Reduced specification of lateral plate mesoderm progenitors underpins heterochronic development of the emu wing

Axel H Newton [1,2*], Sarah M Williams [3,4], Belinda Phipson [5], Andrew T Major [2], Craig A Smith [2]

1. School of BioSciences, The University of Melbourne, Victoria, Australia
2. Biomedicine Discovery Institute, Monash University, Victoria, Australia
3. Monash Bioinformatics Platform, Monash University, Victoria, Australia
4. QCIF, University of Queensland, Queensland, Australia
5. Bioinformatics Division, Walter and Eliza Hall Institute, Victoria, Australia

* Corresponding author: axel.newton@unimelb.edu.au

ORCiDs: 0000-0001-7175-5978 (A. H. Newton); 0000-0002-0944-0622 (S. M. Williams); 0000-0002-1711-7454 (B. Phipson); 0000-0001-9036-6652 (A. T. Major); 0000-0002-1670-5545 (C. A Smith).

Abstract

Between tetrapods the limbs have undergone considerable remodelling to achieve unique adaptive behaviours. The forelimb is an ideal model for exploring the molecular basis of adaptive limb development, as it shows remarkable structural variation and is accessible for experimental manipulation. Of the most striking alterations to limb shape is the evolution of powered flight in birds. However, subsequently the flightless ratites (Paleognathae) have further evolved multiple instances of wing reductions, each utilizing distinct molecular mechanisms and displaying heterochrony with flighted birds (Neoaves). The emu has evolved a greatly reduced wing, consisted of a single digit. Thus, the emu is an excellent model to comparatively determine the cellular and molecular basis of wing heterochrony. We utilize comparative single cell transcriptomics of the developing forelimb field in the emu and chicken, to identify the source of the emus reduced wing. This was observed to occur via reduced specification and commitment of lateral plate mesoderm limb progenitor cells, which was accompanied by differential gene expression, persisting during limb initiation and outgrowth. These data suggest a progenitor allocation model, whereby altered limb morphologies may be achieved through altered commitment of precursor cells which act as an underlying template for pre- and post-patterning mechanisms.

Introduction

The tetrapod limb shows extraordinary diversity in form and function. Throughout the evolution of vertebrates, the paired appendages have evolved a staggering diversity to facilitate complex adaptive behaviours. Yet despite this morphological variation, the underlying structural homology of the limbs and shared gene regulatory networks which drive their development make them an ideal model for examining questions of evolutionary development biology (evo-devo). That is, how morphological diversity is regulated at the genetic and cellular level. The limbs form through a series of embryological events that being with specification of the lateral plate mesoderm (LPM). During early embryogenesis, the primitive LPM undergoes subdivision into the somatic and splanchnic LPM [1]. BMP signals from the overlying ectoderm specify limb fate within the somatic LPM [2], through activation of *TBX4/5* and subsequent *FGF10-FGF8* feedback loop, which initiates and stimulates outgrowth of the limb buds [3–6]. Patterning of the developing limb primordia occur through the interplay of several key transcription factors and morphogens (for extensive reviews see [7,8]), though largely via regulation of *Shh* which influences outgrowth and digit number [910]. Disruption of *Tbx4/5, Fgf10* or *Fgf8* during limb initiation results in failed limb formation [11–13], thus development of divergent limb morphologies typically involves alterations to pre-patterning and post-patterning mechanisms of the established limb primordia [14,15]. Studies across tetrapods have revealed that the spatiotemporal expression domains of *Shh* and/or *Fgf8* and their reciprocal signalling targets underlie the diversity of bat wings, mammalian digit reductions, dolphin fin expansions, jerboa hindlimb expansions and snake limb losses [16–22].

One of the most compelling modifications of the tetrapod limb was evolution of the avian wing to achieve powered flight. This transition was accompanied by a reduction of five to three digits which can be traced across archosaur evolution [23]. However, not all birds have retained this ability throughout evolution. The Paleognathae, or ratites, are a clade of flightless birds that include the ostrich, emu, rhea, cassowary, and kiwi. Palaeognaths and the Neoaves (flighted birds) last shared a common, volant ancestor ~110 million years ago [24,25]. Comparative genomics have revealed the molecular basis of this reduction is enriched within gene regulatory networks associated with limb development [26]. However, ratites show several independent instances of flight loss with variable wing structure, suggesting that the molecular basis of these reductions are driven through different developmental pathways [27]. The Australian emu (*Dromaius novaehollandiae*) shows the most dramatic reduction of forelimb morphology, characterized by a simplified, vestigial wing containing a single clawed digit [28,29]. Embryonic development of the emu closely resembles that of the chicken [30,31] with forelimb initiation occurring at roughly equivalent developmental stages [32]. However, heterochrony of the emu wing becomes rapidly apparent, as while the chicken forelimb bud undergoes steady outgrowth, the emu forelimb shows reductions immediately following limb initiation, and is stunted relative to the hindlimb [30].

The molecular basis of emu wing heterochrony is complex, where multiple mechanisms have been identified which associate with the delayed outgrowth and differentiation [15]. Compared with development of the chicken wing, the emu exhibits delayed differentiation of the somatopleure from the primitive LPM [32], a smaller forelimb forming domain and *TBX5*

expression domain [33], attenuated activation of *FGF10* in the early wing bud mesenchyme [32] , delayed expression of *SHH* and formation of the ZPA [34]; and co-option of the cardiac transcription fact *NKX2.5* into the limb myocytes which act to repress forelimb outgrowth and patterning [35]. However, a common theme amongst these disparate mechanisms is an initial delay in emu somatopleure differentiation and corresponding reduced forelimb domain. Together, these suggest that emu wing heterochrony may arise earlier during development than currently appreciated, prior to formation of the forelimb primordia.

To further explore the ontogenetic basis of wing heterochrony in the emu, we utilized single cell transcriptomics during early limb specification and initiation, drawing evolutionary comparisons with the chicken [2]. Through stage-matched sampling of cells during early LPM development, we resolve ambiguity in previous published mechanisms and determine the onset of the emus delayed limb growth to generate new insights into how heterochrony can occur at the molecular level. We identify a novel mechanism underlying differential limb development, occurring via reduced commitment of somatic LPM limb progenitors, preceding activation of *TBX5*, *FGF10* and the canonical limb development pathway. This was accompanied by reduced expression of known LPM genes, suggesting an overall reduction in the differentiation potential of the emu somatic LPM. These findings add new perspectives into the developmental basis of morphological evolution between vertebrates.

Results

In this study we utilized the 10x Chromium system to generate single-cell transcriptomic sampling of cells within the developing chicken and emu forelimb field. We sampled stage-matched embryonic forelimb field tissues covering Hamburger-Hamilton stages HH10, HH14 and HH18, corresponding to chicken embryonic day e1.5, e2.5 and e3.5 [2], and emu e3.5, e4.5 and e5.5 (Figure 1a). After quality filtering, we identified a total of 15355 chicken cells and 16748 genes, and 16149 emu cells with 15628 genes (Table S1). To remove low read count or low diversity cells, we applied an additional strict filtering for high mitochondrial content, and stringent count threshold of ≥2000 UMI per cell. This produced 3262 high quality cells in chicken, and 2678 cells in emu (summarized in Table S1) which were used in subsequent analyses. Cell transcriptomic relationships were visualized with global t-distributed Stochastic Neighbour Embedding (tSNE) dimension reduction, which produced similar cell type distributions within the chicken and emu presumptive forelimb field datasets, displaying clear separation of clusters comprising the embryonic ectoderm, mesoderm, endoderm and vasculature (Figure 1b-d). Independent unsupervised clustering of the chicken and emu HH10, HH14 and HH18 presumptive forelimb datasets revealed 13 transcriptionally distinct cell clusters (c) in the chicken (Figure 1c) as previously determined [2], and 16 clusters in the emu (Figure 1d). Tissue and cell type identities were assigned to each cluster based on their differential gene expression profiles (observed as differences in log-fold change between clusters; Figure 1e,f) and corresponding spatiotemporal expression patterns observed throughout the developing chicken embryo (*Gallus Expression In situ Hybridization Analysis*: GEISHA [36,37]). Identified cell types were largely in agreement between the two species datasets showing single clusters representing the endoderm, vasculature and RBCs,

two-three clusters representing the ectoderm, and several clusters for the mesoderm, including paraxial, intermediate, and lateral plate derivatives (Figure 1c-d).

The emu exhibits reduced commitment of the somatic LPM

To identify differences underlying emu wing bud heterochrony, the chicken and emu datasets were integrated into a single merged atlas. A common set of genes with 1:1 orthology between both species were identified (n=10415) and used to subset the datasets to only contain common gene IDs. The datasets were then merged based on gene ID and visualized through UMAP dimension reduction, revealing minor batch effects between species. These were removed through mutual nearest neighbours (MNN) correction [38], producing a combined, homogeneous chicken-emu dataset with equivalent clusters corresponding to embryonic ectoderm, mesoderm, endoderm and vasculature (Figure 2a). Given that the limbs form from the embryonic mesoderm, the merged chicken-emu dataset was further subset into cells of mesodermal origin (Figure 2b). Additional unsupervised clustering of the merged dataset revealed 9 sub-clusters, corresponding to all know tissue types within the limb field mesoderm. Namely, cell types comprised early (undifferentiated) mesoderm, paraxial and intermediate mesoderm, undifferentiated lateral plate mesoderm (LPM), somatic and splanchnic LPM, visceral (gut) precursors, limb bud mesenchyme and extra-embryonic / flank LPM (Figure 2b; Figure S1). Of interest, NKX2-5 was not detected in chicken cells, but was found to be activated in emu LPM/somatic LPM prior to formation of the limb bud mesenchyme (Figure S1). Cell differentiation trajectories were determined in the merged dataset with monocle3 [39], supporting the subdivision of the undifferentiated LPM into the splanchnic/gut and somatic/limb lineages, as examined previously [2].

A recent study investigating the molecular basis of emu limb heterochrony reported that the emu shows delayed differentiation of the somatic LPM compared with the chicken [32], however, this observation was not verified. Building on this idea, we investigated differences in cell type proportions between the chicken and emu during LPM subdivision and differentiation of the somatic LPM into the limb mesenchyme. Theoretically, as the LPM undergoes dorsoventral subdivision into the somatic and splanchnic layers, these tissues should arise in a roughly 1:1 ratio to line the embryonic coelom [1]. Shortly thereafter, the somatic LPM begins to proliferate to form the limb bud mesenchyme. Initial examination into proportions of LPM, somatic and splanchnic LPM cells between the chicken and emu revealed that the emu sampling failed to recover early HH10 cell types present in the chicken. Despite attempted stage-matched sampling of chicken and emu tissues based on developmental series' [30,31], the emu embryonic day (E)3.5 (~HH10) cells more closely resembled those of chicken HH14 cell types, albeit with a few early mesodermal progenitors (Figure 2d). By e5.5 (~HH18) however, emu cell types matched those of HH18 chicken samples (Figure 2c-d), suggesting our three emu samples were only equivalent to chicken stages HH14 and HH18. However, as our emu sampling sufficiently captured the undifferentiated LPM, prior to its differentiation into somatic / splanchnic LPM and limb and visceral cell fates, respectively, we were able to examine proportions during LPM differentiation.

Analysis of cell type proportions revealed an overall reduction in the number of emu somatic LPM and limb bud mesenchyme cells, relative to the development of the splanchnic LPM and other tissue types and those observed in the chicken (Figure 2c-d). Prior to subdivision, the chicken and emu had equivalent 1:1 proportions of LPM cells across their three individual stages (Figure 2e). However after LPM subdivision, the emu displayed a marked decrease in proportions of somatic LPM and limb bud mesenchymal cells – relative to the chicken; whilst the extraembryonic MES and splanchnic-derived visceral precursors were increased – relative to the chicken (Figure 2d-e). To validated these observations, we performed immunofluorescent labelling of the somatic LPM marker gene *TWIST1* [2] in chicken and emu forelimb field tissue sections. This confirmed that the emu possesses less TWIST1^{+} somatic LPM cells compared to chicken at equivalent stages, relative to a similar abundance of splanchnic LPM cells, and TWIST1+ cells in the somites (Figure 2f). Together, these genomic and functional results confirm the emu has evolved reduced commitment of the undifferentiated LPM towards a somatic LPM fate at the early stages of limb specification in emu, despite the typical development of surrounding tissues such as the somites, splanchnic LPM and viscera.

Gene expression dynamics underlying wing heterochrony

To understand the molecular basis of the reduced specification of the emu somatic LPM, we investigated differentially regulated genes in the emu LPM, somatic LPM and forelimb bud mesenchyme. This was first analysed through pseudobulk differential expression (DE) analysis of emu cell clusters with the corresponding clusters in the chicken. This revealed differentially expressed emu genes in LPM tissues with known roles in various aspects of LPM development (Figure 3a). Within the undifferentiated LPM clusters, we identified emu-specific downregulation of the early LPM and limb fate marker genes *PRRX1* [40,41] and *OFLML3* [2], increased expression of splanchnopleure marker *OSR2*, and increased expression of the RA-metabolizing enzyme *CYP26A1*. DE analysis of the somatic LPM clusters produced few statistically significant gene expression profiles, likely owing to low numbers of contributing cells in the emu dataset. However, in limb bud mesenchyme several differentially expressed genes were noted between chicken and emu, including emu-specific downregulation of limb-developmental genes *PRRX2, HOXD9, FGFR1* as well as *FGF10* which has previously been identified as a contributor to reduced emu limb outgrowth [32]. This confirms that while several limb bud genes are downregulated in the developing emu limb, LPM genes are downregulated in the emu LPM prior to somatic LPM differentiation and limb bud formation.

To complement our static pseudobulk approach, we further investigated dynamic changes in gene expression during LPM and limb bud development across pseudotime. The root node was set in the undifferentiated mesoderm / LPM boundary, where emu undifferentiated LPM cells were first detected. Cells were ordered in pseudotime along the LPM to limb mesenchyme lineage trajectory (Figure 2b,d). We then examined genes which dynamically changed across pseudotime, and that were differentially expressed between the chicken and emu. Excluding ribosomal genes, this analysis revealed 1199 significantly downregulated,

and 1566 significantly upregulated genes in the emu compared with the chicken. Functional profiling of emu downregulated genes using g:GOSt showed enrichment of gene ontology terms associated with embryonic limb morphogenesis (Figure S3). Of the top downregulated genes, these included known markers of LPM and limb bud development *PRRX1, PRRX2, TWIST1, WNT5A, SALL1, HOXD9, GLI3* and *FGF10*; and RA synthesis/response genes: *ALDH1A2, STRA6, RDH10, RARB* and *CYP26B*1 (Table S4). These gene lists complemented the pseudobulk results, but also revealed additional genes that were not captured by the static analysis.

Visualization of selected marker gene expression across pseudotime confirmed dynamic chicken and emu expression differences during LPM and limb development. Notably, we observed delayed onset and maintained downregulation of somatic LPM markers *PRRX1, TWIST1* and *ALDH1A2* during emu LPM development (Figure 3b), as well as decreased activation and expression of limb initiation markers *TBX5* and *FGF10*, confirming previous observations [32,33]. Whole mount *in situ* hybridisation of the known LPM fate marker *PRRX1* [40,42] confirmed reduced expression in the developing emu forelimb field, appearing in less abundance and smaller domain to that of the chicken (Figure 3c), supporting the hypothesis of a reduced pool of emu somatic LPM progenitors (Figure 2f). Interestingly, the field of *PRRX1* expression appeared largely increased in the developing emu hindlimb field and showing further heterochronic activation compared to the chicken. Together, these data demonstrate that emu wing heterochrony begins at the initial stages of LPM specification through a reduced allocation of limb progenitors, via reduced activation of key LPM genes prior to the activation of the *TBX5* and *FGF10* dependant limb initiation pathway [32,33]. This presents an intriguing evolutionary mechanism underlying adaptive morphological evolution, whereby reduced allocation of progenitor cells limits the development of anatomical structures.

Discussion

Of the flightless ratites, the emu has evolved the most striking wing reduction, both in size and number of digits [24,29]. Several studies have examined the molecular basis of this reduction, each finding unique mechanisms, including altered regulation of *TBX5*, *FGF10* and *SHH*, which contribute to their overall reduced size and shape [15,28,32–35]. However, we find evidence to suggest that earlier mechanisms exist, acting upon the LPM progenitors which underlie development of the presumptive forelimb field. Comparative single cell transcriptomics of the developing forelimb field in emu and chicken embryos [2] revealed differences in cell type proportions and allocation of limb progenitors during limb specification. Emu wing heterochrony was found to be genetically pre-determined at the beginning of LPM specification, driving reduced allocation of somatic LPM progenitors. These findings have novel implications for limb evo-devo, demonstrating that differences in adaptive morphological evolution may not necessarily occur through pre- or post-patterning process. Rather, differences in limb size, shape or proportion may instead occur through early alterations to the allocation of LPM progenitors.

Intriguingly, this progenitor allocation hypothesis may be additionally relevant during development of the emu hindlimb. Ratites possess large and powerful hindlimbs in response to their adaptation to a terrestrial lifestyle. While we did not sample hindlimb cells, the increased spatial and temporal activation of the somatic LPM marker gene *PRRX1* in the developing emu hindlimb, relative to that of the chicken, suggests a heterochronic increase in allocation of hindlimb LPM progenitors to further facilitate their rapid hindlimb growth. This hypothesis may be also be relevant to another classic model of mammalian limb evo-devo, the jerboa, which has significantly expanded hindlimbs [43,44] for their terrestrial hopping.

Our data illustrate that while the molecular regulation of adaptive morphological evolution is complex, it may stem from one or several independent sources. In the instance of emu wing heterochrony, this can be traced back to reduced progenitor commitment and attenuation of the limb initiation program. Deeper analysis into the mechanisms which influence progenitor fate commitment will illustrate how gene regulatory networks can evolve to influence developmental outcomes. Through these, we will gain a greater appreciation of the mechanisms which underlie heterochrony and adaptive morphological evolution.

Methods

Egg incubation, tissue collection, single cell sampling

Fertilized chicken and emu eggs were sourced from Research Poultry Farm, Victoria and Emu Logic, NSW respectively. Chicken sampling has been described previously [2], but was performed concurrently with emu egg collection. Eggs were incubated at 37.5°C using staggered timepoints so that different developmental stages were reached on the same day of collection. Chicken eggs were collected at embryonic day (e) 1.5 (stage 10), e2.5 (HH14) or e3.5 (HH18), while emu eggs were collected at e3.5 (HH10) e4.5 (HH14) and e5.5 (HH18). Embryos were dissected away from extra-embryonic membranes, rinsed in ice-cold DPBS then the LPM dissected. LPM tissues were digested with 0.05% Trypsin / EDTA and incubated at 37°C for 15 minutes, with mechanical dissociation every 5 minutes until no clumps were visible. Enzymatic activity was stopped with addition of 10% FCS. The dissociated cells were spun at 400g for 5 minutes, then resuspended in 1x EDTA / Propidium Iodide in DMEM (Gibco). Cells were filtered through a 70um Flowmi Cell Strainer (Scienceware), and viable cells were isolated through flow cytometry (Flowcore, Monash University).

Chicken (n=3) and emu (n=3) samples were produced on different days and submitted to Micromon Genomics (Monash University) for independent processing through the 10X Genomics Chromium controller, using Single Cell 3' V3.1 chemistry, as per the manufacturer's instructions. Samples were subjected to 10 cycles of PCR for cDNA amplification and 16 cycles for library amplification. Once chicken and emu libraries had been constructed, chick they were sequenced using on the MGITech MGISEQ2000RS platform with MGIEasy V3 chemistry, on two lanes with 100bp paired end reads.

Quantification and statistical analysis, bioinformatics

Chicken data analysis has been described previously [2], but was performed concurrently with emu analysis. Chicken reads were aligned to GRCg6a, and emu to droNov1 reference genomes using CellRanger (v4.0.0, using option: –force-cells 15000). Due to the number of reads observed just downstream of annotated genes, the gene annotation (from Ensembl release 100, gene biotypes: protein coding, lincRNA and antisense) was edited to include 1000bp downstream of each gene. Single cell analysis was performed in R using packages scran [45], scater [46] for QC, monocle3 [39] for trajectory analysis, and iSEE for interactive viewing [47]. Gene names were used for analysis and, where they mapped to multiple Ensembl ids, the Ensembl ID with the highest number of counts was kept. Cells with low total UMI counts (<2000) were excluded. Cell cycle was annotated with cyclone in the scran package [45] using the mouse reference from [48] mapped to its one-to-one chicken orthologs.

For each dataset, the top 1000 genes with the highest biological variance were identified with modelGeneVar function of scran [45], blocked on the sequencing sample, and excluded mitochondrial genes or genes on the Z or W chromosomes to minimise sex effects. PCA was calculated on these, and the first 15 PCs used to generate a global chicken and emu tSNE layout. Clusters were defined with the walktrap method, on a SNN graph (k=10) [45], and

cluster identities were determined from gene logFC changes and spatial expression profiles in the Gallus Expression *In Situ* Hybridization Analysis (GEISHA) database [36,37]. Each dataset showed similar cells cluster labels, albeit with minor differences.

Dataset integration and batch correction

To integrate the full chicken [2] and emu datasets, these were filtered to contain shared 1:1 gene IDs. Chicken and emu gene ID were intersected to produce a common 'universe' of shared features (n = 10415), and each dataset was subset to only contain these gene IDs and synchronized to contain the same metadata columns. Each dataset was rescaled to adjust for sequencing depth using batchelor , and integrated with Mutual Nearest Neighbours (MNN) correction [38] with k=20 neighbours and d=50 dimensions. This produced a corrected matrix of cell co-ordinates which was used as input for PCA, UMAP and tSNE dimension reduction, which showed good mixing of chicken and emu samples, recapitulating all of the known tissue types within the early forelimb field [2]. Raw and log count assays and cell metadata was re-added into the corrected matrix, yielding a final merged and batch corrected dataset.

Mesoderm analysis and lineage reconstruction

Mesodermal cell clusters were subset from the combined chicken and emu dataset for comparative analyses. Briefly, mesodermal clusters were subset to a new object, PCA and UMAP dimension reduction was recalculated using the top 250 highly variable genes and corrected for cell cycle effects. Next, the object was imported into Monocle3 [39,49] for k-means clustering (k=4). Cluster labels were confirmed by identifying differentially expressed marker genes through regression analysis implemented in monocle *fit_models* function and compared to those identified previously [2]. This resulted in 9 sub-clusters covering the major cell types within the developing limb field mesoderm.

Lineage trajectories throughout the mesoderm were constructed in Monocle3 using reverse graph embedding (k=4, minimum branch length = 15, rann.k = 50), which produced 3 major lineages that originated in E1.5 cells and terminated in E3.5 cells. Lineage bifurcation points were corroborated using know LPM marker gene expression through the *plot_cells* function, then pseudotime was calculated by selecting the origin of the lineages using the *order_cells* function. Cell type proportions were investigated along lineage trajectories and visualized as relative between chicken and emu across each stage.

Differential expression analysis

To perform differential expression between chicken and emu homologous clusters we used a pseudobulk approach. Cluster-specific cells were aggregated across sample, and samples with < 5 cells were removed. DE between chicken and emu clusters was performed using the *pseudoBulkDGE* function in *scran*, using stages as replicates. DE genes were visualised as volcano plots and through the *plot_cells* function in Monocle3. To identify genes that dynamically changed in expression across pseudotime, the LPM > somatic LPM > limb bud mesenchyme lineage was subset using the *chose_graph_segments* function in Monocle3.

Pseudotime values for each cell were extracted and DE expression between chicken and emu genes along pseudotime were calculated using R packages '*splines*' and '*limma*'.

Gene expression analysis by in situ hybridization and immunofluorescence

Whole mount *in situ* hybridization for spatial mRNA expression was carried out as described previously [2] with minor modifications. Briefly, whole HH8-HH22 chicken embryos were fixed overnight in 4% paraformaldehyde, dehydrated in methanol, and rehydrated in PBTX (PBS + 0.1% Triton X-1000). Tissues were permeabilized in 10µg/mL proteinase K for up to 1 hour, depending upon stage, then re-fixed in glutaraldehyde/ 4% PFA. Tissues underwent pre-hybridization (50% formamide, 5x SSC, 0.1% Triton X-100, 0.5% CHAPS, 5mM EDTA, 50mg/mL Heparin, 1mg/mL yeast RNA, 2% blocking powder) overnight at 65°C. The complete chicken *PRRX1* ORF was used as the riboprobe template, and amplified from limb and whole embryo cDNA using primers F-ATGGCGTCCAGCTATGCC and R-TTAATTGACTGTGGGCACTTG [2]. Fragments were resolved by 1% agarose electrophoresis, excised, and purified using a Nucleospin PCR clean-up kit and subcloned into p-GEM T-easy (Promega). Antisense RNA probes were synthesized using SP6 RNA polymerase and the DIG-labelling kit (Roche, #11277073910) as per the manufacturer's instructions. Precipitated probe was added to pre-hybridized tissues (approx. 5mL/ tube) and hybridization was carried out overnight at 65°C. Tissues were then subjected to stringency washes, blocked in BSA, then treated overnight with anti-DIG antibody conjugated with alkaline phosphatase. Tissues were exposed to BCIP/NBT colour reaction at room temperature for up to 3 hours (340mg/mL NBT and 175 mg/mL BCIP in NTMT (100mM NaCl, 100mM Tris-HCl, pH9.5, 50mM MgCl2, 0.1% Tween-20).

Chicken and emu embryos were fixed in 4% PFA/PBS for 15 minutes at room temperature then cryo-protected in 30% sucrose. Embryos were snap frozen in OCT and 10mm frozen sections were cut. Antigen retrieval was performed by the Monash Histology platform, where slides were heated at 60 degrees for 30 minutes followed by retrieval in citrate solution (pH 6). Sections were blocked and permeabilised in 1% Triton X-100, 2% BSA/PBS for 1-2hr at room temperature, then incubated with mouse anti-TWIST1 primary antibody (ab50887, Abcam) at 1:100 concentration in 0.5% Triton X-100, 1% BSA/PBS overnight at 4°C. An anti-mouse secondary (A21202, Invitrogen) was added at 1:1000 the following day and left at room temperature for one hour. Slides were washed once with DAPI diluted in PBS, followed by two washes in PBS.

Acknowledgements

We thank Micromon Genomics for assistance with sequencing design, Monash University Flowcore for flow cytometry, and Monash Histology platform for histological processing. We additionally thank Nadia Davidson, Alex Combes and Kieran Short for helpful suggestions for choice of bioinformatics methods, and constructive comments from all members of the Smith Lab. This work was supported by the Australian Research Council Discovery Project scheme (DP190100890 to Craig Smith).

Figure Legends

Figure 1. Identification of cell types in the chicken and emu forelimb field. (A) Cells were isolated from chicken embryonic day - E1.5, E2.5 and E3.5, and emu E3.5, E4.5 and E5.5 to sample all major tissues in the developing forelimb field. (C,D) tSNE visualisation separated chicken and emu cells based on stage and germ layer origin (B). (E,F) Unique gene expression profiles were detected for each major cluster of chicken and emu cells, which were largely in agreement, recapitulating shared cell clusters with overlapping marker genes.

Figure 2. Comparisons of cell type proportions in the chicken and emu developing forelimb field. (A) The complete chicken and emu datasets were integrated to produce a unified atlas of the developing forelimb field, recapitulating tissues from all germ layers. (B) Mesodermal cells were subset and clustered to produce (C) 9 clusters covering known mesodermal cell types. Total proportions of cells were identified for each cluster, corresponding to each species and stage. (D) A trajectory was fit from the undifferentiated LPM to the limb bud mesenchyme (B) and the presence and numbers of cells for cluster, per species per stage were identified. This revealed a reduced allocation of somatic LPM cells in the emu from undifferentiated LPM precurosrs (arrow), relative to the splanchnic LPM and to the chicken (E). Immunofluorescent labelling of the somatic LPM marker TWIST1 corroborated the reduced proportions of emu somatic LPM and limb bud cells, showing an overall reduced number of TWIST1+ cells in the emu, relative to the splanchnopleure and the chicken at roughly equivalent stages. som = somatopleure/somatic LPM; spl = splanchnopleure/splanchnic LPM; dm = dermomyotome; scl = sclerotome; nc = notochord; nt = neural tube.

Figure 3. Gene expression dynamics underlying emu wing heterochrony. (A) Pseudobulk differential expression of emu lateral plate mesoderm and forelimb bud mesenchyme, relative to the chicken, revealed downregulation of several LPM markers (i.e. *PRRX1*) prior to limb bud formation, and limb genes (i.e. *FGF10*) post limb-initiation. (B) Visualization of differential expression of genes across pseudotime revealed downregulation of LPM and limb markers in the emu relative to the chicken across development. *PRRX1*, *TWIST1*, *TBX5*, *FGF10* and *ALDH1A2* each showed stage-specific decreases in expression, relative to chicken. (C) Spatial and temporal *in situ* visualization of *PRRX1* gene expression during chicken and emu forelimb development. This emu possessed reduced and delayed activation of PRRX1 in the developing forelimb field (black arrow) despite covering a similar number of somites, relative to the chicken. In contrast, *PRRX1* expression was increased and accelerated in the emu hindlimb relative to the chicken (white arrow). Together these suggest the emu has delayed and accelerated LPM specification in the forelimb and hindlimb, respectively.

Figure S1. Individual UMAP visualization and clustering of chicken and emu mesoderm. Cell types were in agreement between the two datasets, show all mesoderm-derived cell types in the developing forelimb field. NKX2-5 has previously been noted to be co-opted in the developing emu forelimb [35], which we find expressed in early somatic LPM cells and the somites, which is not present in the chicken.

Figure S2. Labelling of chicken mesoderm clusters on integrated chicken and emu dataset. Clusters were largely in agreement between the chicken atlas [2] and merged data.

Figure S3. Gene ontology enrichment of emu downregulated genes across pseudotime. Enriched terms included forelimb morphogenesis and development.

References

1. Funayama, N., Sato, Y., Matsumoto, K., Ogura, T. & Takahashi, Y. Coelom formation: Binary decision of the lateral plate mesoderm is controlled by the ectoderm. *Development* **126**, 4129–4138 (1999).

2. Newton, A. H., Williams, S. M., Major, A. T. & Smith, C. A. Cell lineage specification and signalling pathway usage during development of the lateral plate mesoderm and forelimb mesenchyme. *Development* **149**, 2022.01.13.475748 (2022).

3. Agarwal, P. *et al.* Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. *Development* **130**, 623–633 (2003).

4. Ng, J. K. *et al.* The limb gene Tbx5 promotes limb initiation by interacting with Wnt2b and Fgf10. *Development* **129**, 5161–5170 (2002).

5. Ohuchi, H. *et al.* The mesenchymal factor, FGF10, initiates and maintains the outgrowth of the chick limb bud through interaction with FGF8, an apical ectodermal factor. *Development* **124**, 2235–2244 (1997).

6. Takeuchi, J. K. *et al.* Tbx5 and Tbx4 genes determine the wing/leg identity of limb buds. *Nature* **398**, 810–814 (1999).

7. Tickle, C. How the embryo makes a limb: Determination, polarity and identity. *J. Anat.* **227**, 418–430 (2015).

8. Mcqueen, C. & Towers, M. Establishing the pattern of the vertebrate limb. *Development* **147**, (2020).

9. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).

10. Roberts, D. J. *et al.* Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development* **121**, 3163–3174 (1995).

11. Sekine, K. *et al.* Fgf10 is essential for limb and lung formation. *Nat. Genet.* **21**, 138–141 (1999).

12. Rallis, C. *et al.* Tbx5 is required for forelimb bud formation and continued outgrowth. *Development* **130**, 2741–2751 (2003).

13. Moon, A. M. & Capecchi, M. R. Fgf8 is required for outgrowth and patterning of the limbs. *Nat. Genet.* **26**, 455–459 (2000).

14. Petit, F., Sears, K. E. & Ahituv, N. Limb development: A paradigm of gene regulation. *Nat. Rev. Genet.* **18**, 245–258 (2017).

15. Newton, A. H. & Smith, C. A. Regulation of vertebrate forelimb development and wing reduction in the flightless emu. *Dev. Dyn.* **250**, dvdy.288 (2020).

16. Hockman, D. *et al.* A second wave of Sonic hedgehog expression during the development of the bat limb. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16982–16987 (2008).

17. Shapiro, M. D., Hanken, J. & Rosenthal, N. Developmental Basis of Evolutionary Digit Loss in the Australian Lizard Hemiergis. *J. Exp. Zool. Part B Mol. Dev. Evol.* **297**, 48–56 (2003).

18. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633-642.e11 (2016).

19. Leal, F. & Cohn, M. J. Loss and Re-emergence of Legs in Snakes by Modular Evolution of Sonic hedgehog and HOXD Enhancers. *Curr. Biol.* **26**, 2966–2973 (2016).

20. Thewissen, J. G. M. *et al.* Developmental basis for hind-limb loss in dolphins and origin of the cetacean bodyplan. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8414–8418 (2006).

21. Lopez-Rios, J. *et al.* Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* **511**, 46–51 (2014).

22. Cooper, K. L. *et al.* Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* **511**, 41–45 (2014).

23. Kawahata, K. *et al.* Evolution of the avian digital pattern. *Sci. Rep.* **9**, 1–7 (2019).

24. Harshman, J. *et al.* Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc. Natl. Acad. Sci.* **105**, 13462–13467 (2008).

25. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

26. Sackton, T. B. *et al.* Convergent regulatory evolution and loss of flight in paleognathous birds. *Science (80-.).* **364**, 74–78 (2019).

27. Huynen, L. *et al.* Reconstruction and in vivo analysis of the extinct tbx5 gene from ancient wingless moa (Aves: Dinornithiformes). *BMC Evol. Biol.* **14**, 1–8 (2014).

28. De Bakker, M. A. G. *et al.* Digit loss in archosaur evolution and the interplay between selection and constraints. *Nature* **500**, 445–448 (2013).

29. Maxwell, E. E. & Larsson, H. C. E. Osteology and myology of the wing of the emu (Dromaius novaehollandiae), and its bearing on the evolution of vestigial structures. *J. Morphol.* **268**, 423–441 (2007).

30. Nagai, H. *et al.* Embryonic development of the emu, Dromaius novaehollandiae. *Dev. Dyn.* **240**, 162–175 (2011).

31. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).

32. Young, J. J. *et al.* Attenuated Fgf Signaling Underlies the Forelimb Heterochrony in the Emu Dromaius novaehollandiae. *Curr. Biol.* **29**, 3681-3691.e5 (2019).

33. Bickley, S. R. B. & Logan, M. P. O. Regulatory modulation of the T-box gene Tbx5 links development, evolution, and adaptation of the sternum. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17917–17922 (2014).

34. Smith, C. A. *et al.* Limb patterning genes and heterochronic development of the emu wing bud. *Evodevo* **7**, 1–17 (2016).

35. Farlie, P. G. *et al.* Co-option of the cardiac transcription factor Nkx2.5 during development of the emu wing. *Nat. Commun.* **8**, 1–11 (2017).

36. Bell, G. W., Yatskievych, T. A. & Antin, P. B. GEISHA, a whole-mount in situ

hybridization gene expression screen in chicken embryos. *Dev. Dyn.* **229**, 677–687 (2004).

37. Darnell, D. K. *et al.* GEISHA: an in situ hybridization gene expression resource for the chicken embryo. *Cytogenet. Genome Res.* **117**, 30–35 (2007).

38. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).

39. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

40. Kuratani, S. *et al.* The expression pattern of the chick homeobox gene gMHox suggests a role in patterning of the limbs and face and in compartmentalization of somites. *Dev. Biol.* **161**, 357–369 (1994).

41. Nohno, T. *et al.* A chicken homeobox gene related to drosophila paired is predominantly expressed in the developing limb. *Dev. Biol.* **158**, 254–264 (1993).

42. Martin, J. F., Bradley, A., Olson, E. N. & Eric, N. O. The paired-like homeo box gene MHox is required for early events of skeletogenesis in multiple lineages. *Genes Dev.* **9**, 1237–1249 (1995).

43. Cooper, K. L. The Lesser Egyptian Jerboa, Jaculus jaculus: A Unique Rodent Model for Evolution and Development. *Cold Spring Harb. Protoc.* **2011**, pdb.emo066704 (2011).

44. Saxena, A. *et al.* Interspecies transcriptomics identify genes that underlie disproportionate foot growth in jerboas. *Curr. Biol.* **32**, 289-303.e6 (2022).

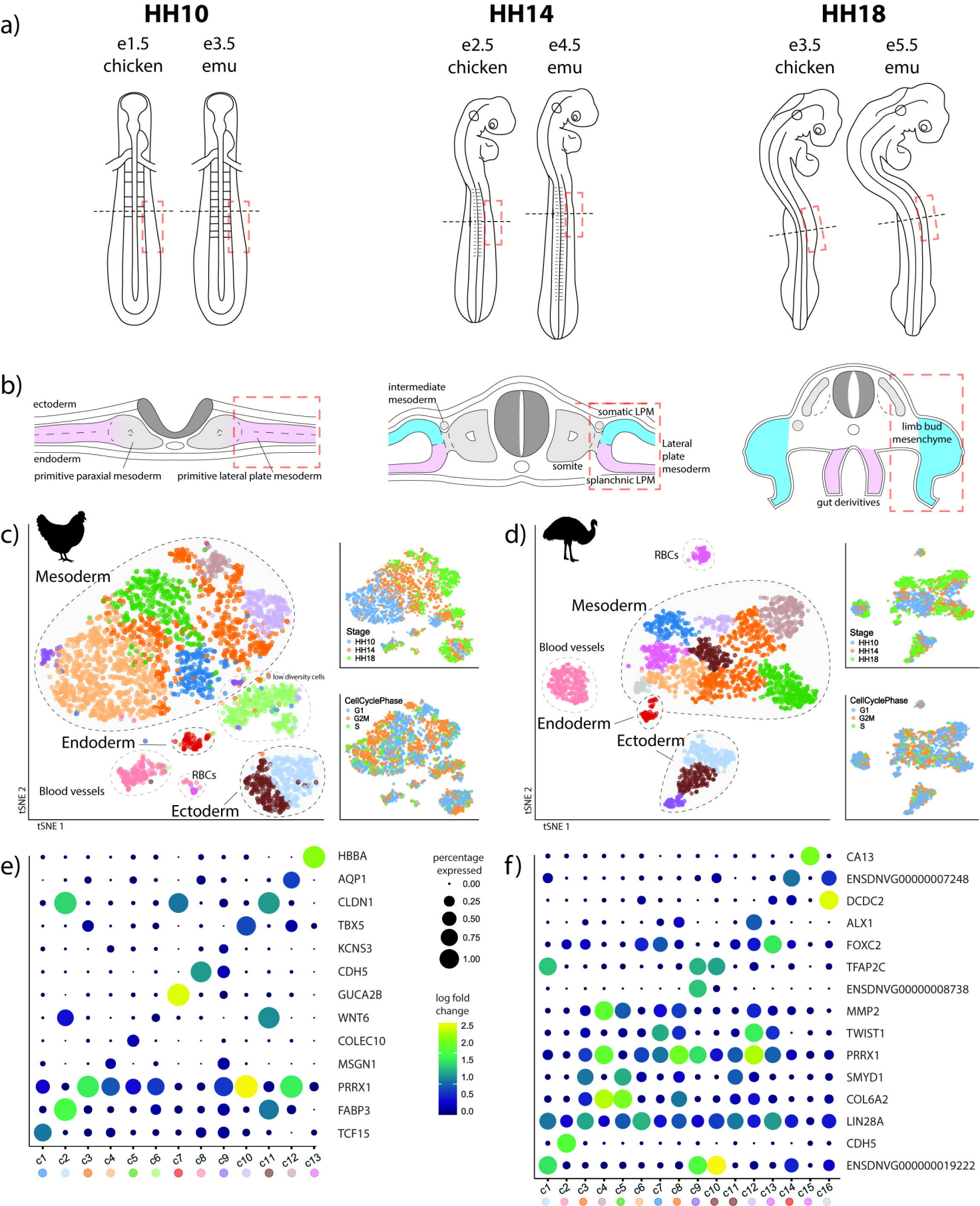
45. Lun, A. T. L., Mccarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor [version 2□; referees□: 3 approved , 2 approved with reservations]. *F1000Research* **5**, 1–69 (2016).

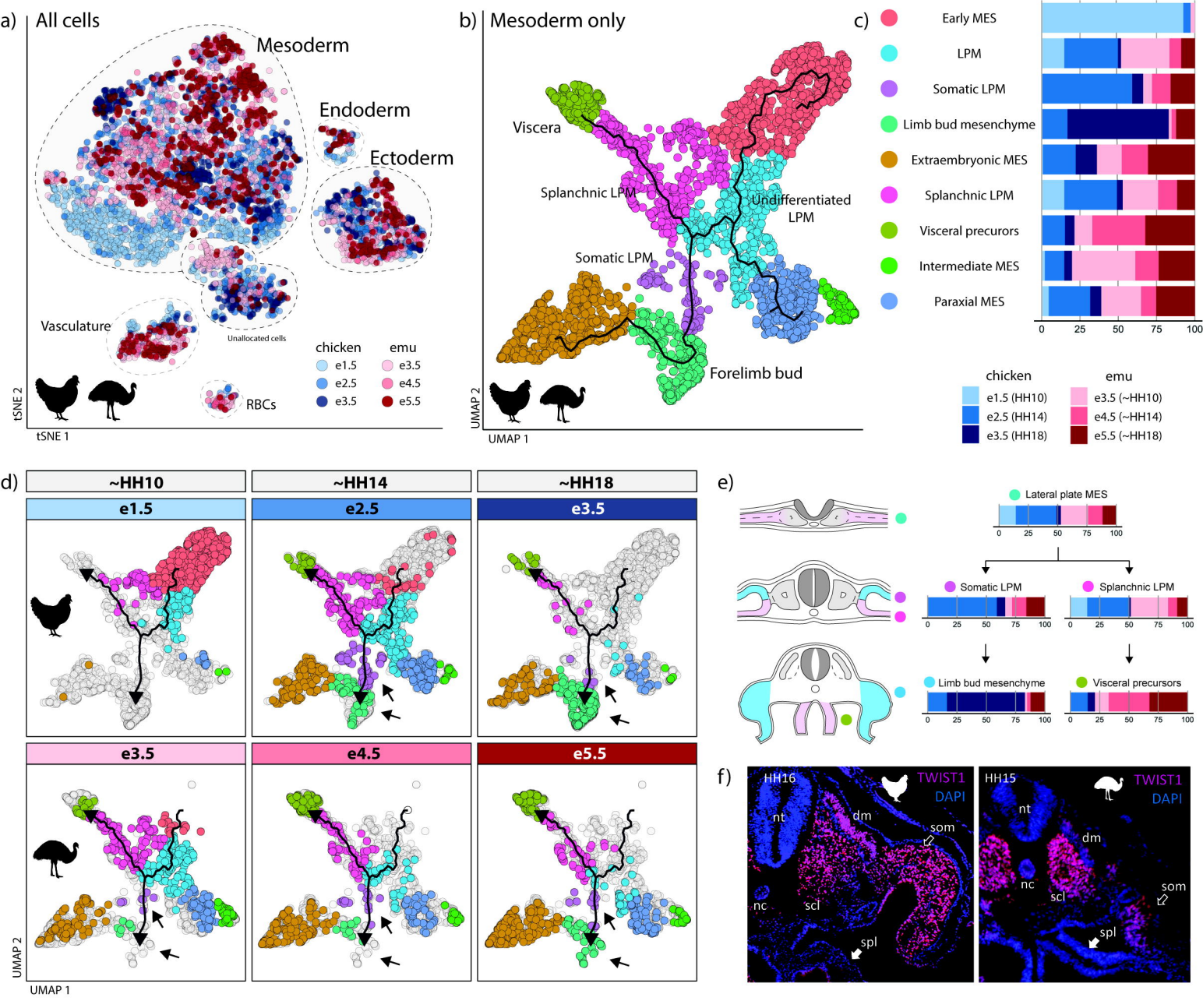
46. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).

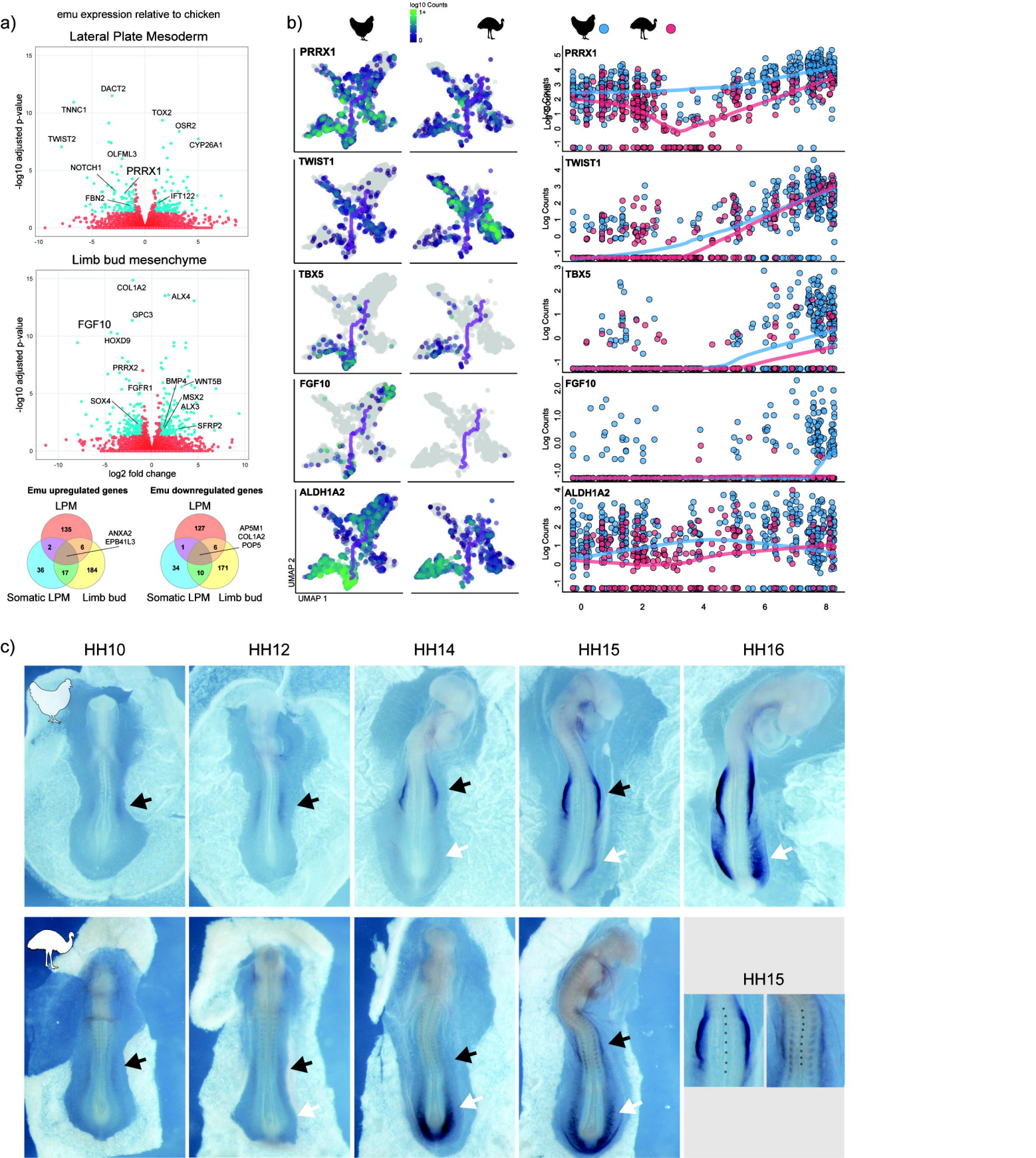
47. Lun, A. T. L., Rue-Albrecht, K., Marini, F. & Soneson, C. iSEE: Interactive SummarizedExperiment Explorer [version 1; referees: 2 approved]. *F1000Research* **7**, (2018).

48. Scialdone, A. *et al.* Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54–61 (2015).

49. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).


a)
HH10
e1.5 chicken
e3.5 emu
HH14
e2.5 chicken
e4.5 emu
HH18
e3.5 chicken
e5.5 emu
b)
ectoderm
endoderm
primitive paraxial mesoderm
primitive lateral plate mesoderm
intermediate mesoderm
somatic LPM
Lateral plate mesoderm
somite
splanchnic LPM
limb bud mesenchyme
gut derivitives
c)
Mesoderm
low diversity cells
Endoderm
RBCs
Blood vessels
Ectoderm
tSNE 1
tSNE 2
Stage
HH10
HH14
HH18
CellCyclePhase
G1
G2M
S
d)
RBCs
Mesoderm
Blood vessels
Endoderm
Ectoderm
tSNE 1
tSNE 2
Stage
HH10
HH14
HH18
CellCyclePhase
G1
G2M
S
e)
HBBA
AQP1
CLDN1
TBX5
KCNS3
CDH5
GUCA2B
WNT6
COLEC10
MSGN1
PRRX1
FABP3
TCF15
c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 c13
percentage expressed
0.00
0.25
0.50
0.75
1.00
log fold change
2.5
2.0
1.5
1.0
0.5
0.0
f)
CA13
ENSDNVG00000007248
DCDC2
ALX1
FOXC2
TFAP2C
ENSDNVG00000008738
MMP2
TWIST1
PRRX1
SMYD1
COL6A2
LIN28A
CDH5
ENSDNVG000000019222
c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 c13 c14 c15 c16


a)
All cells
Mesoderm
Endoderm
Ectoderm
Vasculature
Unallocated cells
RBCs
chicken
e1.5
e2.5
e3.5
emu
e3.5
e4.5
e5.5
tSNE 2
tSNE 1
b)
Mesoderm only
Viscera
Splanchnic LPM
Undifferentiated LPM
Somatic LPM
Forelimb bud
UMAP 2
UMAP 1
c)
Early MES
LPM
Somatic LPM
Limb bud mesenchyme
Extraembryonic MES
Splanchnic LPM
Visceral precurors
Intermediate MES
Paraxial MES
0
25
50
75
100
chicken
e1.5 (HH10)
e2.5 (HH14)
e3.5 (HH18)
emu
e3.5 (~HH10)
e4.5 (~HH14)
e5.5 (~HH18)
d)
~HH10
~HH14
~HH18
e1.5
e2.5
e3.5
e3.5
e4.5
e5.5
UMAP 2
UMAP 1
e)
Lateral plate MES
Somatic LPM
Splanchnic LPM
Limb bud mesenchyme
Visceral precursors
f)
HH16
TWIST1
DAPI
nt
dm
som
nc
scl
spl
HH15
TWIST1
DAPI
nt
dm
nc
scl
som
spl


a)
emu expression relative to chicken
Lateral Plate Mesoderm
-log10 adjusted p-value
DACT2
TNNC1
TOX2
OSR2
TWIST2
CYP26A1
OLFML3
NOTCH1
PRRX1
FBN2
IFT122
Limb bud mesenchyme
COL1A2
ALX4
GPC3
FGF10
HOXD9
PRRX2
BMP4
WNT5B
FGFR1
MSX2
SOX4
ALX3
SFRP2
log2 fold change
Emu upregulated genes
LPM
135
2
6
36
17
184
ANXA2
EPB41L3
Somatic LPM
Limb bud
Emu downregulated genes
LPM
127
1
6
34
10
171
AP5M1
COL1A2
POP5
Somatic LPM
Limb bud
b)
log10 Counts
1+
0
PRRX1
TWIST1
TBX5
FGF10
ALDH1A2
UMAP 2
UMAP 1
Log Counts
c)
HH10
HH12
HH14
HH15
HH16
HH15