**A single cell transcriptomic and clonal analysis depict valvulogenesis**

Batoul Farhat[1], Ignacio Bordeu[2,3,4], Bernd Jagla[5], Hugo Blanc[6], Karine Loulier[7], Benjamin D. Simons[2,3,8], Emmanuel Beaurepaire[6], Jean Livet[7] and Michel Pucéat[1]

*1 INSERM U1251/MMG Aix-Marseille University, Marseille, France*

2 *Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK*

3 *Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK*

*4. Departamento de Física, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile*

*5 Pasteur Institute UtechS CB & Hub de Bioinformatique et Biostatistiques, C3BI, Paris*

*6 Laboratory for Optics and Biosciences, Ecole polytechnique, CNRS, INSERM, IP Paris, Palaiseau, France*

*7 Sorbonne Université, INSERM, CNRS, Institut de la Vision, Paris, France*

8 Wellcome Trust-Medical Research Council Stem Cell Institute, Jeffrey Cheah Biomedical Centre, University of Cambridge, Cambridge CB2 A0W, UK

**Mitral valve prolapse is often associated with several congenital heart malformations. During embryonic development, endocardial cells of the atrioventricular canal (AVC) undergo endothelial to mesenchymal transition (EndMT) to give rise to the different mitral valvular cells. However, both the identity and fate decisions of these endocardial cells during development remain unknown. Here, using single-cell RNA sequencing (scRNA-seq) of genetically labeled AVC endocardial cells in embryonic murine hearts, we uncover and genetically characterize a restricted population of pro-EndMT cells. We further identify distinct populations of valve progenitors that together contribute to different types of endothelial and interstitial valvular cells in late embryonic and early postnatal murine mitral valve. Moreover, using clonal genetic tracing with multicolor reporters of the endocardial cells, we observed specific modes of growth of endocardial cell-derived clones in the mitral valve. Collectively, our data reveal the identity of specified subtypes of endocardial cells and the distinct types of clonal contribution of these cells to the formation of the mitral valve. They further open the path towards finding new cellular targets for a specific therapeutic approach of congenital and acquired valve diseases.**

## INTRODUCTION

During evolution, as organisms adapted to new environments, the heart has grown in size and acquired a complex structure, ranging from a simple contractile vessel in Amphoxius to a four-chambers organ in mammals. These changes have taken place in concert with the development of valves that ensure unidirectional blood flow.

Malformation of valves accounts for around 10% of congenital heart diseases[1]. Mutations of a few genes have been reported as responsible for valves defects. These include genes encoding cytoskeleton proteins fibrillin and filamin A[2,3], TGFβ receptors and the downstream signaling smad pathway[4] and, more recently, Dachsous, potentially LMCD1, tensin and DZIP[5,6]. Altogether, mutations of these genes account for less than 5% of genetic valvulopathies. This suggests that, besides mutation of genes, other patho-biological processes such as defects in cell fate determination, in lineage specification and/or in cell migration might be responsible for valve diseases.

Valve development is a complex biological process involving several cell lineages first specified and determined by specific gene transcriptional changes, influenced by many extrinsic factors and cellular events including growth factors, cell-cell and cell-matrix interactions and mechanical cues[7]. Valves form as early as embryonic stage E9.0-E10 in the mouse embryo and develop as cushions in two discrete cardiac regions, the atrioventricular canal (AVC) and the outflow tract (OFT). The AVC valves comprise the mitral and tricuspid valves, while the OFT or semilunar valves include the aortic and pulmonary valves[5]. The mitral valve is the most affected valve in cardiac congenital diseases.

Valve formation commences with the delamination of endocardial cells that become separated from the myocardium by a cardiac jelly. This event is followed by endocardial cell transition toward a mesenchymal phenotype. Cell proliferation within an extracellular matrix forms the two cushions. The endothelial-to-mesenchymal transition (endMT) is triggered by a series of morphogenetic factors and signaling pathways including TGFβ/BMP-smad, Wnt/TCF, and Notch4 secreted by the myocardium as well as NFATc mediated pathways[8] likely governed by a set of specific genes and transcriptional pathways. The most intriguing feature of this biological process is that it is restricted within a set of a few endocardial cells. Besides the signaling pathways, little is known about the biological processes and transcriptional networks required for the cell to initiate local EndMT. Furthermore, the

specific transcriptomic signature or biological properties that confer cells a competency for EndMT are still unknown.

Afterwards, a series of morphogenetic events allows the cushion cells to further differentiate into specific cell types (e.g., fibroblast, smooth muscle cells, chondrogenic cells) that proliferate and expand to build up the valve leaflets. It remains to be known how early lineage restriction occurs and how do cells expand, move and distribute within the valve leaflet[9].

To address these questions, we used a single cell-RNA sequencing approach combined with clonal analysis with multicolor reporters and uncovered the origin and transcriptomic signature of valvular endocardial cells, and characterized the behavior of the valvular endothelial (VECs) and interstitial (VICs) cells during formation of mitral valve leaflets.

Our results show that endocardial cells constitute a heterogeneous cell population within the AVC. A subset of these cells[10], which feature a unique gene expression profile, are able to acquire EndMT competency and commit toward a specific valvular cell type. Based on our clonal and histological analyses, we estimate the number of self-renewing progenitors required to building a valve leaflet, and we observed multiple behaviors of VICs within the leaflet allowing for the formation of a functional valve.

## RESULTS

### Single-cell transcriptomics reveals cell specification in the atrioventricular canal

To interrogate the endocardial cell heterogeneity within the AVC, we first used a single cell-RNA sequencing approach. The AVC was carefully dissected out from 60 (E9.5) embryos (19-22 somites) from B6.Cg-Gt(ROSA)26Sortm9(CAG-tdTomato)Hze/J) females bred with Tie2$^{Cre}$ males.

Cells were enzymatically dissociated, FACS sorted, and loaded into the chromium (*10X genomics*). 1500 cells were sequenced and 1200 which passed the quality control test used in the analysis. The algorithm for dimension reduction-generated Uniform Manifold Approximation and Projection (UMAP) derived plot (Fig 1a) revealed heterogeneity of the whole cell population. The AVC cell population featured an average of 5530 genes/cell and comprised one population of endocardial cells, two populations of hybrid cells undergoing EndMT, a subpopulation of late hybrid /mesenchymal cells as well as chondro-osteogenic progenitors and hemogenic cells. A small cluster of macrophages was also observed (Fig.1a).

The heatmap (Fig. 1b) further revealed a broad spectrum of gene expression expected to mark future VIC and VEC cell populations. More specifically, we found one subpopulation

of *eng*$^{+}$*, pecam*$^{+}$*, egfl7*$^{+}$*, cdh5*$^{+}$*, npr3*$^{+}$*, tek*$^{+}$*, tie1*$^{+}$*, emcm*$^{+}$ endocardial cells (cluster 3), a subpopulation of *zeb2*$^{+}$*, prrx1/2*$^{+}$*, twist1,2*$^{+}$*, msx1*$^{+}$*, cldn7*$^{+}$*, epcam*$^{+}$*, efnb1*$^{+}$*, efna2*$^{+}$ cells starting the transition from endothelial to mesenchymal fate (early hybrid, cluster 5), two subpopulations$^{-}$ of high *tagln*$^{+}$*, col2a1*$^{+}$ *and twist1,2*$^{-}$ hybrid cells (clusters 0 and 2), and a subpopulation of *tagln*$^{+}$*, col2a1*$^{+}$*, fbln1*$^{+}$*, and alcam+* late hybrid cells transitioning toward mesenchymal cells. We also observed subpopulations of *cdh11*$^{+}$*, sox9*$^{+}$*, sparc*$^{+}$*, Tbx18+, col3a1*$^{+}$*, cthrc1*$^{+}$*, Gata6*$^{+}$*, itgb5*$^{+}$*, cxcl12*$^{+}$*, maoa*$^{+}$ chondro-osteoblastic progenitor cells (cluster 4), of *tal1(scl)*$^{+}$*, Gata1*$^{+}$, *nfe2*$^{+}$, *klf1*$^{+}$ hemogenic cells (cluster 1) and a small cluster of *Plek*$^{+}$*, tyrobp*$^{+}$*, Gata1*$^{+}$, *nef2*$^{+}$*, fcer1g*$^{+}$ macrophages.

EndMT is the first step of valvulogenesis. We thus first addressed the question of the EndMT competency of these cells. Epithelial-to-mesenchymal transition (EMT) is no longer understood as a binary process but has been described as a continuum transition[10]. EndMT is likely to follow such a step-by-step transition from endocardial to mesenchymal cell phenotype. We first interrogated a few genes that we surmised to play a role in EMT. The panel plots in Figure 1c shows only endocardial cells (marked by *pecam1, eng, cdh5* and *egfl7*) and early hybrid cells (marked by *twist1, twist2*, *prrx1, prrx2* ). The plots reveal that only part of endocardial cells expressed *twist1, twist2,* together with *prrx2*, an early marker of EMT. Interestingly these cells also expressed *snai1*, *hand2*, *klf2* and *msx1*, known to play a role in EndMT[11,12]. These cells lost claudin 5, an adherens junction gene while early hybrid cells gained claudin 6 and 7, more reminiscent of migrating mesenchymal cells. *TGF$\beta$1* and *Notch1* were lost in early hybrid cells while expressing *Notch3*, *Wnt2*, *Jag1*, *ptch1* and *pdgfr$\alpha$.*

A full analysis of differential gene expression between cluster 3 (endocardial cells) and 5 (early hybrid cells) can be found in extended data 1.

To further dissect this transition, we used an inference trajectory analysis. Interestingly, this bioinformatics approach suggested that two subpopulations of endocardial cells gave rise first to early *twist1*$^{+}$ hybrid cells. The cells then progressed towards hybrid mesenchymal cells still retaining expression of genes reminiscent of endothelial cells. Three branches emerged from these cells that are engaged into the process of EMT. They gave rise to either chondro-osteogenic cells, remained as hybrid cells, or moved towards late hybrid and hemogenic cells (Fig.2a). Macrophages remained detached from the trajectory indicating that they were potentially contaminating cells.

More specifically, we looked at signaling pathways that could be involved in EndMT (Fig. 2b).BMP2 was expressed only in hybrid cells (cluster 0) while TGF*β*2 was found in both early and late hybrid cells (clusters 0 and 5) but not in endocardial cells, which expressed TGF*β*1. While *TGFβR2* receptor was expressed at very low levels, *TGFβR1* receptor was expressed in both early and late hybrid cells but only at low level in endocardial cells. Late hybrid cells included cells co-expressing *TGFβ1*, *TGFβ2* and the *TGFβR1*. The VEGFR *flit1* was expressed only in endocardial cells. The *notch2* ligand and *Jag1* receptor genes were expressed mostly in early hybrid and chondrogenic cells. *SHH* was expressed in early hybrid cells, while its receptor *Ptch1* was more broadly expressed in early hybrid and chondrogenic cells but not in endocardial cells. *IGF1* was also mostly expressed in endocardial cells, while *IGF2* was expressed in all cells. *P2rx4* and *P2yr1*, the only purinergic receptors expressed, were only detected in rare cells (Fig 2b). The violin plots confirmed that a combination of nfatc1, notch1 and SHH pathways whose genes are expressed in some endocardial cells but not any longer in early hybrid cells (Fig 2c) emerge as a cell signaling signature of EMT competent cells

Interestingly, genes encoding enzymes of the fatty acid oxidation process, the acetyl-CoA acyltransferase 2 (*caa2*), the acetyl Co-A transferase (*acly*), the acetyl-CoA decarboxylase (*acss2*), the diacylglycerol O-acyltransferase 1 (*dgat1*), and the mitochondrial enoyl-Co A hydratase (*echs1*) were expressed all along the EMT process in restricted cell populations and in a few endocardial cells Their expression increased in early hybrid *twist1*$^{+}$ cells, and to a low extent in late hybrid cells (Fig.2d). The glycolytic enzymes enolase (*eno3*), phosphoglycerate mutase *pgam2*, acetyl CoA dehydrogenase (*acadl*), and phosphofructokinase (*pfkp*) were co-expressed in hybrid and late hybrid cells but much less in endocardial or early hybrid cells. *Ogdh* encoding Oxoglutarate Dehydrogenase, a citric acid cycle enzyme that plays a key role in replenishing TCA cycle intermediate was expressed all along the EMT process.

We next picked up proteins encoded by genes of interest in the EndMT process and carried out immunofluorescence experiments on sections of E9.5 embryonic heart. Vcam1 was barely or not expressed in endocardial pecam+ cells and Twist1, a protein involved in early EMT, featured a salt and pepper pattern in the endocardium within the AVC as shown in the panel plot (Fig 3). TWIST was still located into the cell cytosol and TWIST+ cells also co-expressed Pecam1. MSX1, another marker of EndoMT turned on by Bmp2 signaling from the myocardium, was observed in delaminated cells migrating from the endocardium to the cushion. The fourth marker, the ATP-citrate lyase (ACLY), was expressed in some

endocardial cells, as well as in cells migrating within the cushion still expressing *pecam1*, namely, early hybrid cells. Two other markers for late hybrid EndoMT and mesenchymal cells were validated: ITGAV was expressed in endocardial and post-EndoMT cells in the cushion, and ITGB3 was only expressed in rare delaminated endocardial cells. MSX1 was also expressed in very few cells of delaminated endocardium (Fig. 3).

These observations demonstrate that an activation of cell metabolism toward the Krebs cycle and fatty acid oxidation together with specific activation of TGFβ, Notch, and SHH pathways confer to *hand2*$^{+}$*, notch1*$^{+}$*, cldn5*$^{+}$*,msx1*$^{+}$ endocardial cells a potential to undergo EMT.

**Transcriptomic profile of embryonic and postnatal mitral valve depicts early valvulogenesis**

We collected embryos and neonates generated from B6.Cg-Gt(ROSA)26Sortm9(CAG-tdTomato)Hze/J) females bred with Tie2$^{Cre}$ males at E16.5 and P1, respectively . Hearts were dissected out and the annulus of the mitral valve including the leaflets were carefully cut out, enzymatically dissociated, and pooled together from 92 embryos at the same stage of development. The isolated cells were sorted by FACS and altogether 954 Tomato+ cells were then subjected to single-cell RNA sequencing and analysis.

The UMAP revealed that E16.5 valvular cells segregated between VEC expressing *pecam1, cdh5, and eng* and VICs expressing markers of fibroblasts, *postn and dcn* (Fig. 4a). VECs distributed into two clusters including one with cells expressing *wt1,* suggesting that these VECs belong to the parietal leaflet in which epicardial cells contribute[13] (Fig. 4a,4b; extended data Fig 1).

E16.5 valvular cells segregated from the neonatal valvular cells as well as AVC cells segregated from E16.5 and P1 valvular cells as shown in the UMAP plot (Fig. 4c). The heatmap including AVC, E16.5 embryonic and P1 neonatal mitral valve cells revealed the presence of numerous valvular genes (Fig. 4d). *Egfl7, pecam1, cdh5, eng, ecscr, tie1, cd93, thbd, and emcn* were expressed in endocardial cells *(*AVC cluster 4*),* and VEC of E16.5 mitral valve (cluster 7) and of neonatal mitral valve (clusters 7 and 8). Interestingly some E16.5 VEC and neonatal VEC (cluster 7 and 8) expressed *CD36* and *fabp4*, two markers of fatty acid metabolism

VICs of both E16.5 and P1 embryonic and neonatal mice (clusters 2 and 6) expressed numerous markers of VIC of adult valve layers (spongiosa, fibrosa and atrialis)[7]. Indeed, we

interrogated a list of genes specific of spongiosa *(dpt, postn, thbs2, eln, col3a1, col1a1), fibrosa (col1a2, thbs1, bgn, Dkk3)* or atrialis enriched in elastin *(eln).* All genes were expressed in most cells of both E16.5 and neonatal mitral valves in clusters 2 (E16.5) and 6 (neonatal) even in cells that did not cluster in a layer-specific manner. Immune cells, including macrophages which expressed *tyrobp, plek, ctss, laptm5, runx1, cd83, aif1, lyz2 and cx3cr1,* were found in cluster 10 (Fig 4d).

We then applied linear trajectory inference via pseudotime analysis (using Elpigraph) to track the steps of valvulogenesis (from day E9.5 until post-birth). The trajectory described the multistep process by which AVC cells at E9.5 transition towards valvular cells. This analysis revealed a subset of *pecam1+, eng+, cdh5+* endocardial cells that give rise to the endothelial layers (VEC) of both E16.5 and neonatal valves. The endocardial cells also contributed to subsets of VICs. One subset at E16.5 led to the VICs in neonatal valve. The endocardial cells also generated hemogenic cells and underwent EMT to give rise to chondrogenic cells (Fig. 4e). We more specifically interrogated ECM-related genes in the three single-cell RNA-seq datasets (AVC, mitral valve at E16.5 and at neonatal stages). The heatmap in extended data Fig. 2 showed that clusters 1 and 4 (including only E16.5 and neonatal valvular cells) were enriched in collagen genes (*col1a1, col1a2, col3a1)*, reminiscent of the fibrosa. Clusters 5, 7 and 8 including E16.5 and neonatal mitral valvular cells were enriched in proteoglycans including perlecan (*hspg2)*, versican (*vcan),* lumican *(lum),* and elastin (*eln*), reminiscent of the spongiosa and atrialis. AVC cells from clusters 0, 2, 3 and 6 specifically expressed the galactose kinase (*galk1*), an enzyme required to synthetize proteoglycans, as well as the hyaluronan receptor CD44 and the hyaluronan synthase 2 (*has2*).

We further used Tempora[14] another cell trajectory inference using biological pathways information. We used a minimum size of gene sets of 50 and a maximum to 300 to allow the algorithm to use the most defined biological pathways.

This trajectory first revealed that cells differentiate step by step at each stage of development to finally contribute to post-natal valve leaflets.

The endocardial cells (cluster 4) contributed directly to the E16.5 VEC (cluster 7) and to neonatal VEC (cluster 8). The hemogenic endocardium (cluster1) also gave rise to E16.5 VEC (cluster 7) and contributes together with hybrid cells (cluster1) to VIC of neonatal valve (cluster 6).

Interestingly, VEC from cluster 5 and 8 from neonatal valve contribute to the VIC (cluster 6) of the valve. This suggests that the VECs are still able to undergo EMT when the leaflet is formed already.

The significant GO terms revealed by the Tempora trajectory inference further pointed the role of the integrin and Nfκb pathways as well as cilia formation[15]. Integrin signaling largely contributed to the process of valvulogenesis. Dynamic of smooth muscle cells and modeling of collagen fibers as well as protein catabolic pathway emerged as important processes within the remodeling developing valve (extended Figure 3).

**AVC-derived valvular cell clones expand and fragment along the leaflets' axis**

The embryonic cell lineage at the origin of endocardial cells is diverse, which could explain their transcriptomic heterogeneity and the competency to EndMT being only restricted to specific cell types. We therefore questioned how post-EMT cells further acquire their fate and distribute within the valve leaflets. To explore the behavior of valvular cells during the formation of the valve leaflets, we used an experimental approach that is complementary to the analysis of their unique transcriptomic profile. We set up a multicolor lineage tracing approach using CAG-*CreER*$^{TM}$ male bred with Cytbow females. The latter line expresses a modified Brainbow transgene that stochastically triggers either cyan, or yellow, or red fluorescent protein expression upon Cre/lox recombination. The recombinase activity was induced at E8.5 using a low dose of tamoxifen (20 µg/g mouse) which sparsely label pre-EMT progenitors of valvular cells (extended data Fig. 4 a,b). Embryos were first collected at E9.5 to check the number of labelled endocardial cells. Extended data Figure 3a shows the labelling of a single endocardial cell at the proposed tamoxifen dosage. Phosphohistone 3 (PH3) immunostaining was used to score the rate of cell division of AVC endocardial cells at E8.5. A low cell proliferation rate was observed at E9.5 (extended data Fig. 4b) implying that cell division right after recombination was limited.

We first monitored labeled cells in the mitral valve of E16.5 embryos via color two-photon microscopy[16] of whole clarified hearts. Large volume 3D microscopic imaging revealed that labeled cells formed elongated clusters that distributed primarily along the long axis of the leaflet (extended data Fig. 5a). This morphology resulted in a high correlation between the 2D-projected area of cell clusters and their 3D distribution (Pearson's $r=0.91$, $p=0.01$, extended data Fig. 5b), which allowed us to base our clonal analysis on tissue sections. (Fig. 5a,b).

Within a septal or mural leaflet we could observe that putative clones of a given color where fragmented into smaller clusters of cells (Fig 5a). To infer whether a group of clusters of cells belonged to a single clone, we followed an established biostatistical strategy[17] that allowed us to estimate the likelihood that a group of clusters was monoclonal (see Methods for details).

Applying this approach to the full clonal data, we first estimated both the overall induction frequency, pN (the frequency at which a single leaflet progenitor is induced, times the total number of leaflet precursors), and the clone fragmentation rate f (the average number of fragments per clone) (Fig. 5b and extended data Fig. 6). By considering the combined data from both mural and septal leaflets, we estimated the overall induction frequency to be pN=0.5±0.3 (95% C.I.). By contrast, the fragmentation rate was estimated independently for each leaflet, giving f=1.6±0.1 for the septal leaflet, and a slightly lower rate of f=1.1±0.2 for the mural leaflet. From this, we found that, in the septal leaflet, groups of 4 or less cell clusters were likely to belong to a single clone while, in the mural leaflet, this threshold was reduced to 3 fragments (see Fig. 5b and Methods for details). Based on this estimate, out of the 106 clone candidates, 83% (54 clones for the septal and 34 clones for the mural leaflet) satisfied the constraint and could be considered as clonal with statistical confidence.

**Fragment and clone size distribution**

Focusing on the clones identified by our statistical analysis, the distributions of cluster size within clones for both leaflets were found to fit well with a log-normal dependence (Fig. 6a), with logarithmic mean $\mu s$ = 1.3 (2) and variance $\sigma s$ = 0.9 (2) for the septal, and $\mu m$ = 1.4 (2) and $\sigma m$ = 1.0 (2) for the mural leaflet. Indeed, such log-normal behavior has been shown to be a hallmark of clone fragmentation/coagulation processes[18-20], lending additional support to the assignment of clones from clusters. The fact that the size distribution of clusters for both leaflets was statistically indistinguishable suggested that both leaflets share a similar growth dynamic.

Similarly, the distribution of total clone sizes (comprising the sum of constituent clusters) could be fit well by a log-normal dependence (Fig. 6b). A log-normal distribution of clone sizes could result from a cooperative growth process in which the stochastic proliferation of cells in the same neighborhood is positively correlated (i.e., clones arise as the “product of a random number of symmetric cell divisions before differentiation”). To challenge this hypothesis, we tested whether the sizes of clones of different color in the same leaflet were positively correlated (see extended data Fig 7). From this, we found that indeed some clones

sizes where positively correlated, while other others were not, suggesting a dependence on the spatial location of individual clones, and observation that was consistent with the suggested growth dynamics.

**Estimation of the number of progenitors of the mural and septal leaflets**

We proceeded to estimate the number of progenitors contributing to the formation of the mural and septal leaflets. To this end, we estimated the average number of clones needed to cover the 2D-projected area of each of the leaflets. For the septal leaflet, we estimated that about 41±8 (SEM) dividing progenitors were required to cover the cross-section of the leaflet, while about 22±4 (SEM) progenitors were required in the case of the mural leaflet. Given the similarities observed in the clonal organization and estimation of number of progenitors in both leaflets, we questioned whether the pool of progenitors could be the same for both leaflets, so that there is a unique common pool of about 40 progenitors that give rise to both leaflets. In this spirit, we analyzed likelihood that both mural and septal leaflets in heart have cells labelled in the same combination of colors and compared this result to the fraction expected from random combinations (null model) estimated from the unicolor and bicolor statistics (see extended Fig. 8). When comparing any two regions, and in particular the septal and mural leaflets, we observed a significant coincidence in the color expression of both leaflets compared to that expected from random combinations (Fig 6c). Taken together, these observations are consistent with a negative-binomial dependence of clone sizes and suggest that both leaflets may originate from a common pool of no less than 40 progenitors. This number was estimated by the number of progenitors required to cover the 2D-projected area of the leaflets, so the true number of progenitors is likely to be larger.

**Discussion**

Single-cell RNA sequencing of AVC endocardial cells and early cushion cells revealed their heterogeneity pointing to their plasticity and multipotency[21]. This further showed that a subset of endocardial cells, competent for EMT, were already primed toward a specific valvular cell type (smooth muscle cell, fibroblast, chondrogenic progenitors, endothelial cells of future endothelial layers of valve leaflet).

First, we found that a small subset of endocardial cells acquired a competency to undergo EndMT. These cells expressed specific markers of EMT recapitulating the different stages of this process, as previously found in the context of cancer biology[22]. Thus, EndMT, which allows cells to migrating out of the endocardium to form the cushions, is a cell transformation

process and likely involves a finely tuned cell reprogramming process, as recapitulated in tumor cells[23]. Pathological EMT has been reported to be regulated by metabolism involving acetyl-CoA[24]. We further found that the endocardial cells competent for EndMT turned on a specific transcriptional program encoding proteins involved in acetyl CoA metabolism such as *acly, acss2* or *dgat.* This transcriptional program, which likely underlies the regulation of EndMT by metabolism[25], was turned on at the earliest step of EndMT and lasted throughout the process, as suggested by the prolonged expression of *acly* (Fig. 2).

*Ptch1* and *TGFβr1* are the two receptor genes together with the genes encoding the ligands *shh, TGFβ2* as well as *notch2* to be expressed in early *twist1*+, *prrx1,2*+ cells pointing to their role in the initiation of EndMT. Interestingly SHH signaling component early expressed during the process of EMT regulates fatty acid metabolism[26], while *Flt1* mainly expressed in endocardial cells prevents this metabolism rather favoring glycolysis[27]. Our data thus suggest that only cells that specifically harbor *TGFβR1* co-expressed with *twist1* as well as *Notch2*$^+$ and *Ptch1*$^+$ are capable of undergoing EMT while activating their metabolism toward fatty acid oxidation, as indicated by expression of *acyl, echs1*, *acaa2*. Thus, a response to combinatorial TGFβ, Notch and SHH signaling pathways, a competence to switch metabolism and a specific transcriptional profile including expression of *hand2* and *msx1* are all required for endocardial cells to undergo EndMT and to be committed toward a valvular cell. This explains why only a few cells follow this program. These cells loss *cldn5*, a claudin that limits endothelial cell motility[28] and turn on expression of *cldn6* and *cldn7* involved in cancer metastasis[29] and in turn in migratory mesenchymal cells.

As recently reported as a general concept, endocardial cell state heterogeneity make cells either low or high responder to growth factors and downstream signaling to ensure specific and context-dependent cellular decision in a multicellular setting[30].

The EndMT transcriptional and cell metabolic reprograming events are then followed in mesenchymal cells by a genetic program specific to future valvular cell types. The latter is designed to allow cells to acquire a capability to secrete specific extracellular matrix proteins (ECM) (Extended data Fig.2). We found that, at E16.5, valvular cells expressed genes encoding ECM proteins found in layers of the valve leaflets, while a subset remained endothelial to form the surrounding layer of the leaflets. The process was more advanced at the neonatal stage, with the mitral valve featuring two different types of VICs expressing collagen genes as well as proteoglycan genes. Interestingly the VECs could be discriminated between parietal and septal valve leaflets depending upon the presence or

absence, respectively, of cells expressing *wt1*. Wt1 lineage tracing indeed was reported to contribute to VICs as well as VECs of the parietal mitral leaflet[11](Extended data Fig 1).
This suggests that some cells of (pro)epicardial origin expressing Tie2 at the time of induction of the recombinase retained expression of *wt1* even after migration into the valve leaflet. Interestingly VEC could still undergo EMT to provide after birth the leaflet with more VICs (Fig 4f). Hemogenic cells from endocardium also contributed to neonatal VIC likely through generation of macrophages to remodel the valve[31]. Altogether, these findings are in agreement with the long process of maturation of valves in early adult[32] in which macrophages from bone marrow also contribute to[33].
The VICs of both parietal and septal leaflets likely originated from the same AVC cells cluster (Fig. 4f) which was in agreement with the clonal analysis that also points to the same pool of progenitors shared by both valve leaflets (Fig. 6). Clonal analysis based on 2D leaflet areas predicted that at least 40 progenitors were required to cover the area of the septal leaflet and around 20 for the parietal leaflet. Considering the 3D morphology of the leaflets, this figure is broadly consistent with the total number of AVC cells that contribute to the E16.5 Mitral leaflets based on single-cell data (obtained by taking the number of cells in each contributing cluster (Fig. 4f) and dividing by the number of embryos, which led to a figure of around 140 AVC progenitors). This implied that the total clone size results from a growth process of differentiating progenitors that expand before terminally differentiating in VICs. Assuming a number of cells in a mitral valve leaflet of 400 in septal and 150 in mural leaflets at birth[15], this suggests that progenitor cells should undergo about three rounds of division on average.
Another interesting feature in the behavior of cells is their distribution along the long axis of the leaflet (extended data Fig 6). Such a cell behavior could explain the deposition of the ECM as layers, a characteristic of these valvular leaflets.
Altogether, our single-cell RNA-sequencing data combined with the clonal analysis revealed that the process of embryonic mitral valve formation starts early during development when AVC endocardial cells distribute between EMT-competent valve progenitors cells and endothelial/endocardial cells that will participate in lining the chambers during their growth, respectively. The EMT process is restricted to cells sharing specific cell identity (i.e., unique transcriptomic profile, competency to respond to specific agonists, and ability to undergo a metabolic cell reprogramming event). Restriction of such a cell population with the previous criteria may prevent any overgrowth of the cushions and lead to a well-tuned size of the leaflets, mandatory to ensure a perfect closure of the future valve.

We found also that cells specifically contribute to each anterior or posterior cushion and later to specific mural or septal leaflet. Other cells from a different embryonic origin[34] migrate later into these respective valve leaflets to complete the whole process of leaflets formation. The 2D distribution of the cell clones and their transcriptomic profile at E16.5 including ECM genes (Fig. 4, extended data Fig 2) indicate that the cell clones are specified toward the spongiosa, the fibrosa or the atrialis. The clones likely contribute to the formation of specific layer of the valve. Dysregulation of any of these processes may lead to a pathological valve.

**Acknowledgments**

This study was funded by the Leducq Foundation (MP, MITRAL network of excellence).

We are also grateful to the Leducq Foundation for generously awarding us for cell imaging facility (MP "Equipement de Recherche et Plateformes Technologiques" (ERPT)).

This study was also partly supported by Agence Nationale de la Recherche (contracts ANR-11-EQPX-0029 Morphoscope2 and ANR-10-INBS-04 France BioImaging, to EB and IHU FOReSIGHT, ANR-18-AHU-01 and ANR-19-CE16-0019 to JL).

We thank HALIODX (Marseille) for running the single-cell RNA-seq experiments

## Methods

### Mouse lines

*Rosa26–tdTomato and Tie2cre* mice were obtained from Jackson laboratory (B6;129S6-*Gt (ROSA)26Sortm14(CAG-tdTomato)Hze and B6.Cg-Tg(Tek-cre)12Flv/J, respectively)*. The WT1$^{CreERT2}$ mouse and *CAG-CreER$^{TM}$* mice[35] were kindly provided by Sylvia Evans (UCSD) and Joshua R. Sanes (Harvard University), respectively. The *Cytbow* mouse line broadly express a *CAG-Cytbow* transgene designed on the model of Loulier et al[36], modified to stochastically express mTurquoise2, mEYFP, or tdTomato from the broadly active *CAG* promoter following Cre recombination (this line will be described elsewhere).

Mice were kept under standardized conditions (22–24°C temperature; 50%±10% humidity) on a 12 h light/12 h dark cycle in an enriched environment (kraft paper enrichment). Food and tap water were provided ad libitum. The parental strains were bred on a C57BL/6 background.

*CAGCreER$^{TM+/-}$* mice were crossed with the transgenic mice *cytbow*. To obtain embryos of the desired stages (E9.5, E16.5), mating females were verified for the presence of vaginal plug every morning. The day of the presence of a vaginal plug was considered as embryonic day E0.5 at noon. Pregnant females were euthanized at the desired stage of gestation.

Tamoxifen (Sigma-Aldrich, France; 20 µg/g) was administrated with corn oil by oral gavage at embryonic day E8.5. Embryos were collected at E9.5 and E16.5, fixed overnight with 4% paraformaldehyde.

**Preparation of frozen sections**

Embryos at E9.5 and hearts at E16.5 were prepared for immunostainings. Samples were embedded in Optimum Cutting Temperature (O.C.T) compound (Tissue-Tek) for cryoprotection. Embryos at E9.5 and E16.5 were dissected in ice-cold PBS and fixed for 2hrs or overnight at 4°C in 4% PFA in PBS. Embryos were then transferred to 6% sucrose/PBS (w/v), 12.5% sucrose/PBS and 25% sucrose/PBS (1hr each) with gentle shaking at 4°C. Then the embryos were transferred into embedding molds, ad frozen in 50:50 (v/v) O.C.T/50% sucrose. Samples were stored at -80°C and sectioned (10µm, 20µm, 30µm) using a Leica Microm HM 500 OM Cryostat (-20°C). Frozen sections were mounted on superfrost plus slides and stored at -80°C before use.

**Immunofluorescence analysis**

Immunofluorescence was performed on frozen embryonic heart sections (Leica cryostat-10 µm). Sections were fixed with PFA for 10 minutes, washed with PBS 5 minutes, then blocked in PBS with 0.5% donkey serum, 5% BAS and 0.1% tween 20. Sections were stained with the following antibodies: anti-CD31 (BD550274 Purified Rat anti-mouse CD31; 1/100), anti-sarcomeric alpha actinin (mouse, Sigma-Aldrich A7811; 1/200) anti Twist1 (mouse 1/300 Novusbio NBP2), anti-ACLY (37364SS ACLY rabbit 1/200 Novusbio NBP2 – 67509), anti-Msx1 (rabbit 1/300 Novusbio NBP2 – 57021), anti-VECAM-1 (rabbit 1/250 Abcam ab-134047), anti-Integrin alpha V (rabbit 1/500 Abcam ab-179475) anti-Integrin beta 3 (rabbit 1/500 Abcam ab-119992). Nuclear staining was performed with DAPI (1/10000).

Images were acquired using a confocal LSM 800 Zeiss microscope equipped with an airryscan of 32 detectors. Light was provided by a laser module 405/488,561 and 640 nm wavelengths, respectively.

All images were acquired using the ZEN ZEISS software. Then some images were deconvoluted using Autoquant and reconstructed in 3D using Imaris software (IMARIS). All samples were mounted in Fluoromount™ (Cliniscience, France)

In all figures, tdTomato/mCherry, mEYFP, and mCerulean/mTurquoise2 are rendered in red, yellow, and blue, respectively.

**Two-photon imaging of whole embryonic hearts**

Large-scale color two-photon microscopy of whole embryonic hearts was performed on a custom-built imaging platform[37] combining trichromatic two-photon excitation through wavelength mixing[38] and 2D mosaicking of XYZ image series. Hearts were clarified following the CUBIC procedure with a modified final solution[39] and mounted between two coverslips and rigid spacers. Simultaneous excitation of CFP, YFP and tomato was performed by mixing 850 and 1150 nm pulses focused by a multiphoton water-immersion objective equipped with a correction collar (25×, 1.05NA, XLPLN25XWMP2, Olympus, Japan). Detection was performed on three independent detectors equipped with appropriate filters (Semrock FF01-475/64-25, Semrock FF01-542/27-25, Semrock FF01-607/70-25). The correction collar was set to obtain optimal contrast at a depth corresponding to the middle of the sample. A total volume of 1580 × 1580 × 300 μm3 was recorded in the form of 4× 4 XYZ stacks of dimensions 420 × 420 × 300 μm3 with voxel size 0.42 × 0.42 × 1.5 μm3. The volume was reconstructed using ** Fiji / bigstitcher**[40] and imported in Imaris (Bitplane, Switzerland) for visualization.

**Assessment of the recombination of the *cytbow* reporter by PCR**

Genomic DNA was extracted by DNA precipitation in NAOH solution (50 mM) from the tail of mice, wild type mouse and adult *cytbow* mouse as a control. Tails of embryos at E16.5 *CAG-CreER$^{TM+/-}$/cytbow* of the whole litters were used to genotype embryonic mice with specific primers.

**Mouse Genotyping Primers**

*CYT forward 5'-CTGTTCCTGTACGGCATGGA-3' and reverse 5'-TTTCAGGTTCAGGGGGAGGT-3'; CAG$^{CreERT2}$ 5'GCTAACCATGTTCATGCCTTC-3'; 5'-AGGCAAATTTTGGTGTACGG-3'; ROSA26$^{tdTomato}$; Wild type, forward: 5'-AAGGGAGCTGCAGTGGAGTA-3' and reverse :5-'CTGTTCCTGTACGGCATGG-3', Mutant, forward :5-'CCGAAAATCTGTGGGAAGTC-3' and reverse :5'-GGCATTAAAGCAGCGTATCC-3'; Tie-2$^{Cre}$* forward *5'-CGCATAACCAGTGAAACAGCATTG-3' and reverse 5'-CCCTGTGCTCAGACAGAAATGAGA-3'; WT1$^{CreERT2}$ forward 5'-TGAAACAGGGGCAATGGTGCG-3' and reverse 5'-CGGAATAGAGTATGGGGGGCTCAG-3'.*

## Single-cell RNA sequencing

Endocardial cells of the atrio-ventricular canal derived from Tie2$^{Cre}$/Tomato positive embryos at E9.5 (18-22 somites) and mitral valve leaflets were collected at E16.5 and P1 pups, dissociated for 10 min with TrypLE$^{TM}$ express enzyme (Gibco; Catalog number 12604013) in the presence of Y-27632 Rock inhibitor (stem cell technologies) into single cells. After sorting the tomato positive cells with FACS, 10x genomics chromium platform was used to capture and barcode cells in order to generate single-cell Gel Beads-in emulsion (GEMs), following the manufacturer's protocol. Thus, using the reverse transcription master mix, cell suspensions of 2 samples of AVC, 1 sample of E16,5,and 2 samples of P0 were loaded onto the 10x Genomics Single Cell 3' Chip. Single cells were then partitioned into GEMs along with gel beads coated with oligonucleotides. Reverse transcription of cDNAs was followed by amplification, and a library was constructed using the Single cell 3' reagent Kit for all the samples, followed by sequencing of all libraries wit Next-seq Illumina sequencer. 10X Genomics pipeline Cell Ranger v1.2.0 was used for sample demultiplexing, barcode processing, and UMI counting. Libraries obtained from Illumina Sequencer were demultiplexed into reads in FASTQ format using the "cell ranger mkfastq" pipeline. In order to generate a gene-barcode matrix for each library, cellranger count pipeline was used. All the reads were aligned to the *mus musculus* reference genome mm10.

Cell ranger pipeline was used to normalize all the libraries to the same sequencing depth.

A first analysis of the data was performed with cell ranger and loop cell 10xGenomics software's. C loop clustered the cells using genes that are differentially expressed comparing to all other cells in other clusters with a threshold of Log2 > or equal to 2. A deep and secondary analysis was performed using the SCDE package and SCShinyhub[39]. Low UMI cells were excluded before clustering. We used: for E9.5 samples: sample: 1200 cells/ median gene per cell: 4617. For E16.5 sample: 452 cells/ median gene per cell: 700/; For P0 2 samples were used for a total of 854 cells/ median gene per cell 900.

## Clonality assessment and clonal grouping

Cell labeling followed tamoxifen administration at E8.5, and samples were collected at E16.5. At collection, labelled cells had expanded into clones that appeared to be fragmented into separate clusters of cells within a leaflet. To infer whether a group of cell clusters belonged to a single clone or were the result of merger of two or more clones, we followed a biostatistical approach first introduced by Lescroart *et al.*[40]. In this approach, one considers

the probability $p$ that a given cell was induced, as shown in extended data Fig.4. The probability $P(m)$, of observing $m$ induction events in a given leaflet can be approximated by a Poisson distribution: $P(m) \approx \frac{(pN)^m}{m!} e^{-pN}$ **(1)**, where $pN$ is the overall induction probability, and N is the number of cells expressing the promoter. If clone fragmentation events are considered to be random and statistically uncorrelated (i.e., Poissonian), the probability *R*(*k*) that a given clone will end up partitioned into *k* fragments is: $R(k) = \frac{f^{k-1}}{(k-1)!} e^{-f}$ **(2)**, where $f$ corresponds to the expected number of fragments of a clone derived from a single cell at the time of measurement (E16.5). As argued in Ref[40], the degree of fragmentation *f* could depend on the total size of the clone. To assess this, we measured the relation between total clone size and number of fragments in unicolor hearts (extended data Fig. 6). Although the number of measurements restricted to leaflets labelled with a single color was low, there was no clear dependence between these two variables. Note that, here, we considered clones in separate leaflets in the same heart to be independent.

Noting that the number of fragmentation events is given by the difference *k-m* ($k \geq m$) between the number of fragments *k* and the number of induced cells *m*, then the conditional probability $S(k \mid m)$ of finding a total of *k* labeled fragments of a single color at day E16.5, given that *m* cells of that same color were induced at day E8.5 is given by $S(k|m) = \frac{(mf)^{k-m})}{(k-m)!} e^{-mf}$ **(3)**. Combining equations **(1)** and **(3)**, we obtained the joint probability *J* (*k*|*m*) of finding a leaflet with initially *m* induced cells giving rise to *k* fragments of a given color is given by

$J(k|m) = S(k|m)P\ (m) = \frac{(mf)^{k-m}\ (pN)^m}{(k-m)!\ m!} e^{-mf-pN}$ **(4).**

Taking into account the ensemble leaflets in which there is at least one induced cell (*m* > 0), we defined the joint distribution of labeled hearts σ: $J_{ind}(m,k) = J(m,k)/(1-e^{-pN})$. Then the distribution for the number of fragments in labeled hearts is given by :

$F(k) = \sum_(m = 1)\hat{}k\, J_ind\ (m,\ k) = e\hat{}(-pN)/(1 - e\hat{}(-pN)\ )\sum\ (\ [\![(mf)]\!]\ \hat{}(k-m)\ \ (pN)\hat{}m)/(k - m)!\ m!\ e\hat{}(-mf)$ **(5).**

To assess the clonality of our lineage tracing data, we used the experimental quantification of the number of fragments *k* and Eq. **(5)**. We first estimated the overall induction frequency *pN* and average fragmentation rate of clone *f* by combining the data from both the mural and septal leaflets. Given the spatial separation between the mural and septal leaflets, we considered the cell clusters from separate leaflets to be independent from each other. Thus,

if in a given sample we observed e.g., 4 red fragments in the septal and 3 red fragments in the mural leaflet, we considered these to be two clonal induction events, one with 4 and the other with 3 fragments. We analyzed the number of fragments in both leaflets of the mitral valve of 40 hearts, from which we obtained fragments for 171 clone candidates. We analyzed the distribution of the number of fragments (extended data Fig. 7) and fitted the theoretical curve (**5**), from which we obtained an overall induction frequency $pN$ = 0.5 (0.3, 95% C.I.) (**2**), and a fragmentation rate $f$ = 1.4 (0.3) (**3**). This implied that clones with 4 or less fragments were likely to be of clonal origin, when the data for both the mural and septal leaflets was pooled together.

Even though the induction frequency $pN=0.5$ can be assumed to be equal for both leaflets, this is not necessarily the case for the fragmentation rate $f$.

To account for possible differences in the fragmentation in the mural and septal leaflets, we analyzed the distribution of number of fragments $F(k)$ of both leaflets independently (Fig. 5) for fixed induction frequency $pN$ = 0.5, as obtained from the combined data (extended data Fig. 7). In the case of the septal leaflet, the threshold for clonality was found to be $k$=4, meaning that groups of four clusters or less were considered clonal, while this threshold was lowered to $k$=3 for the mural leaflet. The difference originates from the lower fragmentation $f$ in the case of the smaller mural leaflet.

## References


1. Hoffman, J.I. & Kaplan, S. The incidence of congenital heart disease. *J Am Coll Cardiol* **39**, 1890-1900 (2002).
2. Kyndt, F. *et al.* Mutations in the gene encoding filamin A as a cause for familial cardiac valvular dystrophy. *Circulation* **115**, 40-49 (2007).
3. Dietz, H.C. *et al.* Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* **352**, 337-339 (1991).
4. Neptune, E.R. *et al.* Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat Genet* **33**, 407-411 (2003).
5. Durst, R. *et al.* Mutations in DCHS1 cause mitral valve prolapse. *Nature* **525**, 109-113 (2015).
6. Toomer, K.A. *et al.* Primary cilia defects causing mitral valve prolapse. *Sci Transl Med* **11** (2019).
7. Luna-Zurita, L. *et al.* Integration of a Notch-dependent mesenchymal gene program and Bmp2-driven cell invasiveness regulates murine cardiac valve formation. *J Clin Invest* **120**, 3493-3507 (2010).

8. Zhou, B. *et al.* Characterization of Nfatc1 regulation identifies an enhancer required for gene expression that is specific to pro-valve endocardial cells in the developing heart. *Development* **132**, 1137-1146 (2005).
9. Combs, M.D. & Yutzey, K.E. Heart valve development: regulatory networks in development and disease. *Circ Res* **105**, 408-421 (2009).
10. Pastushenko, I. & Blanpain, C. EMT Transition States during Tumor Progression and Metastasis. *Trends Cell Biol* **29**, 212-226 (2019).
11. Laurent, F. *et al.* HAND2 Target Gene Regulatory Networks Control Atrioventricular Canal and Cardiac Valve Development. *Cell Rep* **19**, 1602-1613 (2017).
12. Chen, Y.H., Ishii, M., Sucov, H.M. & Maxson, R.E., Jr. Msx1 and Msx2 are required for endothelial-mesenchymal transformation of the atrioventricular cushions and patterning of the atrioventricular myocardium. *BMC Dev Biol* **8**, 75 (2008).
13. Wessels, A. *et al.* Epicardially derived fibroblasts preferentially contribute to the parietal leaflets of the atrioventricular valves in the murine heart. *Dev Biol* **366**, 111-124 (2012).
14. Tran, T.N. & Bader, G.D. Tempora: Cell trajectory inference using time-series single-cell RNA sequencing data. *PLoS computational biology* **16**, e1008205 (2020).
15. Moore, K. *et al.* PDGFRalpha: Expression and Function during Mitral Valve Morphogenesis. *J Cardiovasc Dev Dis* **8** (2021).
16. Mahou, P. *et al.* Multicolor two-photon tissue imaging by wavelength mixing. *Nat Methods* **9**, 815-818 (2012).
17. Wuidart, A. *et al.* Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol* **20**, 666-676 (2018).
18. Bailey, N. *The elements of stochastic processes*. (john wiley & sons. Inc., New York, 1964).
19. Redner, S. *Statistical theory of fragmentation. In Disorder and fracture*. (1990).
20. Chabab, S. *et al.* Uncovering the Number and Clonal Dynamics of Mesp1 Progenitors during Heart Morphogenesis. *Cell Rep* **14**, 1-10 (2016).
21. Zhang, H., Lui, K.O. & Zhou, B. Endocardial Cell Plasticity in Cardiac Development, Diseases and Regeneration. *Circ Res* **122**, 774-789 (2018).
22. Pastushenko, I. *et al.* Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463-468 (2018).
23. McFaline-Figueroa, J.L. *et al.* A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat Genet* **51**, 1389-1398 (2019).
24. Sanchez-Martinez, R. *et al.* A link between lipid metabolism and epithelial-mesenchymal transition provides a target for colon cancer therapy. *Oncotarget* **6**, 38719-38736 (2015).
25. Xiong, J. *et al.* A Metabolic Basis for Endothelial-to-Mesenchymal Transition. *Mol Cell* **69**, 689-698 e687 (2018).
26. Bhatia, B., Hsieh, M., Kenney, A.M. & Nahle, Z. Mitogenic Sonic hedgehog signaling drives E2F1-dependent lipogenesis in progenitor cells and medulloblastoma. *Oncogene* **30**, 410-422 (2011).
27. Seki, T. *et al.* Ablation of endothelial VEGFR1 improves metabolic dysfunction by inducing adipose tissue browning. *J Exp Med* **215**, 611-626 (2018).
28. Yang, Z. *et al.* The tight junction protein Claudin-5 limits endothelial cell motility. *J Cell Sci* **134** (2021).
29. Tabaries, S. & Siegel, P.M. The role of claudins in cancer metastasis. *Oncogene* **36**, 1176-1190 (2017).
30. Kramer, B.A., Sarabia Del Castillo, J. & Pelkmans, L. Multimodal perception links cellular state to decision-making in single cells. *Science* **377**, 642-648 (2022).
31. Shigeta, A. *et al.* Endocardially Derived Macrophages Are Essential for Valvular Remodeling. *Dev Cell* **48**, 617-630 e613 (2019).
32. Hulin, A. *et al.* Maturation of heart valve cell populations during postnatal remodeling. *Development* **146** (2019).
33. Kim, A.J., Xu, N. & Yutzey, K.E. Macrophage lineages in heart valve development and disease. *Cardiovasc Res* **117**, 663-673 (2021).

34. Farrar, E.J. *et al.* OCT4-mediated inflammation induces cell reprogramming at the origin of cardiac valve development and calcification. *Sci Adv* **7**, eabf7910 (2021).
35. Guo, C., Yang, W. & Lobe, C.G. A Cre recombinase transgene with mosaic, widespread tamoxifen-inducible action. *Genesis* **32**, 8-18 (2002).
36. Loulier, K. *et al.* Multiplex cell and lineage tracking with combinatorial labels. *Neuron* **81**, 505-520 (2014).
37. Abdeladim, L. *et al.* Multicolor multiscale brain imaging with chromatic multiphoton serial microscopy. *Nat Commun* **10**, 1662 (2019).
38. Mahou, P., Vermot, J., Beaurepaire, E. & Supatto, W. Multicolor two-photon light-sheet microscopy. *Nat Methods* **11**, 600-601 (2014).
39. Jagla, B. *et al.* SCHNAPPs - Single Cell sHiNy APPlication(s). *J Immunol Methods* **499**, 113176 (2021).
40. Lescroart, F. *et al.* Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nat Cell Biol* **16**, 829-840 (2014).

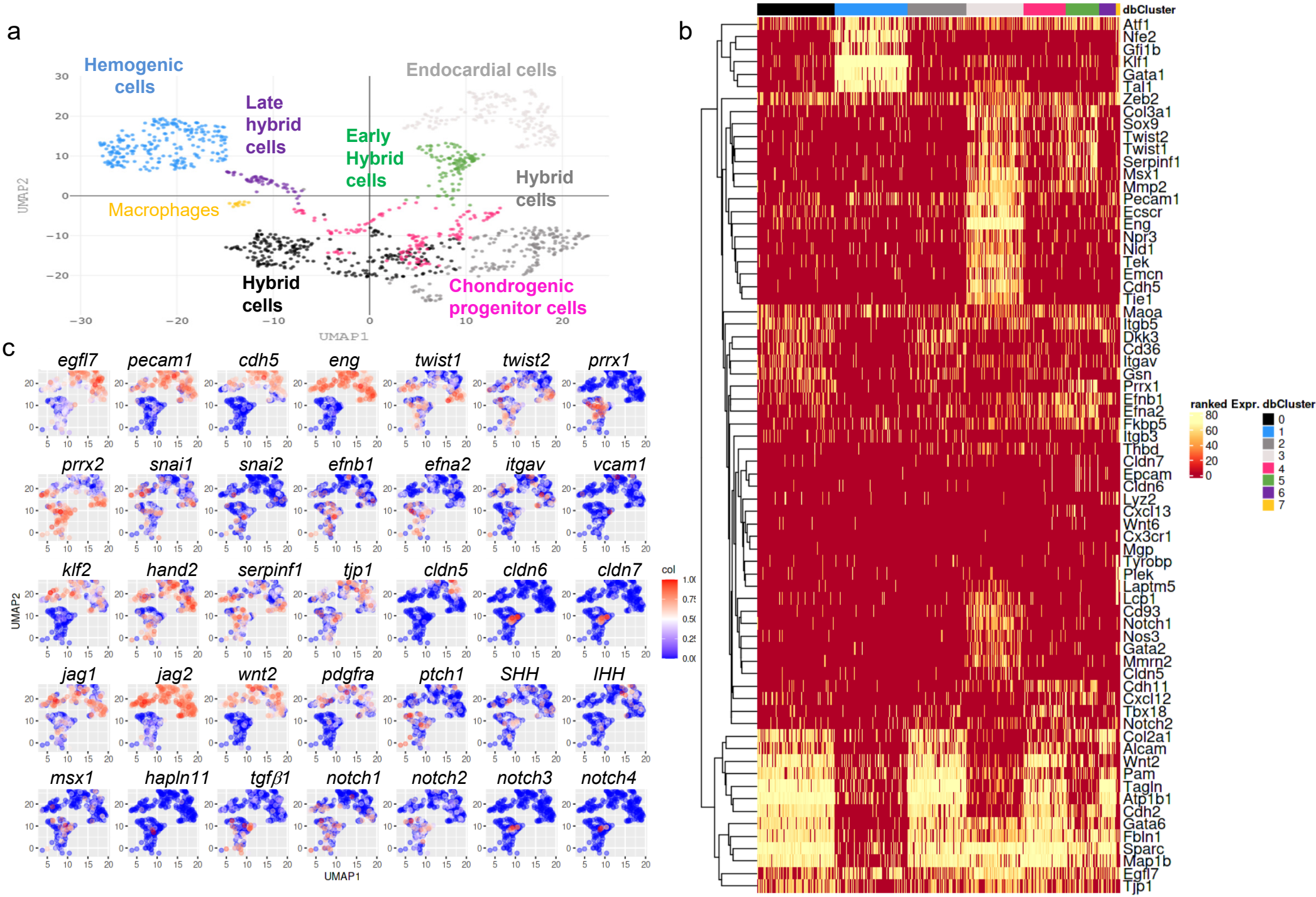



**Figure 1 : Single cell transcriptomic of E9.5 AVC cells.** AVC was dissected from E9.5 embryonic hearts from a *Tie2*$^{cre}$x*B6.Cg-Gt(ROSA)26Sor*$^{tm9(CAG-tdTomato)}$ mice breeding. Enzymatically dissociated cells were FACS-sorted and used for single cell-RNA seq. (a) UMAP showing the different AVC cell types (b) heatmap showing the most expressed genes (c) panel plots of gene expression in specific clusters 3 (endocardial cells) and 5 (early hybrid cells) .

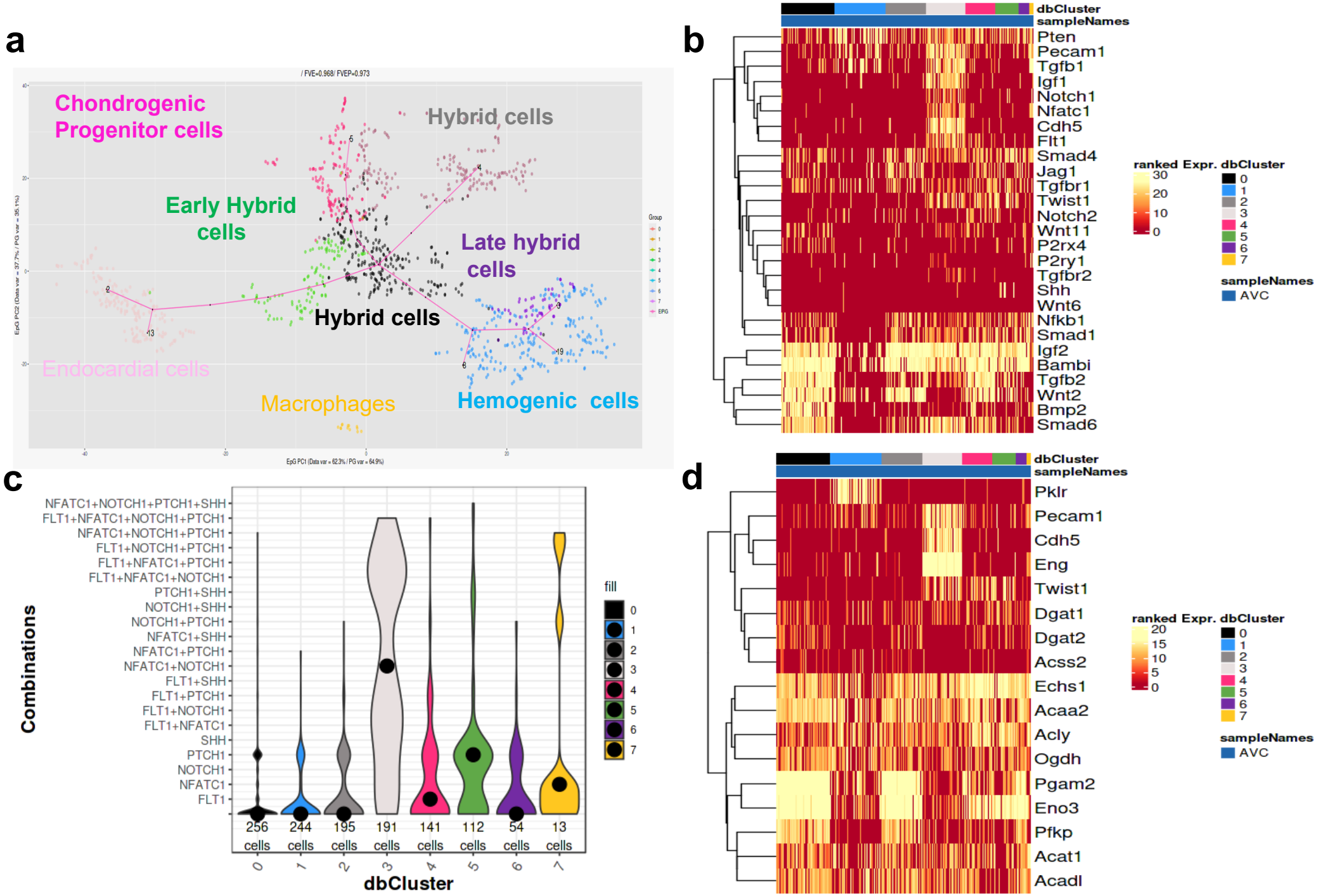



**Figure 2: Trajectory inference analysis and heatmaps of AVC cells.** (a) elpigraph inference trajectory. (b) Heatmap revealing expression of genes specifically involved in signaling pathways.(c) Violin plots illustrating the number of cell/cluster expressing genes specifically involved in signaling pathways (d) Heatmap revealing expression of genes specifically involved in metabolic pathways leading to fatty acid β-oxydation.

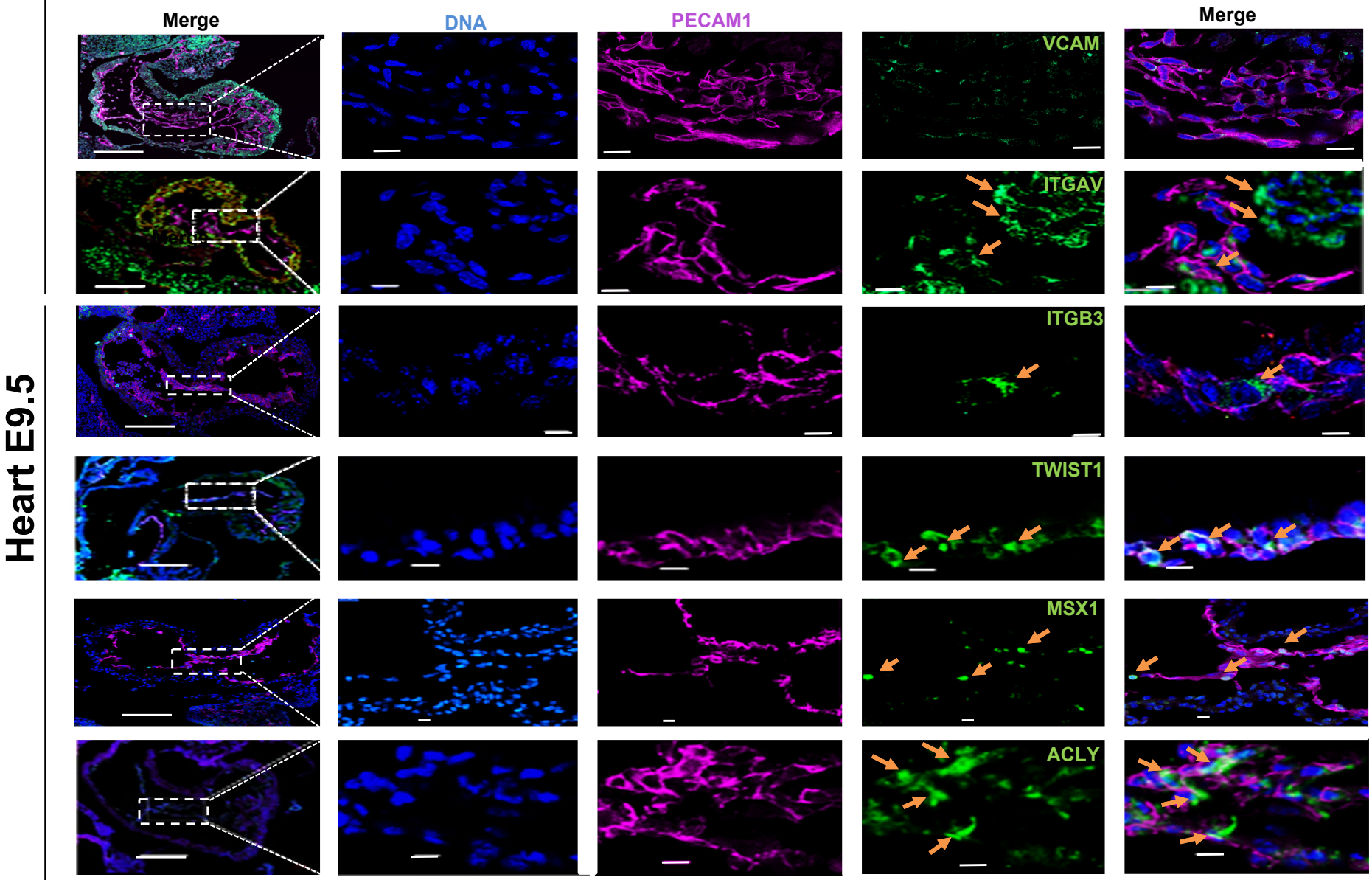



**Figure 3: Immunofluorescence of proteins expressed in AVC cells in E9.5 mouse heart.** The left panels show low magnifications of sections of whole heart. The square marks the AVC only imaged in the other panels. Positive cells for the marker of interest are pointed by orange arrows. The scale bars in the left column is 100 μm and the others are 10 μm

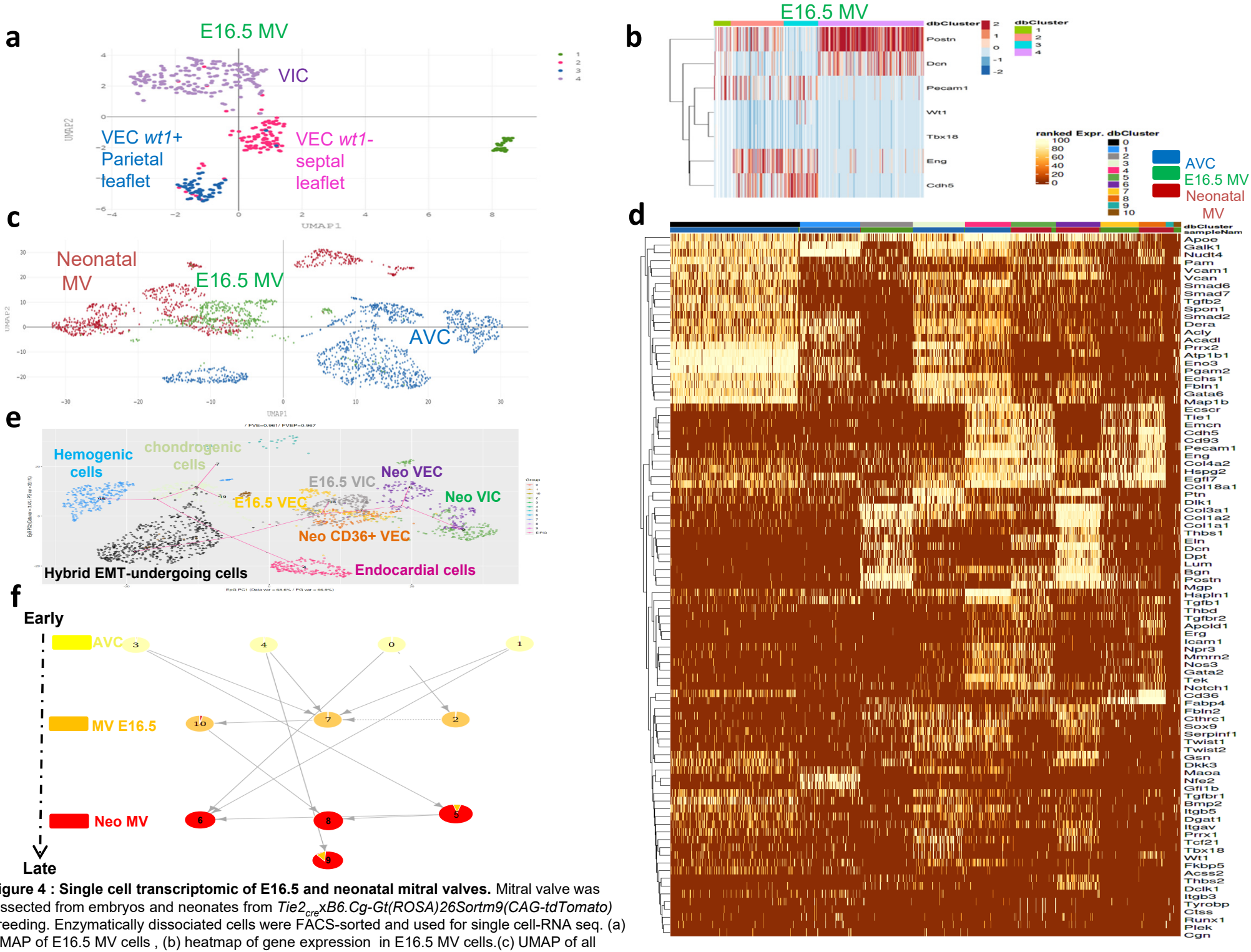
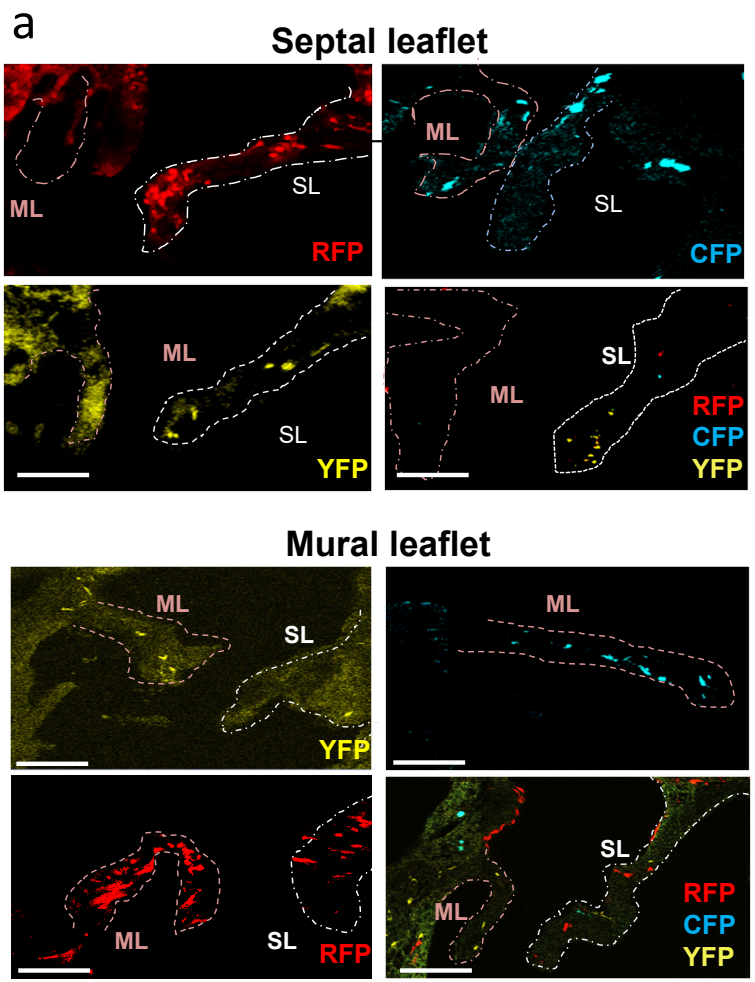



**Figure 4 : Single cell transcriptomic of E16.5 and neonatal mitral valves.** Mitral valve was dissected from embryos and neonates from *Tie2$_{cre}$xB6.Cg-Gt(ROSA)26Sortm9(CAG-tdTomato)* breeding. Enzymatically dissociated cells were FACS-sorted and used for single cell-RNA seq. (a) UMAP of E16.5 MV cells , (b) heatmap of gene expression in E16.5 MV cells.(c) UMAP of all samples (AVC, E16.5 and neonatal mitral valve cells). (d) heatmap of all samples (e) elpigraph inference trajectory including all datasets. (f) tempora inference trajectory of all samples.



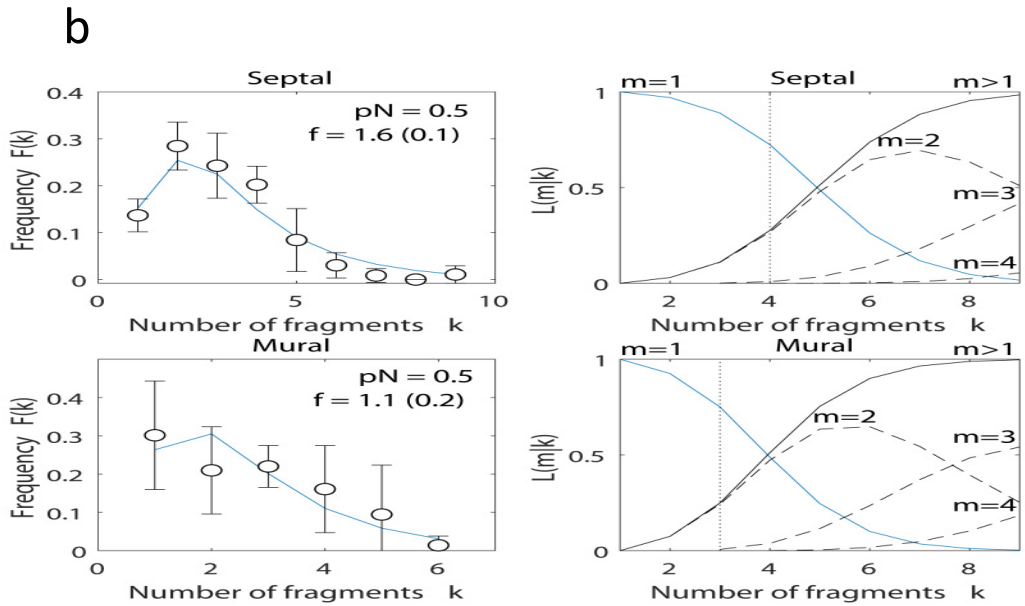




**Figure 5. cell clone distribution in embryonic mitral valve** (a) E16.5 mitral valve leaflets images with cell clones from CAG$^{Cre}$ x cytbow embryos ( scale bar 50 µm) (b) Probability F(k) of observing a total of k fragments of any given color, and (right) conditional probability L(m|k) that m cell were induced given that we observe k fragments in the anchoring point for the (top) septal and (bottom) mural leaflets. Markers correspond to experimental measurements, error bars correspond to the standard deviation from the measurement of F(k) for the three different colors. The solid line was obtained from fitting Eq. (5) in methods for pn = 0.50 and free parameter f.

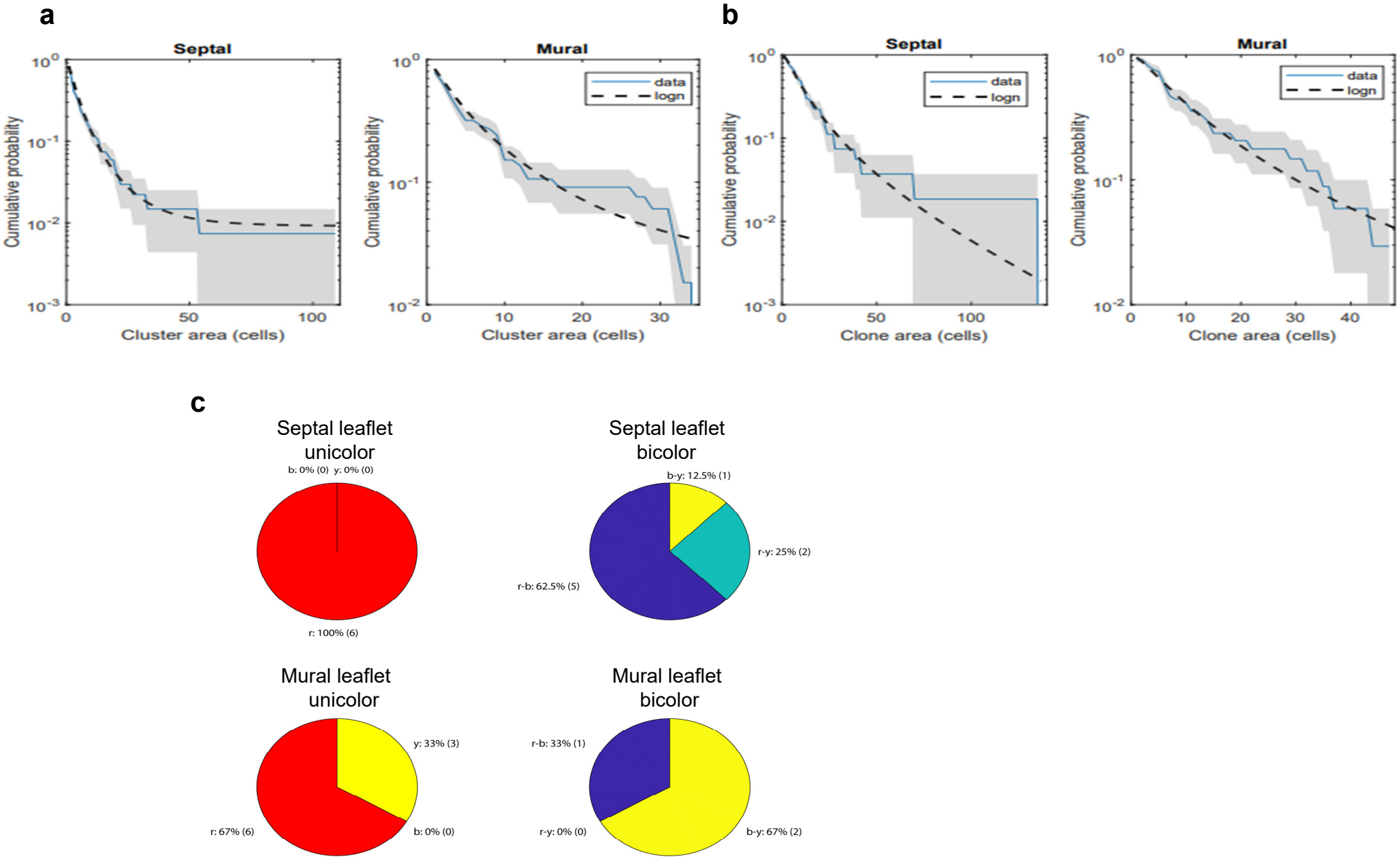
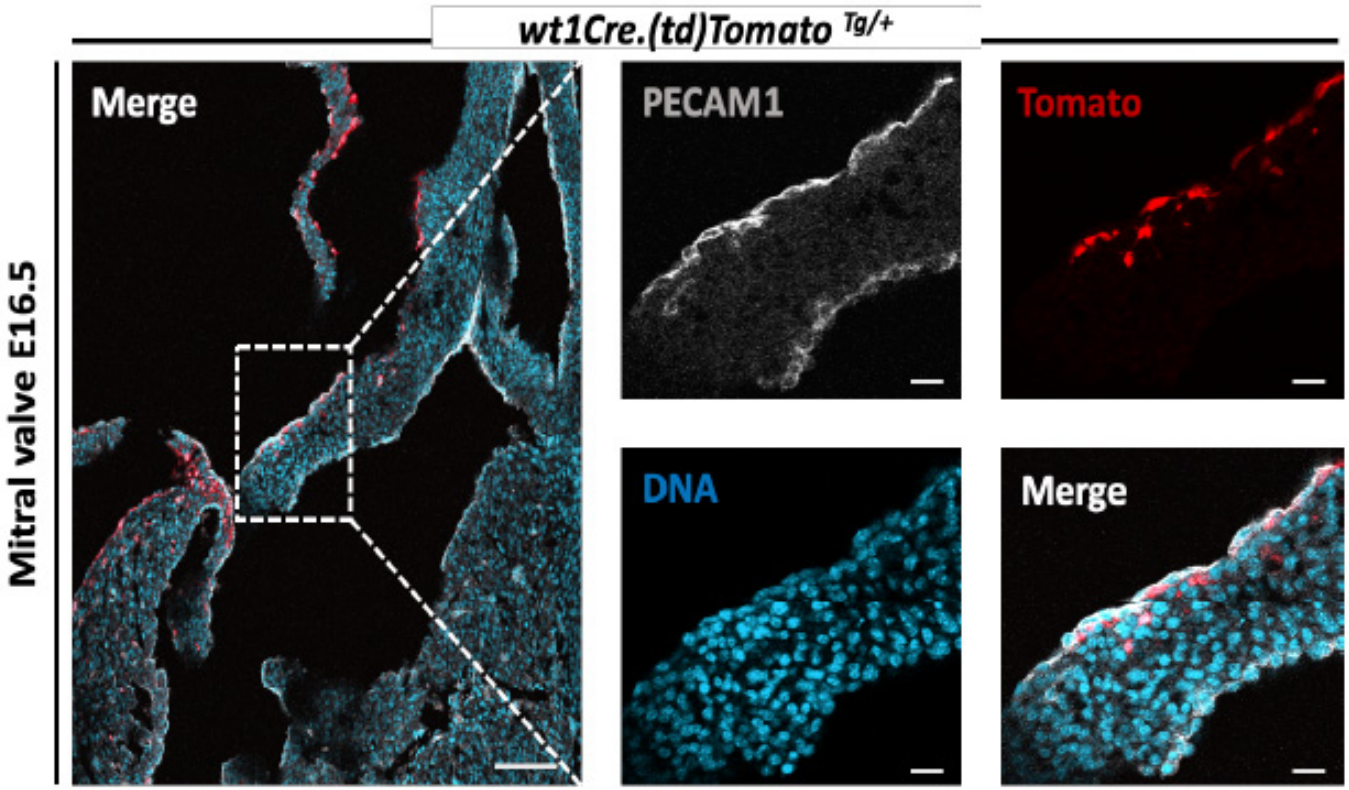
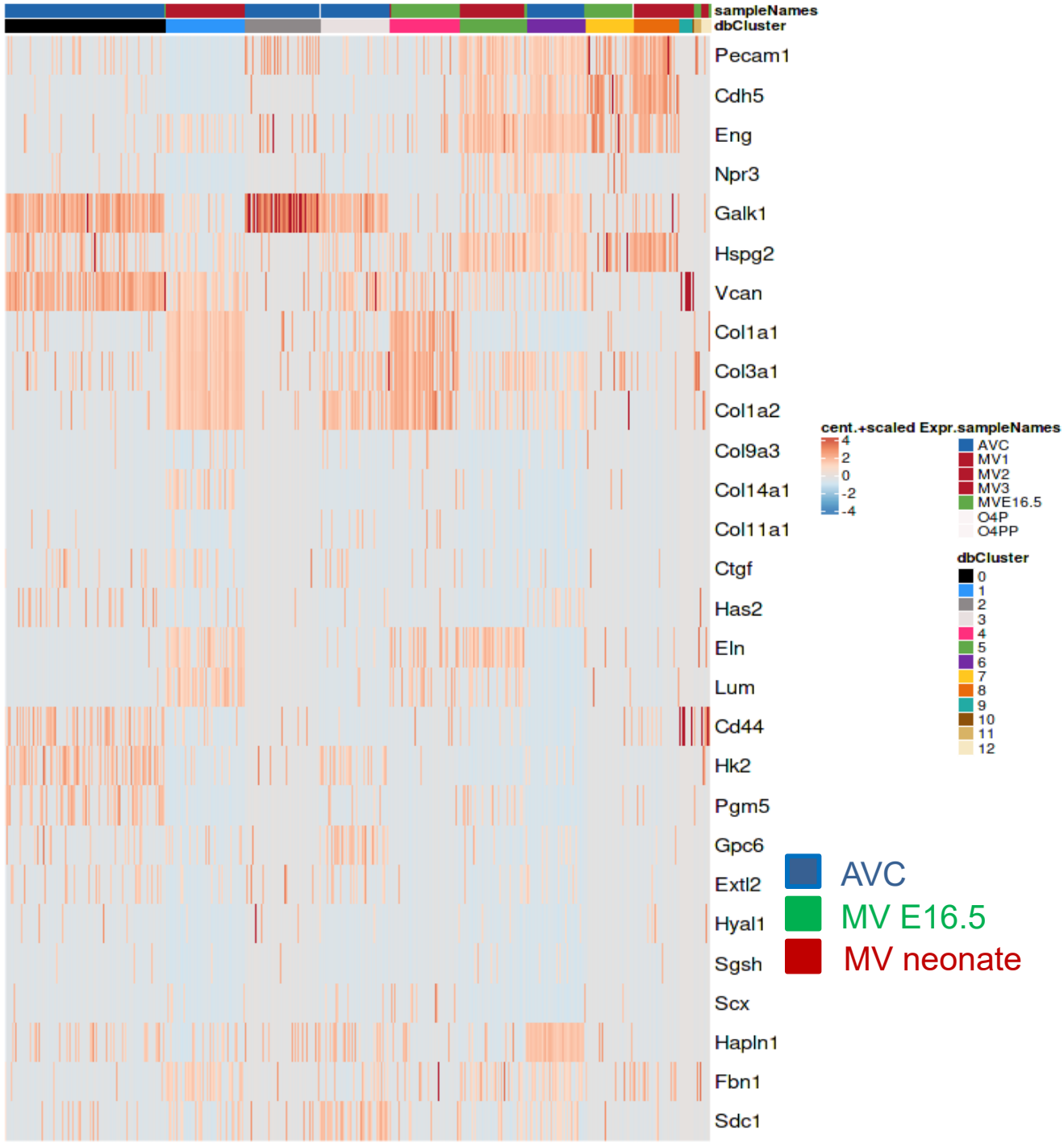


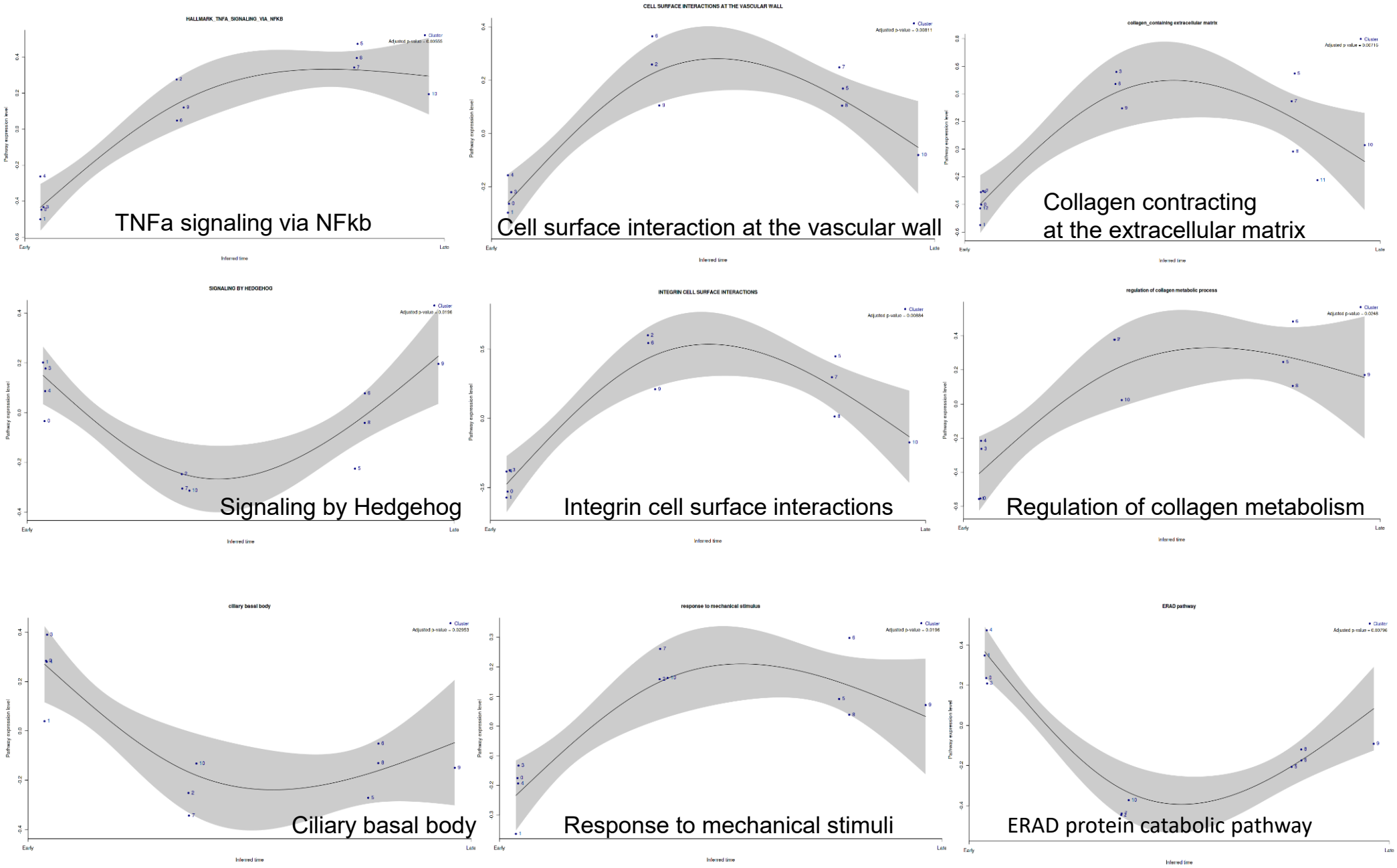

**Figure 6** : **Distribution of cell clusters (a) and cell clone (b) sizes in both septal and mural leaflets of E16.5 mitral valve** (logn:log normal distribution) (c) observed color expression in both septal and mural leaflets compared with ramdom combinations



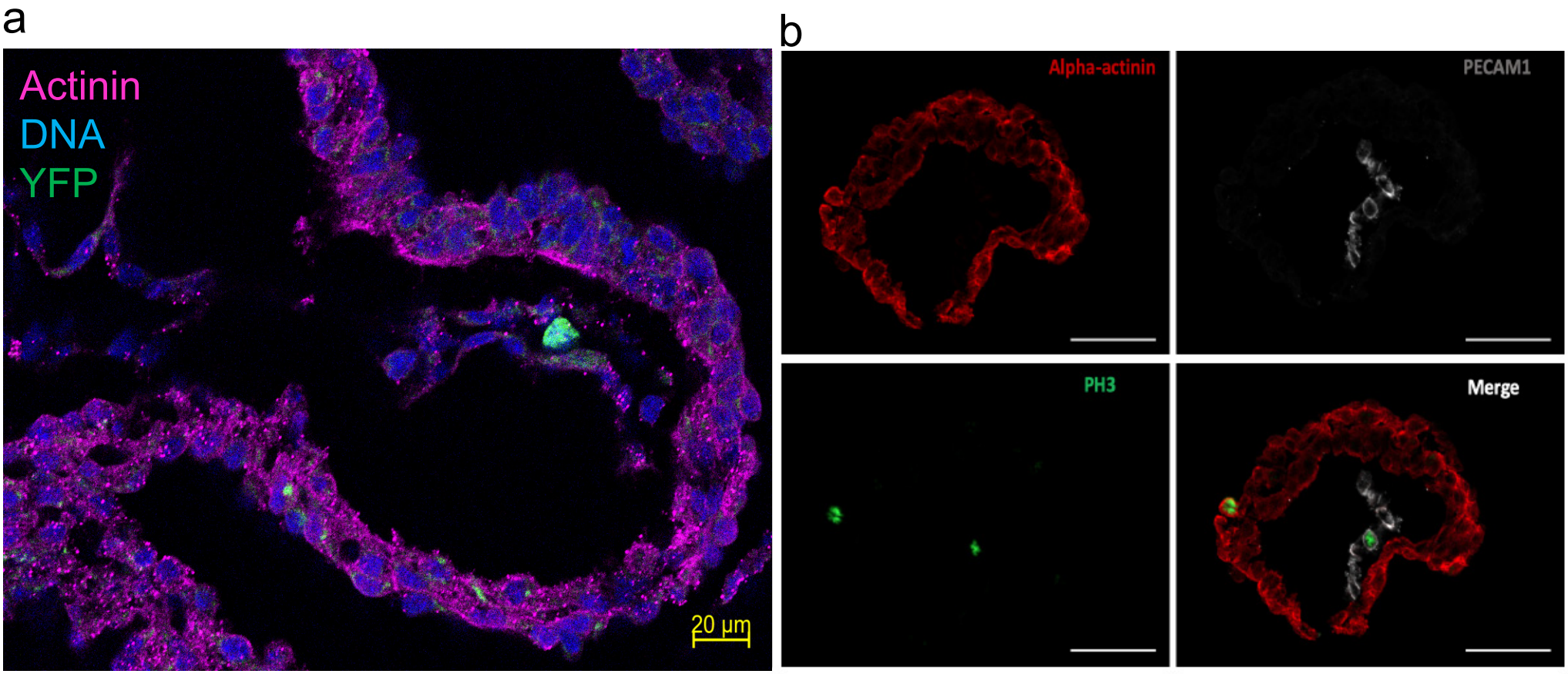

**Extended data Fig 1: Lineage tracing of epicardial cells.** Wt1$^{creert2}$ bred with Rosa26 $^{tdTomato}$ mice. The cre was induced by tamoxifen gavage to the mother at E8.5. The embryonic hearts were collected at E16.5 and the mitral valve imaged. The parietal leaflet is magnified in the right images




**Extended data Fig 2: Heatmap of cells revealing expression of genes involved in extracellular matrix** ( ECM) **production**

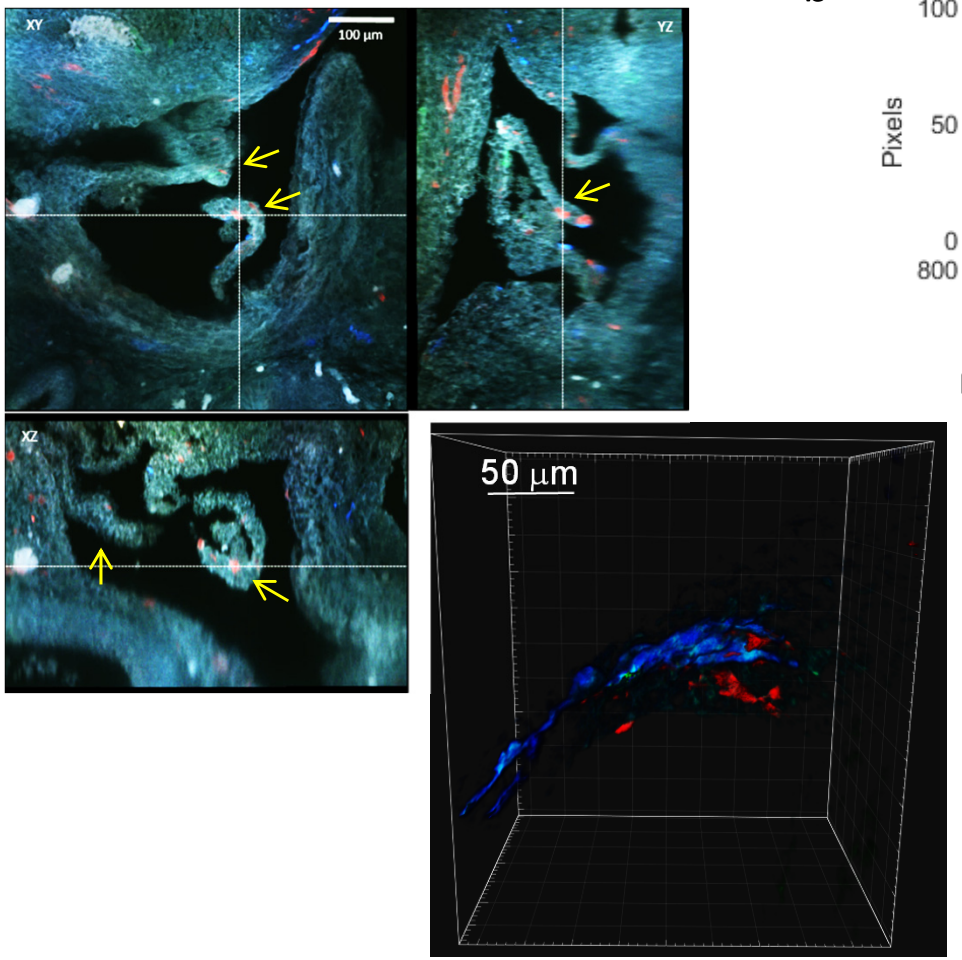



**Extended data Fig 3: Tempora trajectory inference significant pGo terms** (p value < 0.02)

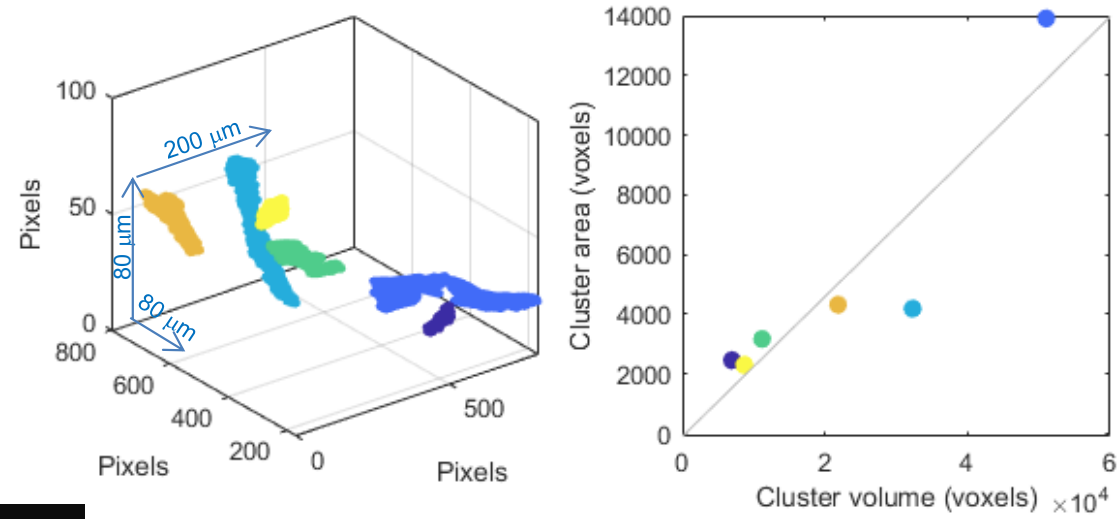



**Extended data Fig 4:** (a) **E9.5 embryonic heart from CAG$^{creERT2}$ x Rosa26$^{TCY}$ breeding mice revealing a single YFP+ cell in the AVC**.The female was given a low dose of tamoxifen (20 µg/g) at E8.5 of gestation. (b) E9.5 heart section stained with a anti-PH3 andtibody reveals a low cell division rate in the endocardium

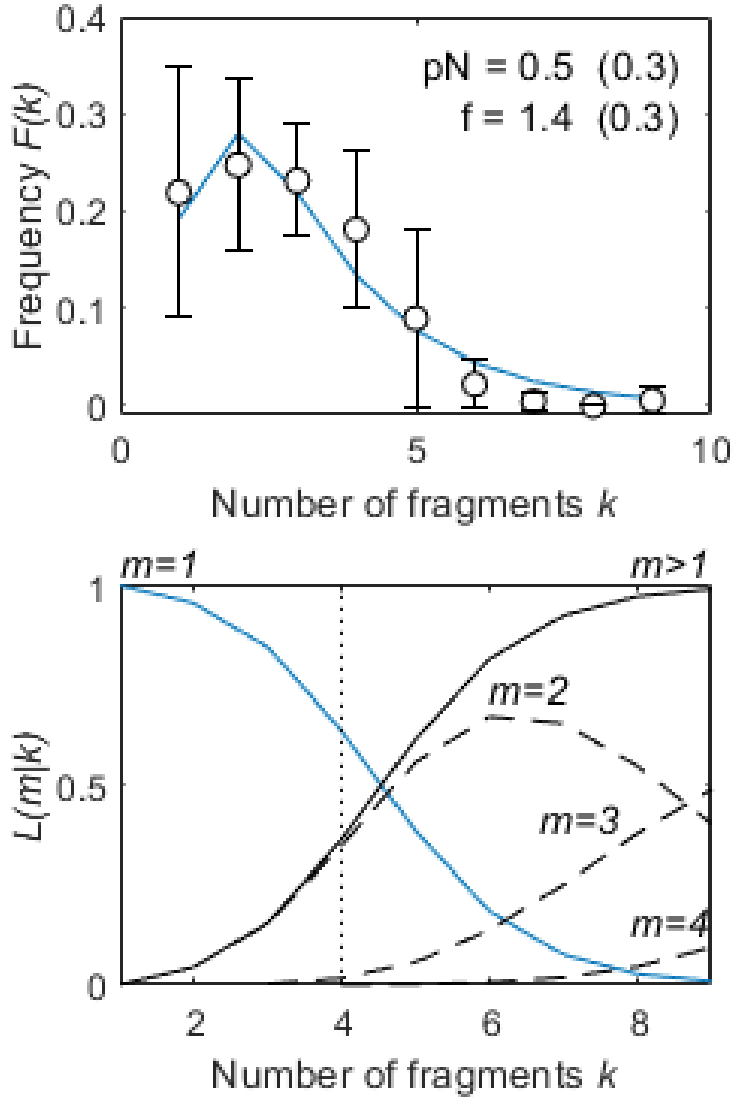
a



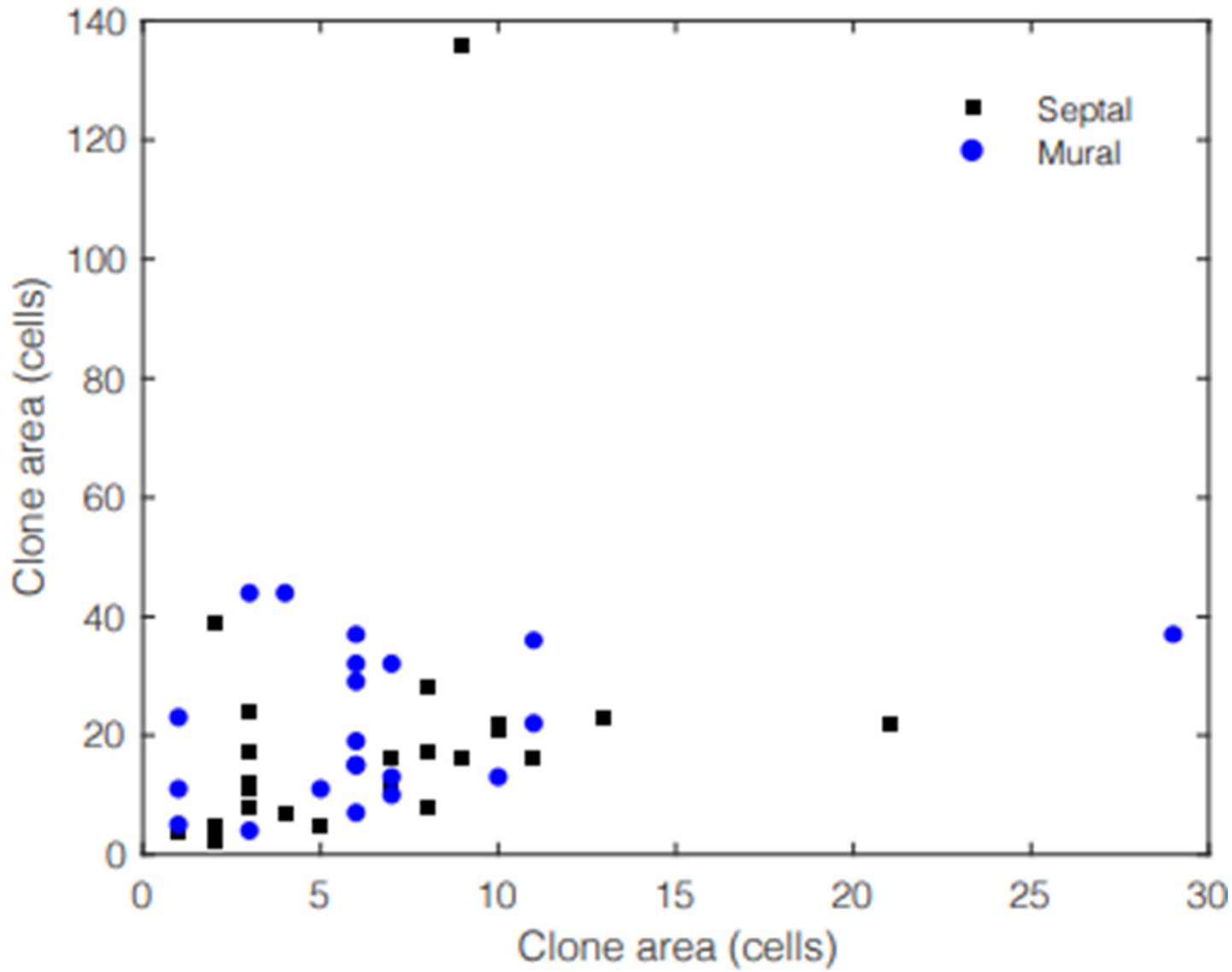

b



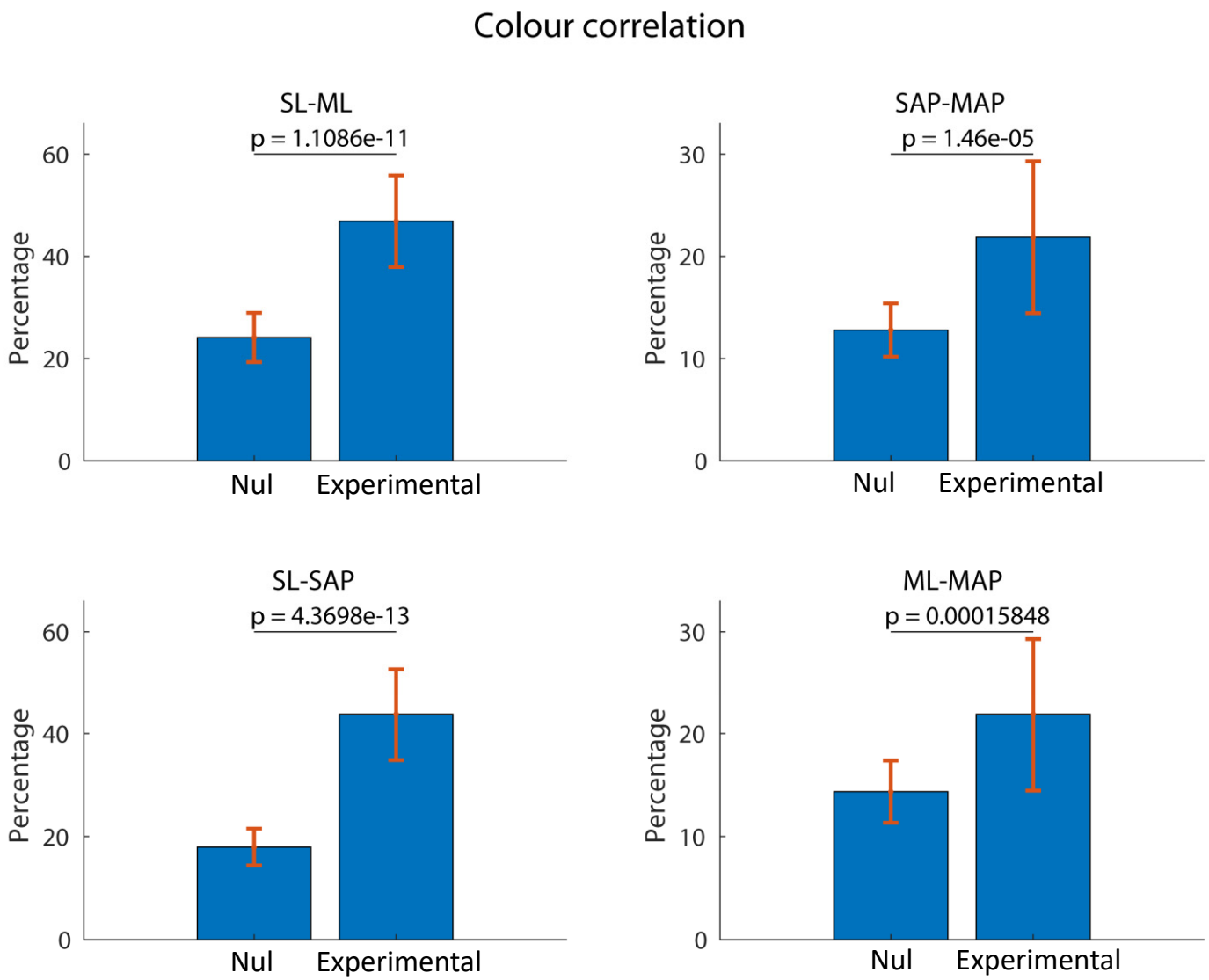

**Extended data Fig 5**: a: top panel **Mitral valve leaflets** (indicated by yellow arrows) within the orthoslices of 2-photons microscopy acquisition of E16.5 whole heart . Bottom panel : 3D reconstruction of two photon image of clones within a leaflet. b: correlation between the 2D-projected area of cell clusters and their 3d volume




**Extended data Fig 6: Overall induction frequency versus the clone fragmentation rate f**




**Extended Figure7**: **correlation of the sizes of clones of different colors in the same leaflet**




**Extended data Fig8: comparison of the fraction expected from random combinations (null model) with estimation (experimental) from the unicolor and bicolor statistics.**