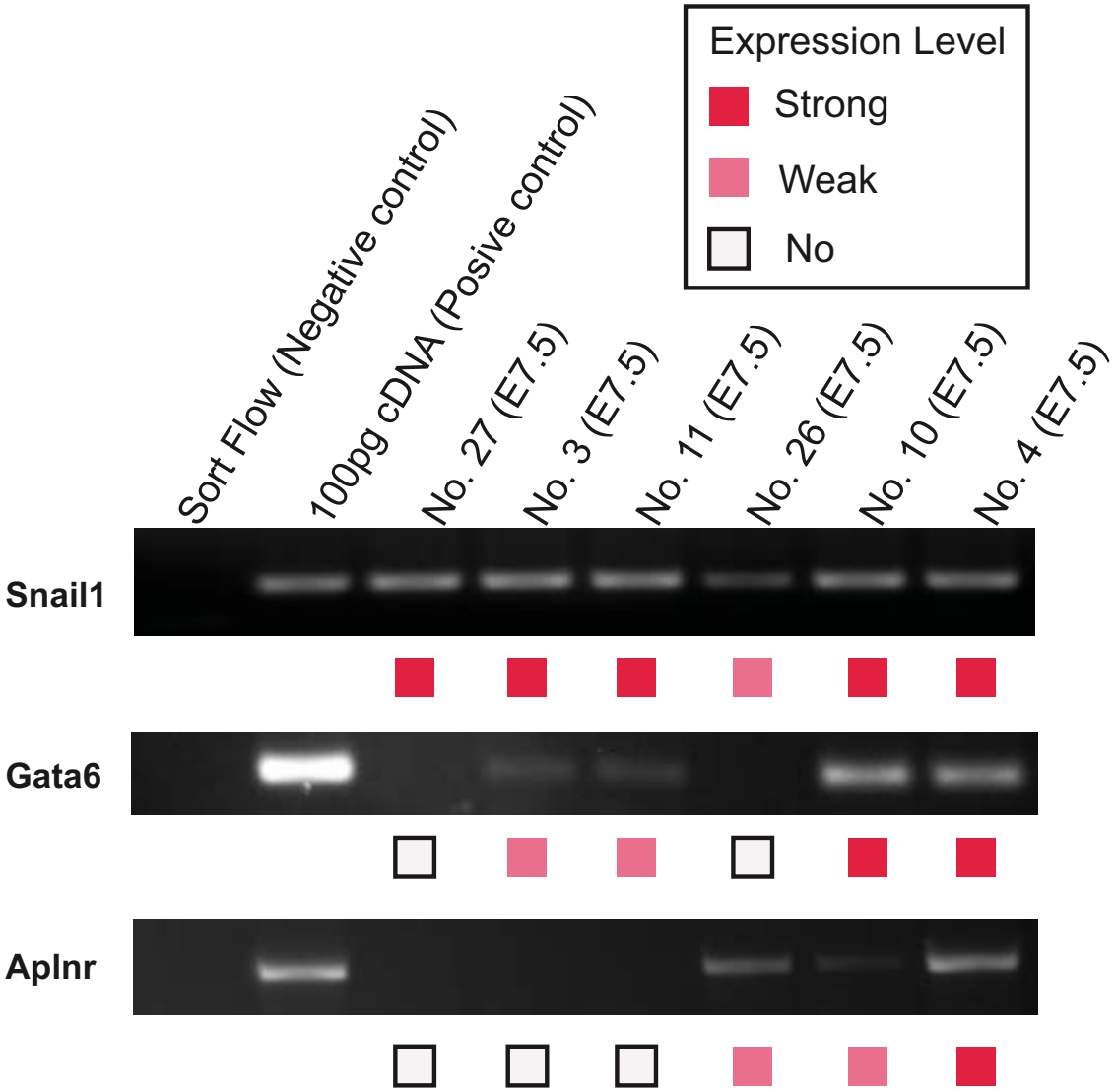

**Supplementary Figure 3** Biostatistical modeling of the clonal fate data. **a.** The likelihood function F gives the probability of the experimental data for different values of the induction frequency pN and the fragmentation rate f. The numeric values have been rescaled such that the maximum of the likelihood function corresponds to 1. Color denotes the value of F, such that red signifies a large value and blue signifies small value. Lines of equal values are indicated on the bottom of the figure. One sees that the maximum value of F is relatively featureless along a curve in the pN-f-plane. To infer the values of pN and f we must therefore refer to an independent measurement of one of the two parameters. **b.** The multicolour labelling strategy allows us to independently infer the induction frequency pN=1.3 by evaluating the abundances of hearts with a given number of colours. With this, we are left with a slice through the pN-f-plane and the fragmentation rate can be determined with a higher accuracy. **c.** Monoclonal datasets (n=89) identify two subpopulations in *Mesp1* expressing cells: FHF progenitors, which contribute to the LV and SHF progenitors, which contribute to OFT and IFT. The plot shows the probabilities of monoclonal fragments in the different heart compartments. **d.** Values for the induction frequency, pN, and the fragmentation rate, f, for the two FHF (n=188) and SHF (n=102) precursors. While the overall induction frequency is higher for FHF precursors, which we attribute to highest expression of *Mesp1* at the early time points, the fragmentation rate is higher for SHF precursors. **e.** We may use the distribution of monoclonal fragments (c) to predict the distribution of fragments in all hearts (n=263). We find an excellent agreement with the notable exception of the RV, which might suggest the existence of an independent pool of progenitors contributing to RV morphogenesis. Error bars indicate one sigma (c and e) or 95% (d) confidence intervals.




**Supplementary Figure 4** Late *Mesp1* progenitors also contribute to the head. **a-a'.** Example of a *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* embryo with co-labelled head (a) and heart (a') at E12.5. Scale bars: 200 µm. **b-b'.** Sections of a *Mesp1-rtTA/TetO-Cre/Rosa-Confetti* labelled embryo showing labelling in the head in an extraocular muscle (EOM) (b) as well as in the right ventricle RV (b'). Scale bars: 200 µm. **c.** Temporal appearance of head muscle labelling inferred from all datasets (n=105 independent embryos translating to n=181 embryos by colour). Plotted is the fraction of head muscle labelling induced at each induction time for a given colour. Head muscles are preferentially labelled at the late time points. **d.** Regional contribution of head progenitors in monoclonal datasets (n=5), showing the co-labelling of the head with the heart and preferentially SHF derivatives or the right ventricle (RV). OFT, outflow tract; RV, right ventricle; LV, left ventricle; RA, right atrium; LA, left atrium. Errors bars indicate one sigma confidence intervals.




**Supplementary Figure 5** Cohesive versus dispersive mode of growth of the myocardium and the endocardium. **a-b.** Sections of E12.5 *Mesp1-rtTA/tetO-Cre/Rosa-Confetti* unicolour hearts. **a.** Example of a compact myocardial YFP-labelled clone showing cohesive growth of the myocardium. **b.** Example of a dispersed endocardial RFP-unicolour clone showing dispersive mode of growth of the endocardium. Scale bars: 200 µm. The number on the upper right in each panel refers to the ID of the labelled heart.




**Supplementary Figure 6** Semi-quantification of single cell RT-PCR analysis. Examples showing strong, weak and no gene expression in single cells.

nature cell biology

DOI: 10.1038/ncb3024

**Supplementary Table legends**

**Supplementary Table 1** Table summarizing the clonal fate data according to their regional contribution and their probability of being monoclonal. Description of the labelling in *Mesp1-rtTA/tetO-Cre/Rosa-Confetti* induced heart. For each region, the number of labelled clusters is indicated. OFT, outflow tract, RV, right ventricle, LV, left ventricle, RA, right atrium, LA, left atrium, IFT, inflow tract.

**Supplementary Table 2** Up-regulated genes in *Mesp1* GFP+ cells *in vivo.* Description of genes displaying a change in expression of >2 fold between *Mesp1*-GFP+ and *Mesp1*-GFP- cells at E6.5 and 7.5. (Fold change over GFP- cells at E6.5 ; Fold change over GFP- cells at E7.5) in 2 independent biological replicates. A gene ontology analysis was used to classify the up-regulated genes in the following categories : Transcription Factors/Chromatin Remodelling, Signaling pathways, Migration/Polarity/Guidance and Others (all biological function related to early embryo development that we can not put in any previous classes). **In bold** (overexpressed in Mesp1 GOF ESC) Underlined (overexpressed in Mesp1-GFP ESC).

**Supplementary Table 3** Gene up-regulated in both *in vivo* and *in vitro* arrays.

**Supplementary Table S4** primer sequences.

# Supplementary Note

The interpretation of clonal fate data in the developing heart is complicated by the potential of clones to "fragment" into disconnected "clusters" as the tissue expands and cells rearrange, making the assignment of clonal progeny potentially ambiguous. However, by implementing a mathematical framework to analyze the statistics of the resulting clonal fragments, it is possible to faithfully recover information on the lineage potential and timing of the marked Mesp1 expressing cells. In the following supplementary note section, we detail the analytical program, presenting only the summary of the method in the main text.

We begin by introducing a simple stochastic framework to model clone fragmentation. Making use of the multicolor labeling strategy, we then use the observed clonal fate data to infer the induction frequency and the fragmentation probability of Mesp1 expressing cells. This enables us to assign clonal fragments to single-cell induction events. As a result, we show that Mesp1 expressing cells consist of two discrete subpopulations, one committed to the first heart field (FHF) derivatives and the other committed to the second heart field (SHF) derivatives. Further, with this assignment, we then show that these subpopulations are temporally distinct: while FHF precursors are mostly induced during the earliest two induction time points, SHF precursors are mainly induced during the latest two induction time points.

We note that the present scheme provides a general framework, which can be used to decipher the fate behavior and potency of progenitors using inducible genetic labeling methods in other developing tissues.

## Induction frequency and clone fragmentation

The fragmentation of clones into separate clusters complicates the interpretation of clonal fate data. As both genetic labeling of cells and clone fragmentation happen in a stochastic manner, one finds a broad distribution of fragment numbers in labeled hearts. The number of precursors associated with such fragments is therefore not straightforwardly inferable from the data, as neither the induction frequency nor the degree of fragmentation is known (Fig. 3c). Fortunately, by addressing a statistical ensemble of labeled hearts, we can make use of statistical inference to assign with known confidence the provenance of the observed fragments. To this end, we first identify the induction frequency and degree of fragmentation for the heart as a whole by pooling data from all of the labeled hearts. With this result, we can then identify which of the collection of monochromatic patches in a heart are derived from a single induced cell. The basic strategy is illustrated in Figs. 3d,f. Restricting our analysis to monoclonal fragments, we then address the question of lineage potential. In addition, by analyzing the data by induction time using all labeled hearts, we also reveal the timing of lineage specification in the first and second heart field.

### *Induction frequency of Mesp1 progenitors following inducible labeling*

To analyze the clonal fate data, let us begin by defining the probability, $p$, that following Dox administration, an early Mesp1 expressing progenitor cell becomes induced. Of course, this probability may vary according to the specific color of the fluorescent reporter gene. However, for now, let us consider just one of the colors and later generalize to multiple color

combinations. Then, if there are a total of $N$ Mesp1 expressing progenitor cells at the time of induction, if the induction probability of each cell is considered statistically uncorrelated with its neighbors, the probability distribution for the number of induced cells for a given color is given by the binomial distribution,

$$P(m) = \binom{N}{m} p^m (1-p)^{N-m},$$

where the binomial coefficient is defined by $\binom{n}{k} = n!/[k!\,(n-k)!]$. Then, if the induction probability is clonal (i.e. $p$ is of the order of $1/N$), we can make a Poisson approximation,

$$P(m) \approx \frac{N^m}{m!} e^{-pN} p^m = \frac{(pN)^m}{m!} e^{-pN}.$$

In particular, the probability that the tissue remains completely unlabeled is given by $e^{-pN}$ and, as expected, the mean number of induced cells is $\langle m \rangle = pN$. Let us now consider the potential for Mesp1 cell-derived clones to undergo fragmentation.

*Clone fragmentation*

Once a precursor cell has been induced, in the course of its clonal expansion through cell proliferation, cells may disperse and the clone may fragment into multiple subclones. To account for this process of fragmentation, we may once again model these events as a statistically uncorrelated Poisson random process, so that the probability that an individual clone ends up in $k$ fragments (i.e. it undergoes $k-1$ fragmentations) is given by

$$R(k) \approx \frac{f^{k-1}}{(k-1)!} e^{-f},$$

where $f$ denotes the degree of fragmentation, defined as the average number of fragmentations experienced by a single cell-derived clone over the time course from induction to analysis. The degree of fragmentation represents the time-integral of the underlying fragmentation rate, which may itself vary over time. Of course, the degree of fragmentation may depend on the total size of the clone, i.e. large clones may fragment more than small clones. To investigate this, we calculated the surface area (SA) of clones in unicolor hearts, i.e. the percentage of the heart's surface clones cover ($n = 18$). We indeed found that clones vary significantly in SA at each induction time. However, comparing the size of these clones to the number of fragments did not show any significant correlation (Spearman's rank correlation, $\rho = 0.19$, $p = 0.45$). Therefore, since we will later see that most of these hearts are monoclonal, there is no evidence in the data that the degree of fragmentation depends on the size of clones.

With this definition, what then is the probability distribution of finding a total of $k$ labeled fragments if $m$ cells of a common color have been induced? In this case the number of fragmentation events is given by the total number of fragments minus the number of induced cells, $k-m$. Then, taking the fragmentation and induction events to be statistically independent, the branching probability, $S(k|m)$, is described by a Poisson process with an effective rate $m \cdot f$, and

$$S(k|m) = \frac{(mf)^{k-m}}{(k-m)!} e^{-mf},$$

where $k \geq m$. Therefore, with this result, we can infer the joint probability distribution for finding a heart with *m* induced cells giving rise to a total of $k$ fragments of a given color,

$$J(m,k) = S(k|m)P(m) = \frac{(mf)^{k-m}}{(k-m)!}\frac{(pN)^m}{m!}e^{-mf-pN},$$

with $k \geq m$. (Note that, as defined here, the number of fragments must obviously be bound by the number of induced cells.)

In practice, in any given experiment, only the total number of labeled fragments (of a given color) is accessible – the underlying number of induced cells (clones) cannot be recovered for any given cluster of fragments. Moreover, we only have access to the frequency of clone fragments when at least one cell has been induced. The frequency of non-induced hearts is not recorded. Therefore, we should exclude the contribution of $m = 0$ from the statistical ensemble. In this case, the joint size distribution of "labeled" clones is therefore given by

$$J^{labeld}(m,k) = \frac{1}{(1-e^{-pN})}J(m,k)$$

where $m > 0$. For these persisting clones, since we measure only clonal fragments, we should combine all possible induction outcomes, from which we obtain the persisting fragment distribution,

$$F(k) = \sum_{m=1}^{k} J^{labeled}(m,k) = \frac{e^{-pN}}{1-e^{-pN}}\left[\sum_{m=1}^{k}\frac{(mf)^{k-m}}{(k-m)!}\frac{(pN)^m}{m!}e^{-mf}\right].$$

From this expression, we find that the average number of labeled fragments is given by

$$\langle k\rangle^{labeled} = \frac{pN(1+f)}{1-e^{-pN}}.$$

Hence, the average number of fragments in labeled hearts increases linearly with the degree of fragmentation $f$ and, for moderately large values of $pN$, linearly with the induction frequency.

*Fitting the data*

Already at this point we may try to give an estimate on the values of $pN$ and $f$ for the heart as a whole. For our analysis, we do not take into accounts hearts, which are labeled in a specific color in the epicardium. The reason for this choice is that the outer unicellular layer is formed by cell migration quite late compared to the induction time (at E9.5), leading to very dispersed cells across the epicardium. This makes it difficult to distinguish labeled cells in the epicardium from those in the IFT and OFT. Making use of the formula for $F$ and explicitly denoting its dependence on the parameters, we calculate the probability that the observed fragment numbers are found for any given degree of fragmentation, $f$, and induction frequency, $pN$. As the observations $k_1, k_2, \ldots$ are statistically independent this probability is given by:

$$F(\{k_1, k_2, \ldots\}|pN, f) = F(\{k_1|pN, f) \cdot F(k_2|pN, f) \cdot \ldots .$$

Treating $F$ as a function of $f$ and $pN$ (which is then generally called the *likelihood*), we may now ask for the maximum of this likelihood function: the values of $pN$ and $f$ that yield the experimental data with highest probability. We consider these values to be the best estimate for the degree of fragmentation and induction frequency. From this analysis, we find that $pN = 1.7 \pm 0.8$ and $f = 1.3 \pm 1.0$ (95% confidence intervals). The large confidence intervals reflect the fact that the maximum of the likelihood function cannot be precisely determined along a curve in the $pN$-$f$-plane (Supplementary Fig. S3a). In other words, all of these parameters fit the experimental data equally well within the limits of statistical significance. In the following we therefore develop an independent approach to further constrain the two fitting parameters.

The multicolor labeling assay provides a means to independently infer the induction frequency, $pN$. To understand how, consider first the probability that a heart remains unmarked in any one of the three colors following Dox administration (we do not consider the GFP+ contributions as the induction frequency of these cells is found to be negligible – only one heart was found to contain any GFP+ cells. If the relative induction frequency of the three colors (YFP, RFP, and CFP) is equal, then this probability is given by $J(0,0)^3 = e^{-3pN}$. Therefore, the probability that an induced tissue involves all three colors (regardless of the number of fragments) is given by the probability that the tissue is clonally labeled in all three colors divided by the probability that the tissue is labeled at all:

$$C_{tricolor}^{labeled} = \frac{[1 - J(0,0)]^3}{1 - J(0,0)^3}.$$

Similarly, the chance that an induced tissue involves two out of three colors is given by

$$C_{bicolor}^{labeled} = \frac{3J(0,0)[1 - J(0,0)]^2}{1 - J(0,0)^3},$$

while those that involve only one color is set by,

$$C_{unicolor}^{labeled} = \frac{3J(0,0)^2[1 - J(0,0)]}{1 - J(0,0)^3}.$$

Since these probabilities are independent of the fragmentation probability, $f$, they can be used to provide an independent estimate of the induction frequency, $pN$.

To estimate the induction frequency of cells, we could immediately apply the results above to investigate the relative frequency of unicolor, bicolor and tricolor clones. However, in this case, we have to exercise some caution: Analysis of the unicolor clones shows that the induction frequency of the CFP is significantly smaller than the RFP and YFP with only 3 CFP+ clones out of a total of 23. By contrast, both the bicolor and tricolor hearts have a roughly equal representation of the three colors: In hearts which are labeled in any of the two heart fields, the multiplicity of RFP:YFP:CFP is 23:22:13 for bicolor and 55:57:66 for tricolor. While the multiplicity of colors is far from perfectly equal in the bicolor case, the comparison of the fitted distribution with the experimental data will further validate our approach. Then, since

$$\frac{C_{bicolor}^{labeled}}{C_{tricolor}^{labeled}} = \frac{3\,J(0,0)}{1 - J(0,0)},$$

we can use the ratio of the number of bicolor to tricolor clones to infer the induction probability, $pN$. With a measured ratio of 1.10, we find that $pN = 1.31 \pm 0.05$. We performed the same calculations by explicitly taking the lower induction frequency of CFP in bicolor hearts into account. With this approach we obtain an average induction frequency, pN, of roughly 1.4 for all fluorescent markers in tricolor hearts and RFP and YFP in bicolor hearts, i.e. we find only a minor deviation for most observed hearts. The induction frequency of CFP in bicolor hearts involving CFP would correspondingly be about 0.7. Therefore, the pN value only changes significantly for CFP in bicolor hearts. As we will see below, this only marginally influences the outcome of the statistical analysis.

Therefore, on average, pooling all of the data from the three induction times, we expect that approximately 1.3 cells are induced per color in each heart. However, this estimate includes hearts where there are no marked clones at all. To obtain the induction frequency for clones that contain at least one marked cell we have to divide $P(m)$ by the probability that a heart is unlabeled in a given color, $1 - P(0)$. Therefore, when restricting attention to labeled hearts, the probability of $m$ induction events is given by

$$P^{labeled}(m) = \frac{P(m)}{1 - e^{pN}}.$$

Consequently, we obtain for the mean number of induced cells in labeled hearts

$$\langle m \rangle^{labeled} = \frac{pN}{1 - e^{-pN}},$$

which, in the present case, is approximately equal to 1.8.

With this estimate for $pN$, we may now turn to consider the probability distribution of fragment numbers, $F(k)$, and the estimate of the degree of fragmentation, $f$. By fixing the value of $pN$ we can restrict the possible parameters to a slice through the $pN$-$f$-plane (Supplementary Fig. S3a and b). Making use of the formula for $F(k)$, analysis of the maximum likelihood shows that $f = 1.6 \pm 0.2$. Both, the induction frequency and the degree of fragmentation are in agreement with the values obtained above.

As a further consistency check, we may note that, with these values of $pN$and $f$, the average number of fragments labeled hearts, $\langle k \rangle^{labeled}$, is given by 4.6, which compares excellently with the experimentally measured value of 4.7. Indeed, with the fitted values, the predicted fragment number distribution compares favorably with the measured distribution, as indicated in Fig. 3e.

Note that, to estimate $pN$, we made use of the fact that the induction frequency is roughly equal for all fluorescent markers for bicolor and tricolor hearts. However, for unicolor hearts, the frequencies of the different colors are manifestly different. If the statistical weights of the different colors in unicolor hearts were representative for all hearts, this might lead one to conclude that induction frequencies are also different for multicolor hearts. Since the degree of fragmentation should not depend on the color of fluorescent label, we may analyze the total fragment numbers, $\langle k \rangle^{labeled}$, to test whether or not the overall induction frequency does indeed depend on the color of the fluorescent marker. If unicolor hearts were representative for all hearts, one would expect that $\langle k \rangle^{labeled}$ depends sensitively on the induced color. Taking all labeled hearts (uni-, bi- and tricolor), we find that

the average number of fragments is $4.6 \pm 0.3$, $4.1 \pm 0.4$ and $5.4 \pm 0.3$ for YFP (n=87), CFP (n=83) and RFP (n=92), respectively, which suggests that the induction frequency is only weakly dependent on the color of the fluorescent marker. We attribute the apparently non-representative induction of unicolor hearts to a thresholding effect in which the sensitivity of different colors to induction is amplified when the level of Cre expression is low.

## Mesp1 positive cells are restricted to either the first or the second heart field

With the value of the induction frequency and degree of fragmentation fixed, we may now make an informed decision on which hearts are monoclonal. To begin, we note that the probability that $k$ patches derive from $m$ clones is given by

$$L(m|k) = \frac{J(m,k)}{F(k)}.$$

Therefore, the probability that $k$ fragments are of clonal origin is given by $L(m = 1|k) = J(m = 1|k)/F(k)$. Similarly, the probability that $k$ patches derive from more than one clone is obtained by summing over all induction outcomes larger than one,

$$L(m > 1|k) = \sum_{m=2}^{\infty} \frac{J(m,k)}{F(k)} = 1 - L(m = 1|k).$$

To make a decision on the maximum number of fragments we consider to be of clonal origin, we compare these two: we consider $k$ fragments to be monoclonal if the probability that that they stem from a single cell, $L(m = 1|k)$, is larger than the probability that they stem from more than one cell, $L(m > 1|k)$. Specifically, in the spirit of the theory of Bayesian inference, we compute the logarithm of these two probabilities and multiply by -2,

$$D = -2\ln\left[\frac{L(m > 1|k)}{L(m = 1|k)}\right].$$

With this definition, fragments are considered monoclonal if $D > 0$. Taking the values for $pN$ and $f$ obtained in the previous section, we find that 3 or less fragments of a single color are likely monoclonal, cf. Fig. 3 g and h. Indeed, with this classification, we expect that some 12% of hearts designated as monoclonal would in fact be polyclonal.

How does the approximation of equal induction frequencies in bicolor hearts affect this threshold value? If the type of fluorescent protein does not influence the degree of fragmentation, this would mean that the likelihood that a given number of patches is monoclonal is higher for CFP than for the other fluorescent markers in bicolor hearts. In other words, treating the induction frequency separately would allow us to treat some bicolor hearts as monoclonal, which have slightly more than 3 patches. Hence, treating the induction frequencies of different colors separately only marginally increases the sample size of monoclonal hearts.

With these results we may now restrict our analysis to hearts, in which a single clone has been labeled per color. Remarkably, we find that, of the 89 cases of hearts that are deemed to have marked fragments of clonal origin in either the FHF (LV) or the SHF (OFT and IFT), all are restricted to one or the other heart field. None of these clones contribute to both heart

fields. (We note that this apparently perfect segregation of clones is further assisted by the histogenesis which, as we will see below, leads to the temporal separation in the specification of progenitors of the two heart fields.) By contrast, of the 69 clones that have fragments in the FHF, 15% also have fragments in the other heart compartments (i.e. the RV and the RA). Similarly, of the 20 clones that have fragments in the SHF, 55% have fragments in other heart compartments. Figure. 3i shows the percentage of clones that contribute to the different heart compartments given that they contribute to the FHF (left) or SHF (right), respectively. We conclude that, by the time of induction, Mesp1+ cells are already lineage restricted, contributing to either the first or second heart field, but not both. However, both Mesp1+ subpopulations are able to contribute to cells in the remaining heart compartments.

To further scrutinize the properties of Mesp1+ cells we calculated the non conditional probabilities with which these cells contribute fragments to the different heart compartments (Supplementary Fig. 5c). For example, we find that about 10% of the fragments of FHF precursors end up in other compartments. This means, as 85% of FHF precursors exclusively contribute to the FHF, that from the remaining 15% approximately two out three fragments end up in other heart compartments.

These clones have an average number of fragments of $k_F = 2.10 \pm 0.01$ and $k_S = 2.60 \pm 0.02$ for FHF and SHF precursors, respectively. Taking into account the fact that, by introducing a threshold of $k = 3$, we neglect clones with a large number of fragments (i.e. those lying in the tail of $L(m = 1|k)$), this result agrees well with the predicted value for the overall population, viz. $f + 1 = 2.6$. This also tells us that fragmentation of SHF precursors is slightly higher than fragmentation of FHF precursors, which raises the question of whether the former might migrate more.

As a consistency check we may estimate the induction frequency and the degree of fragmentation of the two types of precursors independently by following the steps from the previous section. Since most of the contributions of these cells go into the FHF (LV) and the SHF (OFT and IFT), respectively, we restrict our analysis to fragments in these compartments. For the FHF precursors we find that $pN = 1.07 \pm 0.07$ and $f = 0.78 \pm 0.17$ for FHF precursors and $pN = 0.39 \pm 0.16$ and $f = 1.00 \pm 0.22$ for SHF precursors (the values for $pN$ are shown in Supplementary Fig. S3d). On the one hand, these results are in agreement with a higher degree of fragmentation of SHF precursors. On the other hand, this tells us that the induction frequency is significantly higher for FHF precursors. Moreover, noting that, as the overall induction frequency is the sum of the individual induction frequencies, the individual values for $pN$ are in good agreement with the values obtained for the whole population of Mesp1 expressing cells. The fragmentation rates are, expectedly, lower, as we neglected fragments located in heart compartments other than the LV, OFT and IFT.

With the probability of single clones to contribute to the different heart compartments defined, we may now predict the overall distribution of fragments in all hearts. To this end, we may account for the neglect of large monoclonal clusters by calculating the effective degree of fragmentation of the two subpopulations as follows: $\bar{k}_{F,S} = 2k_{F,S}/(k_F + k_S) \cdot (f + 1)$. In other words, we use the monoclonal data to infer the relative deviation of each subpopulation from the average number of fragments of a single clone, $f + 1$, in the overall population. From this we obtain an estimate for the fragmentation rates of each subpopulation, viz. $f_F = 1.4 \pm 0.2$ and $f_S = 1.9 \pm 0.3$ (Supplementary Fig. S3d). Since the 95% confidence interval of the difference between these two values does not contain 0, this

difference in fragmentation rates is statistically significant. With this result, we are then able to predict the experimentally observed distribution of fragments in all hearts with remarkable accuracy, cf. Supplementary Fig. S3e. One notable exception is the number of fragments in the RV, which is twice as large as that expected. We attribute this apparent discrepancy to the fact that there are clones that exclusively contribute to the RV. These are not included in the analysis of the monoclonal hearts, but they do, of course, contribute to the overall distribution of fragments.

## Temporal induction of the FHF and SHF progenitors

To investigate the temporal order of fate specification we now take into account the time point of Dox administration. First, we address the proportion of FHF and SHF precursors that are labeled at each induction time. From the previous results, we know that FHF and SHF precursors are mutually exclusive with respect to their contributions to the LV (FHF) on the one hand and the OFT and IFT (SHF) on the other. As the labeling of clones can be considered statistically independent, the average number of induced FHF and SHF precursors is proportional to the average number of fragments in these compartments $\langle m_{F,S}\rangle(t) = p_{F,S}(t)N_{F,S}(t) \propto \langle k_{F,S}\rangle(t) = K_{F,S}(t)/H$, irrespective of whether or not the hearts are monoclonal. Here, $p_{F,S}(t)$ denotes the induction probability of a single FHF or SHF derivatives, respectively, $N_{F,S}(t)$ are the total numbers of FHF or SHF derivatives in the early tissue, and $\langle k_{F,S}\rangle(t)$ signifies the average number of fragments in the corresponding heart compartments. The total number of heterozygotic mice, $H$, is, by the design of the experiment, independent of the induction time. With this, we can calculate the proportion of induced cells at each time point $t$ of Dox administration, $r_{F,S}(t)$. To this end, we divide the average number of clones that were induced at time $t$, $\langle m_{F,S}\rangle(t)$, by the total number of induced clones, to obtain

$$r_{F,S}(t) = \frac{K_{F,S}(t)}{\sum_t K_{F,S}(t)} = \frac{p_{F,S}(t)N_{F,S}(t)}{\sum_t p_{F,S}(t)N_{F,S}(t)}.$$

Then, since the probability $p_{F,S}(t)$ of a single Mesp1 expressing cell to be induced should not depend on the particular time point of induction, we can make the simplification

$$r_{F,S}(t) = \frac{N_{F,S}(t)}{\sum_t N_{F,S}(t)}.$$

Therefore, the ratio represents the proportion of FHF and SHF derivatives that are induced at time $t$. Importantly, this proportion can be estimated by analyzing the total numbers of fragments in *all* hearts. From this analysis, we find that most FHF derivatives are induced at induction times E6.25 and E6.75 (89%) while most SHF derivatives are labeled at induction times E6.75 and E7.25 (95%), cf. Fig. 3k.

Finally, one may also use the analysis of data from the monoclonal fragments alone as a consistency check. Here, the number of induced clones, $m_1(t)$ and $m_2(t)$ are directly accessible. In agreement with the results incorporating all hearts we find that FHF precursors are mostly induced early (E6.25 and E6.75) and SHF progenitors are mostly induced late (E7.25). However, in contrast to the results obtained from the full data set, none or very few of the SHF precursors are induced at E6.75. We attribute this to the fact that, at this time point, Mesp1 is only expressed at low levels in SHF precursors. As a result, these cells will

only be induced at induction frequencies higher than the relatively low doses that define the monoclonal data points.

We may also infer the total numbers of induced clones of each subpopulation for each time point by dividing the total number of fragments, $K_{F,S}(t)$, by the average number of fragments that single clones contribute to the FHF or SHF, respectively, $\bar{k}_{F,S} \cdot \pi_{F,S}$: $m_{F,S}(t) = K_{(F,S)}(t)/(\bar{k}_{F,S} \cdot \pi_{F,S})$. Here, $\bar{k}_{F,S}$ is the corrected overall number of fragments and $\pi_{F,S}$ denotes the probability that a fragment of such a single clone ends up in the FHF (LV) or the SHF (OFT and IFT), respectively, cf. Supplementary Fig. S3c. With $\langle m_{F,S} \rangle(t) = m_{F,S}(t)/H$, of course, as neither $\bar{k}_{F,S}$ nor $\pi_{F,S}$ do not significantly depend on time, this exactly reproduces $r_{F,S}(t)$.

From this analysis, we find that, overall, $254 \pm 22$ FHF and $138 \pm 17$ SHF precursors have been induced. Hence, FHF precursors have roughly twice the induction frequency, $pN$, of SHF precursors, which compares favorably with the estimated induction frequencies for the two subpopulations that we obtained by comparing the numbers of tricolor and bicolor hearts. Given that clone induction is statistically independent, the numbers of FHF and SHF precursors follows a binomial distribution. In this case we may employ *Fisher's exact test* to calculate the probability that differences in the number of induced clones between two induction times are the result of pure chance (Fisher, 1922). We find that this probability is small when comparing any two induction times. The differences in the number of lineage specified cells are statistically highly significant ($p < 0.0001$). Hence, we find that Mesp1 positive cells consist of two temporally distinct subpopulations. While FHF derivatives are largely specified early, most SHF derivatives are induced at the latest time points.

This completes the quantitative statistical analysis of the clonal fate data. In summary, making use of a multicolor labeling strategy, we employed statistical inference to estimate the induction frequency and the degree of fragmentation in a pooled dataset. This allowed us, for a given color, to identify the ensemble of monoclonal hearts. Restricting the analysis to these hearts we showed that Mesp1 expressing cells are already committed to either contributing to the first heart field or the second heart field. We calculated the contribution of these two types of precursors to the different heart compartments and showed that the precursors to the two heart fields are induced in two distinct temporal regimes.

## Bibliography

Fisher, R. A. (1922). On the interpretation of χ2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society , 85* (1), 87-94.