Feregrino et al.

1 **A Single-cell Transcriptomic Atlas of the Developing Chicken Limb**

2

3 Christian Feregrino[1], Fabio Sacher[1], Oren Parnas[2, §], Patrick Tschopp[1]*

4

5 [1]DUW Zoology, University of Basel, Vesalgasse 1, CH-4051, Basel, Switzerland; [2]Klarman

6 Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

7 [§]Present address: The Concern Foundation Laboratories at the Lautenberg Centre for

8 Immunology and Cancer Research, IMRIC, Hebrew University Faculty of Medicine,

9 Jerusalem, 91120, Israel

10

11 Email addresses: christian.feregrino@unibas.ch, fabio.sacher@stud.unibas.ch,

12 orenp@ekmd.huji.ac.il, patrick.tschopp@unibas.ch

13

14 *Corresponding author: Patrick Tschopp

15 Tel.: +41 61 207 56 49

16 patrick.tschopp@unibas.ch

Feregrino et al.

Abstract

Background. Through precise implementation of distinct cell type specification programs, differentially regulated in both space and time, complex patterns emerge during organogenesis. Thanks to its easy experimental accessibility, the developing chicken limb has long served as a paradigm to study vertebrate pattern formation. Through decades' worth of research, we now have a firm grasp on the molecular mechanisms driving limb formation at the tissue-level. However, to elucidate the dynamic interplay between transcriptional cell type specification programs and pattern formation at its relevant cellular scale, we lack appropriately resolved molecular data at the genome-wide level. Here, making use of droplet-based single-cell RNA-sequencing, we catalogue the developmental emergence of distinct tissue types and their transcriptome dynamics in the distal chicken limb, the so-called autopod, at cellular resolution.

Results. Using single-cell RNA-sequencing technology, we sequenced a total of 17,628 cells coming from three key developmental stages of chicken autopod patterning. Overall, we identified 23 cell populations with distinct transcriptional profiles. Amongst them were small, albeit essential populations like the apical ectodermal ridge, demonstrating the ability to detect even rare cell types. Moreover, we uncovered the existence of molecularly distinct sub-populations within previously defined compartments of the developing limb, some of which have important signaling functions during autopod pattern formation. Finally, we inferred gene co-expression modules that coincide with distinct tissue types across developmental time, and used them to track patterning-relevant cell populations of the forming digits.

Conclusions. We provide a comprehensive functional genomics resource to study the molecular effectors of chicken limb patterning at cellular resolution. Our single-cell transcriptomic atlas captures all major cell populations of the developing autopod, and highlights the transcriptional complexity in many of its components. Finally, integrating our data-set with other single-cell transcriptomics resources will enable researchers to assess molecular similarities in orthologous cell types across the major tetrapod clades, and provide an extensive candidate gene list to functionally test cell-type-specific drivers of limb morphological diversification.

Keywords

scRNA-seq, Gene Expression, Cellular Transcriptomics, Autopod Patterning, Digits, Interdigit, Perichondrium, Phalanges

Background

51 Embryonic pattern formation relies on the tight coordination of numerous developmental

52 processes, across multiple scales of complexity. From seemingly homogenous progenitor

53 populations, different cell types get specified and arranged in intricate patterns, to give rise to

54 functional tissues and organs. As progenitors mostly share a common genome, this

55 phenotypic specialization relies on the precise execution of distinct gene regulatory networks,

56 to enable cell type specification and ensuing pattern formation [1–3]. Slight deviations in

57 these processes contribute to morphological variations within natural populations. More

58 profound aberrations, however, can cause malformations and ultimately result in death of the

59 embryo. To buffer such fragile balance, many cell type specification and pattering processes

60 rely on complex feedback mechanisms, through tightly interconnected molecular loops

61 between spatially distinct signaling centers [4–6] Hence, integration of multiple signaling

62 pathways across space and time define a molecular coordinate grid to instruct organogenesis

63 at the tissue level. Ultimately, however, these multifaceted signaling inputs have to be

64 incorporated at the cellular level, *via* cell type-specifying gene regulatory networks, as

65 progenitor cells undergo spatially and temporally defined cell fate decisions to contribute to

66 proper pattern formation.

67 Tetrapod limb development has long served as a model to study the genetic and molecular

68 underpinnings of vertebrate pattern formation. Due to its non-essentiality for embryo survival,

69 many fetuses carrying mutations that affect limb development make it to full term.

70 Accordingly, human geneticists have been able to accumulate an impressive catalogue of

71 candidate genes for limb patterning [7–9]. Combined with the easy accessibility of the limb in

72 chicken embryos, and molecular genetic tools in the mouse, decades of experimental work

73 have resulted in an in-depth understanding of many of the molecular mechanisms driving limb

74 formation at the tissue scale [5]. Moreover, given the profound morphological diversifications

75 the basic limb structure has experienced in numerous tetrapod clades, limb development has

76 long attracted the interests of comparative developmental biologists using 'EvoDevo'

77 approaches [10]. This holds especially true for the most distal portion of the limb, the

78 autopod, i.e. hands and feet. There, species-specific adaptations to distinct modes of

79 locomotion have resulted in a diverse array of digit number formulas and individualized digit

80 patterns [11–14].

81 Early in development, proliferation of a lateral plate mesoderm (LPM)-derived mesenchymal

82 progenitor population drives overall limb bud outgrowth. Signaling crosstalk with a

83 specialized structure of the distal overlaying ectoderm, the apical ectodermal ridge (AER),

3

84 controls these dynamics. Concurrently, the major embryonic axes of the limb are defined by
85 the coordinated action of multiple signaling centers [reviewed in 5]. As development
86 progresses, LPM-derived progenitors start to differentiate into skeletal and other connective
87 tissue types [15–17], while muscles cells originating from the somites migrate into the limb
88 bud to complement formation of the musculoskeletal apparatus [18, 19]. For autopod pattern
89 formation, digit numbers and identities are first defined by posteriorly restricted sonic
90 hedgehog (SHH) activity, and altered by modulations therein [10, 14, 20, reviwed in 21].
91 Digit elongation then relies on a specialized distal progenitor population, which supports
92 outgrowth of individual digit bones, the phalanges [22, 23]. Digit-specific phalanx-formulas,
93 and their stereotypic connection patterns *via* synovial joints, are established by signals
94 emanating from the posterior interdigit mesenchyme [24, 25].

95 In this study, capitalizing on the power of droplet-based single-cell RNA-sequencing, we
96 resolve the underlying transcriptional dynamics of autopod tissue formation and pattern
97 emergence at single-cell resolution, across three stages of chicken hindlimb development. In
98 total, we present transcriptomic data for 17,628 cells, allowing us to identify all major tissue
99 types of the developing limb, as well as a substantial amount of molecular heterogeneity
100 therein. Through weighted correlation network analysis, we define distinct gene co-expression
101 modules that track corresponding tissue types across developmental time. Finally, we focus
102 on the molecular make-up of cell populations involved in digit pattern formation and, hence,
103 putative drivers of morphological diversification in the autopod.

104 Collectively, we present a comprehensive genomics resource that for the first time reveals the
105 transcriptome dynamics of the developing chicken foot at cellular level. Our study identifies
106 novel and known marker genes in co-expression modules of patterning-relevant cell
107 populations, thereby providing an extensive catalogue of candidate genes for functional
108 follow-up studies, to elucidate the molecular mechanisms of autopod pattern formation and
109 diversification.

Results

Singe-cell sampling of the developing distal chicken limb

To follow the appearance of patterning-relevant cell populations and their associated transcriptome dynamics, we sampled three developmental stages of the embryonic chicken foot: stage Hamburger-Hamilton 25 (HH25, ~4.5 days of development), stage HH29 (~6 days of development) and stage HH31 (~7 days of development). This time window spans key morphogenetic events that drive species-specific patterns in the developing autopod, particularly for the skeletal apparatus and its associated tissues. Namely, stage HH25 is dominated by overall autopod outgrowth and delineation of the main embryonic axes, at HH29 digit-specific patterns differentiate, and at HH31 digit elongation is phasing out. We designed our tissue sampling strategies accordingly. At HH25, we captured the entire distal part of the growing limb (Fig. 1a), at HH29 we dissected two digits with distinct skeletal formulas, digit 3 and 4, as well as their adjacent interdigit mesenchyme (Fig. 1b), and at HH31 we focused on the tip of digit 4 with its growth-relevant progenitor population (Fig. 1c). We dissociated the micro-dissected tissue pieces using enzymatic digest combined with mechanical shearing and prepared single-cell suspensions for droplet-based high-throughput single-cell RNA-sequencing (*10X Genomics* and *Drop-Seq* [26, 27]). Using the corresponding bioinformatics pipelines, the resulting Next-Generation Sequencing libraries were mapped to the chicken genome, de-multiplexed according to their cellular barcodes and quantified to generate gene/cell read count tables. In total, we sampled over 17,000 cells and obtained single-cell transcriptomic profiles for 5,982 (HH25), 6,823 (HH29) and 4,823 (HH31) individual cells, respectively (Additional file 1: Fig. S1a). Quality-based exclusion of single-cell transcriptomes was implemented based on mean library size, percentage of mitochondrial reads and number of genes detected per cell. Additionally, data normalization as well as batch and cell cycle corrections were performed (for details, please refer to the *Methods* section). On average, we detected 2,879 unique molecular identifiers (UMIs) and 1,081 genes per cell (Additional file 1: Fig. S1b,c).

Autopod tissue composition at cellular resolution

Using unsupervised graph-based clustering, we identified 5, 10 and 5 clusters at stages HH25, HH29 and HH31, respectively. Projecting these clusters onto stage-specific tSNE (t-Distributed Stochastic Neighbor Embedding [28]), plots of our cellular transcriptomes revealed the presence of a dominant bulk of cells, with varying degrees of sub-structure, as well as distinct outlier groups (Fig. 1 d-f). Based on the expression of known marker genes and gene ontology (GO)-term enrichment analyses, we were able to attribute these broadly

5

144   defined cell populations to distinct tissue types (Fig. 1g-f, Additional file 1: Fig. S1a and Fig.

145   S2a-c). At stage HH25, they comprise a largely undifferentiated and proliferating

146   mesenchymal population (red), early skeletal progenitors (blue), muscle cells invading the

147   limb (black), as well as skin (purple) and blood cells (grey) (Fig. 1d,g). We recovered cell

148   populations corresponding to those same five tissue types in our HH29 sample, with the

149   exception that the "blood cluster" was now dominated by white blood cells and not

150   erythrocytes. Additionally, we identified cell populations matching the interdigit mesenchyme

151   (green), non-skeletal connective tissue (nsCT, maroon), cells enriched for markers of the very

152   distal margin of the autopod mesoderm ("distal mesenchyme", yellow), as well as endothelial

153   (brown) and smooth muscle (orange) cells of the forming blood vessels (Fig. 1e,h). At stage

154   HH31, we again find a largely undifferentiated mesenchymal population, the interdigit and

155   distal margin mesenchyme, skeletal and skin cells (Fig. 1f,i). As expected according to our

156   sampling strategy, for spatial and/or temporal context, we did not find all cell populations in

157   every dataset. For example, while sample HH25 is biggest in relative size to the autopod, it is

158   the earliest stage and thus predictably displayed the lowest cellular complexity. We observed

159   the opposite trend in HH31, where the relative size is smallest but development more

160   advanced. Our most complex dataset, in terms of cell number and tissue types identified, is

161   from stage HH29. Collectively, using broad graph-based clustering and molecular profiling on

162   our single-cell transcriptomics data, we catalogued the tissue composition of the developing

163   autopod with cellular resolution, across three developmental stages.

164   **Fine-scale clustering and marker gene expression across developmental time**

165   Although all expected major tissue types were recovered in our primary analyses, smaller cell

166   populations, some well known to be essential for limb outgrowth and patterning, remained

167   elusive. Hence, given our sampling depth, we next examined our data for additional sub-

168   structure. Indeed, upon closer inspection using finer-tuned clustering parameters, we did find

169   additional sub-populations with distinct transcriptional signatures (Fig. 2a-c, Additional file 1:

170   Fig. S1a). Based on differential expression analyses, we identified marker genes for each of

171   these sub-populations (Additional files 2-4). Certain sub-population/marker gene-

172   combinations appeared to be conserved in all three samples, thereby allowing us to assign

173   cellular equivalencies across developmental time (Fig. 2d-f). A subset of marker genes only

174   showed loosely restricted expression patterns, likely a reflection of the largely

175   undifferentiated state of the corresponding sub-population. For example, *PRRX1*, a well-

176   established marker of the limb mesenchyme [16, 29, 30], and *PCNA*, active during DNA

177   replication in proliferating cells [31], showed varying levels of expression beyond the

178 proliferating mesenchyme sub-clusters. Such transcriptional ambiguities, however, seemed

179 progressively lost, as mesenchymal progenitors committed to the different skeletal and non-

180 skeletal lineages that define the emerging autopod patterns (Fig. 2d-f). As expected, cell sub-

181 populations residing outside the LPM-lineage showed more pronounced transcriptome

182 individualizations. For example, at HH25 the ectodermal 'skin' population got split into two

183 distinct sub-clusters, one representing the bulk amount of the embryonic skin covering the

184 autopod (sub-cluster 8), and the other corresponding to the apical ectodermal ridge (sub-

185 cluster 7). Expression of its canonical marker *FGF8* and other highly enriched genes clearly

186 established AER identity, demonstrating that even small cell populations can be successfully

187 captured (Fig. 2d).

Gene co-expression modules and corresponding tissue types

189 To gain further insights into the regulatory programs that maintain these transcriptional

190 signatures, and explore their potential biological significance, we tested for the occurrence of

191 transcriptome-wide gene co-expression patterns using weighted correlation network analysis

192 (WGCNA) [32]. This approach consists of an unsupervised clustering of genes based on their

193 expression pattern across all cells, irrespective of the assigned cell or tissue type. In order to

194 comprehensively screen for relevant gene co-expression modules, we conducted the analysis

195 in our transcriptionally most complex sample at stage HH29. Starting with genes that showed

196 high levels and variation of expression, we calculated an adjacency matrix and its topological

197 overlap to construct a hierarchical tree. The resulting tree was cut to obtain a first set of gene

198 co-expression modules. We then computed the first principal component of each module, to

199 define so-called 'module eigengenes'. For each individual gene, correlation to the respective

200 eigengenes was used to assess module membership. Genes not significantly correlated with

201 any eigengene were discarded, after which the entire process was repeated iteratively with a

202 reduced gene set. Eventually, we identified a total of 836 genes grouped in 16 distinct gene

203 co-expression modules, each designated by a color (Fig. 3a). Final module sizes ranged from

204 15 to 215 genes (Additional file 5).

205 On a cell-by-cell basis, we calculated the average expression for each of the co-expression

206 modules and visualized their distribution on our stage HH29 tSNE plot (Additional file 1: Fig.

207 S3). Compared to our initial clustering of sample HH29, we found co-expression modules

208 specifically enriched in the following cell populations: blood cells (module Black), skin

209 (Blue), blood vessel endothelium (Brown), nsCT (Darkgrey), distal mesenchyme (Magenta),

210 chondrocytes (Red and Turquoise) and muscle (Yellow). Interestingly, GO-terms associated

211 with more broadly distributed modules enabled us to attribute the sub-clustering structure of

212   certain tissues to particular biological processes. For example, HH29 mesenchyme sub-cluster

213   5 showed higher activity for module Green, associated with GO-terms connected to mitosis,

214   whereas sub-cluster 16 was enriched for module Pink, linked to G2/M-transition-related genes

215   (Additional file 1: Fig. Fig. S3). Hence, we reasoned that distinct cell-cycle states underlie the

216   subdivision of the proliferating mesenchyme cluster. Likewise, HH29 interdigit sub-clusters

217   2, 6 and 12 were closely matched by the activities of modules Tan, Olivegreen, Orange and

218   Midnightblue (see below, Fig. 4a-h).

219   To follow the developmental dynamics of the identified modules, we calculated their

220   averaged activities across all the three sampled time points, and visualized similarities across

221   time and tissue types using unsupervised hierarchical clustering (Fig. 3b). Indeed, despite

222   differences in embryonic stages and experimental platforms, we were able to confirm

223   corresponding cell and tissue types between our samples. For example, what we refer to as the

224   "distal mesenchyme" is a population of cells characterized by high activity of the co-

225   expression module Magenta at all time points (Fig. 3c-f). Comparisons to published

226   expression patterns for *TFAP2B, WNT5A*, *MSX1* and *MSX2* confirmed its distal location and,

227   based on those genes' functions, suggested a role for this cell population in controlling distal

228   autopod outgrowth. Using WGCNA thus enabled us to define equivalent cell populations

229   across developmental time, and helped attribute biological functions at the sub-cluster level.

230   **Transcriptionally and spatially distinct sub-populations in the interdigit mesenchyme**

231   As expected by developmental stage, interdigit populations were only recovered in samples

232   HH29 and HH31. In total, we identified four associated co-expression modules (Fig. 4a-d).

233   High Orange and Olivegreen module activities were coinciding with the same interdigit sub-

234   population (Fig. 4e,f), which was recognizable in both HH29 and HH31 samples and marked

235   by *RDH10* expression (Fig. 2e,f). Noticeably, all genes with high membership in module

236   Olivegreen were transcription factors (TFs), while module Orange was enriched for

237   enzymatic activities (Fig. 4a,b). Both, however, scored high for GO-terms related to retinoic

238   acid signaling, an important mediator of interdigit cell death [33]. Module Tan was enriched

239   for skeletogenic and morphogenetic GO-terms, suggesting it might mediate some of the

240   patterning information contained in the interdigit mesenchyme to the adjacently forming

241   digits (Fig. 4c,g). Lastly, module Midnightblue showed multiple TFs and its activity was

242   restricted to HH29 sub-cluster 2 (Fig. 4d,h).

243   Since relevant patterning information is contained in the interdigit, posteriorly adjacent to

244   each forming digit, we next wondered whether some of the sub-clustering structure

245   corresponded to spatially distinct interdigit populations along the anterior-posterior axis of the

8

246   autopod. At HH29, we detected three interdigit sub-clusters (Fig. 4i). Using differential
247   expression analyses, we defined marker genes that distinguish the three sub-clusters from
248   each other (Fig. 4j). To assign putative spatial information to our single-cell interdigit
249   transcriptomes, we reanalyzed a bulk RNA-seq dataset covering stages HH29 and HH31 of
250   the developing chicken hindlimb autopod [34]. This dataset is based on dissections of
251   individual digits, together with their posteriorly associated interdigit mesenchyme, and thus
252   provided an opportunity to identify spatially resolved marker genes. We contrasted their
253   transcriptomic data of digit/interdigit III against digit/interdigit IV and found a total of 54
254   genes to be significantly differentially expressed at both developmental time points (Fig. 4k).
255   Comparing the digit/interdigit IV-specific subset of these genes to our differential expression
256   analysis of sub-cluster 2, and its affiliated module Midnightblue, we found an overlap of
257   seven up-regulated genes (Fig. 4d,j, underlined). In contrast, we couldn't find any other
258   digit/interdigit IV gene in the rest of the interdigit sub-cluster signatures or co-expression
259   modules. We therefore concluded that HH29 sub-cluster 2 consisted of cells of the interdigit
260   mesenchyme posterior to digit 4.

261   **Developing digits and their associated tissues**

262   Of the cell populations directly contributing to the making of digits, a cluster reminiscent of
263   the non-skeletal connective tissue, the nsCT, appeared in all of the samples. In our WGCNA
264   analyses, we identified three modules, Darkgrey, Purple, and Darkgreen, which mapped to the
265   nsCT sub-clusters (Fig. 5a-f). The Darkgrey module was most restricted, in both time and cell
266   numbers, and its activity pattern closely matched the HH29 sub-cluster 4 (Fig. 5d). Cellular
267   retinoic acid binding protein I *CRABP-I*, Aquaporin *AQP1*, *DKK2* and *GLT8D2* were the
268   genes most strongly associated with this module. Modules Purple and Darkgreen showed
269   more widespread activities (Fig. 5e,f), and centered on *COL1A2*, *DCN*, *KCNJ2*, *SALL1*, and
270   *AKR1D1*, *PRRX1*, *TCF12, ZFHX3*. Comparing our differential expression analyses between
271   the respective cell populations, only six genes appeared significantly enriched across all
272   stages (Fig. 5g), five of which also appeared in our nsCT modules. Using *in situ* hybridization
273   for the top-three of these genes, both differential expression- and module membership-wise,
274   allowed us to attribute module activities to discrete nsCT domains along the developing
275   skeletal elements. *CRABP-I* showed highest expression near and around the forming
276   epiphysis, where synovial joints and ligament attachment sites develop (Fig. 5h). *COL1A2*-
277   and *ZFHX3*-positive populations showed a graded distribution along the periskeletal tissue
278   layer, predominantly marking the prospective periosteum and perichondrium domains,
279   respectively (Fig. 5i,j).

9

280   Finally, we identified skeletal progenitor populations at all three time points (Fig. 6a-c).

281   According to the developmental stages we sampled, only cartilage-producing skeletal cells

282   were recovered. In all three samples, we found a cell population resembling early

283   chondrocytes (sub-clusters HH25-4, HH29-15 and HH31-2). At stages HH29 and HH31, a

284   seemingly more mature chondrocyte type emerged (HH29-3, HH31-1), and an additional

285   cartilaginous cluster was evident in the HH29 sample (HH29-17). Concomitantly, we

286   identified two co-expression modules associated with these cell populations, Turquoise and

287   Red (Fig 6d,e). Turquoise is centered on *CD24*, *CHGB* and *SULF1,* whereas module Red

288   displays a core of collagens *COL9A1* and *COL9A3*, *MATN4*, *C9H2ORF82* (also known as

289   *SNORC* in mammals), and *ACAN*. Based on additional marker genes and GO-term enrichment

290   analyses, we inferred the Turquoise module to be related to early chondrocyte proliferation

291   and growth, whereas the Red module reflected chondrocyte maturation and extracellular

292   matrix deposition (Fig. 6f). Interestingly, compared to module Turquoise, the activity of

293   module Red was generally more restricted and specifically excluded from sub-cluster HH29-

294   17 (Fig. 6g,h). Upon closer inspection, we identified high expression of several known

295   synovial joint markers genes in this population, thus identifying it as the forming

296   interphalangeal joints (Fig. 6i, Additional file 3).

297   Hence, through a combination of differential gene expression and GO-term enrichment

298   analyses, as well as gene co-expression modules, we identified spatially and/or temporally

299   distinct sub-populations and transcriptome dynamics in the skeletal and peri-skeletal tissues of

300   the forming digits.

10

Discussion

Singe-cell tissue decomposition of the developing chicken autopod

Here, using single-cell RNA-sequencing, we present a transcriptomic atlas of the developing chicken limb at cellular resolution. Focusing on the distal and morphologically diverse portion of the limb, the autopod, we sampled over 17,000 single-cell transcriptomes with an average of over 1,000 genes detected in each cell. Within our atlas, we identify all major tissue types that constitute and pattern the embryonic appendage across three developmental time points. Additionally, taking advantage of our cellular and transcriptomic sampling depth, we manage to isolate even minute cell populations like the AER and identify novel marker genes in it. We also distinguish transcriptionally discrete sub-populations within known major tissue types, reflecting distinct spatial locations or cellular states. As such, it demonstrates the power of scRNA-seq to molecularly disentangle cell populations of the developing limb that occur in close spatial or 'lineage' proximity. Historically, such populations have proven notoriously difficult to separate and characterize transcriptionally, using either manual tissue dissection or reporter-gene based cell lineage isolation. To what extent all of our tissue sub-clusters indeed correspond to distinct lineage separations [35], or rather represent the extremes of a molecular continuum that follows the inherently stochastic nature of transcription [36, 37], remains to be addressed in future studies. Regardless, however, our results provide a toolbox of candidate genes to tackle this question in a molecularly comprehensive manner. Furthermore, our data enables a characterization of emerging embryonic cell types based on transcriptional signatures, rather than relying on the definitive morphological and/or functional features of their mature counterparts.

Cell type equivalencies across developmental and evolutionary time

Such molecular classification schemes echo recent conceptual frameworks that aim to categorize 'cell types' across developmental and evolutionary time scales, irrespective of morphology or function [2]. If, however, we consider a 'cell type' to be primarily defined by the expression of distinct regulatory programs, then detection of program activities can substantially precede our ability to distinguish morphological or functional specializations. Indeed, our sub-clustering and module analyses across developmental time reveal the appearance of certain prospective cell types long before they become morphologically distinct. For example, already at stage HH25 we recover clear gene expression signatures reminiscent of the future periskeletal nsCT, even though prominent cartilage anlagen have yet to form (Fig. 2d, Fig. 3b). As such, it suggests an early lineage priming, without necessarily implying a definite switch in cell fate or clear morphological distinctions. In agreement with

335 this, our *ZFHX3*-containing module Darkgreen appears to be the most basic and least specific

336 of the co-expression modules that coincide with the nsCT population. We detect its activity at

337 all three time points, marking the prospective nsCT as well as parts of the *PRRX1*-positive

338 mesenchymal progenitor population (Fig. 5c,f). Only later do more mature and restricted

339 nsCT sub-divisions and their corresponding co-expression modules occur, as exemplified by

340 the activity of module Darkgrey and some of its members known to be involved in the

341 formation of periskeletal tissues and tendon attachment sites (Fig. 5a,d) [38, 39].

342 Moreover, combining such transcriptome-based 'cell type' classification schemes with

343 comparative scRNA-seq datasets allows for a molecular assessment of homologous cell types

344 between species, across evolutionary time scales [40, 41]. This has important implications

345 when trying to elucidate the impact of cell type-specifying gene regulatory networks on

346 pattern formation and diversification at its relevant cellular scale. Namely, how progenitor

347 populations exactly perceive and process patterning-relevant cues can be modulated by

348 species-specific alterations in the respective cell type-specifying networks. In this context, it

349 is worth noting that we detect *RSPO3* as one of the main markers of the chicken AER (Fig.

350 2d, Additional file 2). R-spondins, a family of secreted ligands involved in WNT-signaling,

351 have previously been implicated in AER maintenance and control of limb outgrowth.

352 However, in mammals only *RSPO2,* and not *RSPO3*, seems to be implicated in AER function

353 [42–44]. Similarly, species-specific modifications in the gene regulatory networks driving

354 skeletal cell type maturation have been reported [45, 46]. Together with recent scRNA-seq

355 studies in other vertebrate model organisms [30, 47, 48], our dataset now opens new avenues

356 for a comprehensive assessment of molecular similarities and divergences in patterning-

357 relevant cell populations of the developing limb, across all major tetrapod clades.

358 **Digit growth and patterning at cellular resolution**

359 Variations in digit number, size and individual digit patterns in the autopod skeletal structure

360 reflect functional specialization of tetrapod hands and feet. During development,

361 condensations of mesenchymal cells first give rise to early skeletogenic progenitors, to then

362 differentiate into distinct skeletal lineages such as chondrocytes, osteocytes or synovial joint

363 cells [49–51]. However, unlike for skeletal elements at more proximal locations of the limb,

364 individual phalanx condensations are sequentially added and expanded at the distal tip of each

365 forming digit, through proliferation of an evolutionary conserved progenitor population [22,

366 23, 52]. Hence, identifying regulators of growth rates, as well as for the relative temporal

367 sequence at which the different skeletal cell types are specified, becomes paramount when

368 trying to understand digit-specific phalanx patterns [25, 53].

369 Early autopod outgrowth, and later digit elongation, is controlled through complex signaling
370 interactions at the distal margin of the limb, involving the concerted action of FGFs, BMPs
371 and WNTs [reviewed 5]. Coinciding with this distal domain, we identify a distinct sub-
372 population of mesenchymal cell types in all of our samples, marked by elevated activity of
373 module Magenta with *TFAP2B*, *WNT5A* and high BMP signaling (Fig. 3c-f). Certain module
374 members have been functionally implied in regulating autopod growth and digit elongation
375 [24, 54–56], yet others remain completely unexplored in this context.

376 Moreover, we identify distinct sub-populations of interdigit mesenchyme cells in our HH29
377 and HH31 samples, with four associated gene co-expression modules (Fig. 4a-h). Module
378 Olivegreen contains SNAI and ID genes, known to be expressed in interdigits, and likely
379 relates to the various BMP-driven processes in this tissue [57–62]. On the other hand, module
380 Orange is dominated by *RDH10*, implicated in mouse interdigital apoptosis [63]. Before its
381 apoptotic disappearance at later stages of development, interdigit mesenchyme is known to
382 instruct the specific phalanx-formulas of its anteriorly adjacent digit [24, 25]. Moreover, we
383 manage to spatially attribute a distinct co-expression module (Midnightblue) to interdigit 4,
384 i.e. posterior to a digit with known regulatory individualization in tetrapods [64].

385 Finally, across all developmental time points we sampled, we identify skeletogenic cell
386 populations. At those stages, the forming skeletal elements still consist exclusively of early
387 progenitors, maturing chondrocytes, and developing synovial joints. Accordingly, we only
388 find three distinct sub-populations, associated with two co-expression modules. Module Red
389 shows enrichment for many canonical markers of chondrocyte maturation (Fig. 6e) [45, 51].
390 On the other hand, genes in module Turquoise do not, for the most part, evoke a classical
391 chondrogenic transcriptional profile (Fig. 6d). Again, this module might rather reflect an early
392 transcriptional priming, only this time towards the skeletogenic lineage. In agreement with
393 this, we only detect low expression levels for the canonical early skeletogenic marker *SOX9* in
394 HH25 sub-cluster 4 (Fig. 2d), which itself is specifically enriched for Turquoise activity.
395 Likewise, our synovial joint-like HH29 sub-cluster 17 shows high activity for Turquoise,
396 while excluding the more mature chondrocyte module Red (Fig. 6g-i).

Conclusion

398 Our single-cell transcriptomic atlas provides a comprehensive genomics resource to study
399 chicken limb development in unprecedented detail. Thereby, it complements a classical
400 experimental model of vertebrate pattern formation with molecular data at cellular resolution.
401 We curate molecular catalogues to provide an in-depth description of the embryonic autopod,
402 through the assembly of cell population-specific lists of candidate marker genes. Combined

13

403 with the power of viral overexpression screens and recent CRISPR/*Cas9* genome

404 modifications technologies, this resource will provide a roadmap for the functional

405 elucidation of cell type specification programs in patterning-relevant populations. Moreover,

406 by constructing cell population-specific gene co-expression modules, we provide a tool to

407 follow tissue dynamics across developmental and evolutionary time scales. Thereby, it will

408 enable insights into the molecular underpinnings of homologous cell types across all major

409 tetrapod clades, and their ensuing developmental impact on pattern formation and

410 diversification in the vertebrate autopod.

Methods

Tissue sampling

We collected tissue samples from embryonic hind limbs at different developmental stages (Fig. 1,a-c). Limbs were dissected in cold PBS, and chopped coarsely with a razorblade. Dissociation into single cells was done using 2.5% trypsin in DMEM and incubation for 15 minutes at 37°. Occasional mechanical shearing by careful pipetting was applied during the incubation time.

scRNA-seq library preparation

Single-cell suspensions of samples HH25 and HH31 were fed into a *10X Genomics Chromium* Single Cell System (*10X Genomics*, Pleasanton, CA, USA) aiming for a concentration of 4000 cells per microliter. Cell capture, cDNA generation, preamplification and library preparation were done using *Chromium Single Cell 3' v2* Reagent Kit according to the manufacturer instructions. For stage HH29 the cells were processed with the *DropSeq* method according to the original protocol [26]. Once the cDNA was obtained from all the samples, the sequencing proceeded on Illumina *NextSeq 500* platforms as recommended by the developers at 75bp and an average depth of 400 million reads per sample.

Data processing

Using either the *Cell Ranger* software v2 (*10X Genomics*) or the *DropSeq* pipeline v1 (https://github.com/broadinstitute/Drop-seq/releases) we performed base calling, adaptor trimming, mapping to the chicken ENSEMBL genome assembly and annotation Gallus_gallus-5.0 [65], de-multiplexing of the sequences and generation of the gene / cell count matrices.

Filtering thresholds for mapped data were adapted for each sample, depending on the different library complexities. Cells with an UMI count of more than 4 times the sample mean or less than 20% of the sample median were filtered out, cells with a mitochondrial or ribosomal contribution to UMI count of more than 10% were also filtered out. Using the R package Seurat v2.3.2 [66] the UMI counts were then Log-normalized and any variation due to the library size or mitochondrial UMI counts percentage was then regressed via a variance correction using the function ScaleData.

The cell cycle stage of each cell was inferred using the R package SCRAN [67] and gene pairs that covariate with cell cycle stages in mouse [68]. The gene pairs were translated to orthologous chicken genes [69] and a cell cycle stage score was obtained cell-wise for stages S, G1 and G2/M, the difference between the G2/M and S scores (δG2M/S) was calculated to be accounted for in later steps.

Feregrino et al.

Dimensionality reduction and visualization

Significant principal components were determined for each sample as those falling outside of a Marchenko-Pastur distribution [35]. A dimensionality reduction step was carried out, using the t-SNE algorithm [28] to visualize the data and clustering of the cells based on transcriptomic similarities. The cells were clustered using the Louvain method for community detection from large networks and the Jaccard similarity coefficient to compare similarity and diversity of the sets, implemented in the FindClusters function in Seurat using data which was additionally variance-corrected for δG2M/S. A first, broad, clustering step was done using a resolution of 0.4 for samples HH31 and HH29 and 0.5 for HH25; a second clustering was done to find sub-clusters within the data, this time using resolutions of 1.4 and 1.1 for the corresponding samples. All clustering steps were done using a k number of 20 and the significant principal components of the sample.

Differential expression analysis

Differential expression analyses based on the negative binomial distribution were performed with Seurat, using the δG2M/S as a covariate and only genes expressed in at least 15% of any compared population (Additional files 2-4); genes expressed in at least 25% of the cells and showing differences with a log fold-change > 0.5 and an adjusted p value < 0.05 were used for GO analyses. To find expression signatures for every cell cluster, in a first step, a phylogenetic tree was obtained for the cell clusters in each sample; all directly paired clusters were tested for differential expression. Any pair of clusters with less than 15 differentially expressed genes were collapsed recursively. In a second step, specific genes for each cluster were obtained contrasting each cluster against the rest of the cells in their sample. To find genes differentially expressed genes between the interdigit clusters (Fig. 4j), we compared each of the sub-clusters against the rest of the cells in the other two clusters.

Marker genes for digit/interdigit 3 and 4 were defined using the DESeq2 R package v1.20.0 [70]. We analyzed bulk RNA data sets of digit/interdigt 3 and 4 from stage HH28/29 and HH31 of a previous study [34]. After normalization based on size factors and dispersion, we performed the differential expression analysis using a Wald test and the contrast design ~Stage+Digit to use the different stages as pseudo-replicates of the digit. We filtered for differential expression with a p-value < 0.05. For visualization, we subtracted the fold changes of early and late stages and plotted a heatmap using heatmap3 R package v1.1.1 [71] using hierarchical clustering of the genes.

Weighted co-expression analyses

478   A weighted correlation network analysis was done using the WGCNA R package v1.6.6 [32].
479   Using the function FindVariableGenes from Seurat, we calculated the genes with high
480   variation (dispersion > 0.5) across all the cells in sample HH29, and were subsequently used
481   in WGCNA. Adjacencies and signed topological overlaps were calculated with an inferred
482   soft-thresholding power of 8. A hierarchical tree was constructed using the "average" method
483   and then cut using the "tree" method at height 0.9957 and minimum module size of 15. The
484   eigengenes of the resulting modules, as well as the membership and a Correlation Student p
485   value of the membership of each gene to its module were calculated. All genes not
486   significantly (p value >0.01) correlated with any module were discarded. The process was
487   repeated recursively, until all genes were significantly associated with a module; the only
488   change made in every iteration was the module minimum size, set to the smallest that would
489   yield at least the same number of modules as the first analysis.

490   The output of WGCNA was exported to the Cytoscape v3.7.0 software [72] where the node
491   size was coded to represent the membership, and the edge thickness and color intensity to
492   represent the weights of each gene-pair coexpression. For visualization purposes, the scales of
493   thickness, color and size were made relative to the minima and maxima found in each
494   network. Furthermore, a transparency gradient was added to the edges, which was scaled to
495   hide unimportant edges and avoid edge saturation, the threshold was always adjusted to make
496   visible at least one edge per node. In only one case (module midnightblue), an edge with an
497   outlier weight was coded to be red and thicker than any other edge, and the color/size re-
498   scaled to the second highest weight.

499   **Gene Ontology**

500   Gene Ontology analyses were conducted with the R package *limma* [73]. We used the list of
501   genes in the expression signature of each computed cell cluster, and the genes members of
502   each co-expression module as input. For each case we used all the genes detected in the
503   corresponding sample as the contrast universe.

504   ***In situ* hybridization**

505   Probes for *CRABP-I* and *COL1A2* were described previously [38]. Primers for the *ZFHX3*
506   probe were designed using primer3 [74]. An AA overhang and an *EcoRI* restriction site were
507   added to each of the primers at the 5' end. ZFHX3 (fw: [5'-
508   AAGAATTCAGCCGTACCGGGTGCAATGAGC-3'], rev: [5'-
509   AAGAATTCAGCGCTTCCTCTTCCCGTAGAGC-3']). *In situ* hybridization was
510   performed using standard protocols [75]

511   **Abbreviations**

Feregrino et al.

512   **EvoDevo:** Evolutionary developmental biology

513   **LPM:** Lateral plate mesoderm

514   **AER:** Apical ectodermal ridge

515   **HH:** Hamburger-Hamilton stages

516   **UMIs:** Unique molecular identifiers

517   **tSNE:** t-distributed stochastic neighbor embedding

518   **GO:** Gene ontology

519   **nsCT:** Non-skeletal connective tissue

520   **TFs:** Transcription factors

521   **scRNA-seq:** Single-cell RNA sequencing

Declarations

Funding

Work in the Tschopp laboratory is supported by the Swiss National Science Foundation (SNSF project grant 31003A_170022), the University of Basel and the *Forschungsfonds* of the University of Basel. These funding bodies had no role in the design of the study, collection, analysis, and interpretation of data, and in writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. Raw sequencing data has been deposited in the SRA (accession numbers: TBD)

Authors' contributions

PT conceived and designed the study. CF, OP and PT conducted the scRNA-seq experiments. CF conducted data analyses and *in situ* experiments. CF and FS conducted the bulk RNA-seq re-analysis. CF and PT drafted the manuscript. All of the authors read and approved the final manuscript.

Ethics approval and consent to participate

In accordance with Swiss national guidelines (Swiss Animal Protection Ordinance; TSchV, chapter 6, Art. 112), no formal ethics approval was required, as all experiments were carried out prior to the third trimester of incubation.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

Acknowledgements

Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing center at University of Basel. CF and PT wish to acknowledge Katja Eschbach and Christian Beisel for help with *10X Genomics Chromium* and sequencing. OP and PT thank Tyler Burks for help with *DropSeq* experiments. PT and OP would like to acknowledge the generous support of Cliff Tabin and Aviv Regev, in whose labs this project was initiated (with help of NIH grant HD03443 to Cliff Tabin).

Bibliography

553   1. Stathopoulos A, Levine M. Genomic regulatory networks and animal development.

554   Developmental Cell. 2005.

556   2. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M,

557   Schlosser G, Widder S, Laubichler MD, Wagner GP. The origin and evolution of cell types.

558   Nat Rev Genet. 2016;17:744–57.

559   3. Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic

560   landscapes. Nature Reviews Genetics. 2016.

561   4. Eldar A, Dorfman R, Weiss D, Ashe H, Shilo DZ, Barkal N. Robustness of the BMP

562   morphogen gradient in Drosophila embryonic patterning. Nature. 2002.

563   5. Zeller R, Ríos JL, Zuniga A. Vertebrate limb bud development : moving towards

564   integrative analysis of organogenesis. Nat Rev Genet. 2009;10:845–58.

565   6. Perrimon N, Pitsouli C, Shilo B-Z. Signaling Mechanisms Controlling Cell Fate and

566   Embryonic Patterning. Cold Spring Harb Perspect Biol. 2012;4:a005975–a005975.

567   7. Biesecker LG. Polydactyly: How many disorders and how many genes? 2010 update.

568   Developmental Dynamics. 2011.

569   8. Stricker S, Mundlos S. Mechanisms of digit formation: Human malformation syndromes

570   tell the story. Developmental Dynamics. 2011.

571   9. Zuniga A, Zeller R, Probst S. The molecular basis of human congenital limb

572   malformations. Wiley Interdisciplinary Reviews: Developmental Biology. 2012.

573   10. Petit F, Sears KE, Ahituv N. Limb development: A paradigm of gene regulation. Nat Rev

574   Genet. 2017;18:245–58.

575   11. Fedak TJ, Hall BK. Perspectives on hyperphalangy: Patterns and processes. Journal of

576   Anatomy. 2004.

577   12. Sears KE, Behringer RR, Rasweiler JJ, Niswander LA. Development of bat flight:

578   Morphologic and molecular evolution of bat wing digits. Proc Natl Acad Sci. 2006.

579   13. De Bakker MAG, Fowler DA, Oude K Den, Dondorp EM, Carmen Garrido Navas M,

580   Horbanczuk JO, Sire JY, Szczerbińska D, Richardson MK. Digit loss in archosaur evolution

581   and the interplay between selection and constraints. Nature. 2013.

582   14. Cooper KL, Sears KE, Uygur A, Maier J, Baczkowski KS, Brosnahan M, Antczak D,

583   Skidmore JA, Tabin CJ. Patterning and post-patterning modes of evolutionary digit loss in

584   mammals. Nature. 2014.

585   15. Wachtler F, Christ B, Jacob HJ. On the determination of mesodermal tissues in the avian

586   embryonic wing bud. Anat Embryol (Berl). 1981.

587   16. Logan M, Martin JF, Nagy A, Lobe C, Olson EN, Tabin CJ. Expression of Cre

588   Recombinase in the developing mouse limb bud driven by a Prxl enhancer. Genesis. 2002.

589   17. Pearse R V., Scherz PJ, Campbell JK, Tabin CJ. A cellular lineage analysis of the chick

590   limb bud. Dev Biol. 2007.

591   18. Chevallier A, Kieny M, Mauger A. Limb-somite relationship: origin of the limb

592   musculature. J Embryol Exp Morphol. 1977.

593   19. Christ B, Jacob HJ, Jacob M. Experimental analysis of the origin of the wing musculature

594   in avian embryos. Anat Embryol (Berl). 1977.

595   20. Riddle RD, Johnson RL, Laufer E, Tabin C. Sonic hedgehog mediates the polarizing

596   activity of the ZPA. Cell. 1993.

597   21. Lopez-Rios J, Duchesne A, Speziale D, Andrey G, Peterson KA, Germann P, et al.

598   Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. Nature. 2014.

599   22. Montero JA, Lorda-Diez CI, Gañan Y, Macias D, Hurle JM. Activin/TGFβ and BMP

600   crosstalk determines digit chondrogenesis. Dev Biol. 2008.

601   23. Suzuki T, Hasso SM, Fallon JF. Unique SMAD1/5/8 activity at the phalanx-forming

602   region determines digit identity. Proc Natl Acad Sci U S A. 2008;105:4185–90.

603   24. Dahn RD, Fallon JF. Interdigital regulation of digit identity and homeotic transformation

604   by modulated BMP signaling. Science (80-). 2000.

605   25. Huang BL, Trofka A, Furusawa A, Norrie JL, Rabinowitz AH, Vokes SA, Taketo MM,

606   Zakany J, Mackem S. An interdigit signalling centre instructs coordinate phalanx-joint

607   formation governed by 5'Hoxd-Gli3 antagonism. Nat Commun. 2016;7:1–10.

608   26. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR,

609   Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A,

610   McCarroll SA. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using

611   Nanoliter Droplets. Cell. 2015;161:1202–14.

612   27. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively

613   parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:1–12.

614   28. Van Der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-sne. J Mach

615   Learn Res. 2008;9:2579–605.

616   29. Cserjesi P, Lilly B, Bryson L, Wang Y, Sassoon DA, Olson EN. MHox: a mesodermally

617   restricted homeodomain protein that binds an essential site in the muscle creatine kinase

618   enhancer. Development. 1992.

619   30. Gerber T, Murawala P, Knapp D, Masselink W, Schuez M, Hermann S, Gac-Santel M,

620   Nowoshilow S, Kageyama J, Khattak S, Currie JD, Camp JG, Tanaka EM, Treutlein B.

621   Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration.

622   Science (80-). 2018;362.

623   31. Bravo R, Frank R, Blundell PA, Macdonald-Bravo H. Cyclin/PCNA is the auxiliary

624   protein of DNA polymerase-δ. Nature. 1987.

625   32. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network

626   analysis. BMC Bioinformatics. 2008;9.

627   33. Dupé V, Ghyselinck NB, Thomazy V, Nagy L, Davies PJA, Chambon P, Mark M.

628   Essential roles of retinoic acid signaling in interdigital apoptosis and control of BMP-7

629   expression in mouse autopods. Dev Biol. 1999.

630   34. Wang Z, Young RL, Xue H, Wagner GP. Transcriptomic analysis of avian digits reveals

631   conserved and derived digit identities in birds. Nature. 2011;477:583–7.

632   35. Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X,

633   Levin JZ, Nemesh J, Goldman M, McCarroll SA, Cepko CL, Regev A, Sanes JR.

634   Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics.

635   Cell. 2016;166:1308–1323.e30.

636   36. Nicolas D, Phillips NE, Naef F. What shapes eukaryotic transcriptional bursting?

637   Molecular BioSystems. 2017.

638   37. Urban EA, Johnston RJ. Buffering and Amplifying Transcriptional Noise During Cell

639   Fate Specification. Front Genet. 2018;9 November:1–14.

640   38. Bandyopadhyay A, Kubilus JK, Crochiere ML, Linsenmayer TF, Tabin CJ. Identification

641   of unique molecular subdomains in the perichondrium and periosteum and their role in

642   regulating gene expression in the underlying chondrocytes. Dev Biol. 2008.

643   39. Witte F, Dokas J, Neuendorf F, Mundlos S, Stricker S. Comprehensive expression

644   analysis of all Wnt genes and their major secreted antagonists during mouse limb

645   development and cartilage differentiation. Gene Expr Patterns. 2009.

646   40. Marioni JC, Arendt D. How Single-Cell Genomics Is Changing Evolutionary and

647   Developmental Biology. Annu Rev Cell Dev Biol. 2017;33:537–53.

648   41. Tschopp P, Tabin CJ. Deep homology in the age of next-generation sequencing.

649   Philosophical Transactions of the Royal Society B: Biological Sciences. 2017.

650   42. Nam JS, Park E, Turcotte TJ, Palencia S, Zhan X, Lee J, Yun K, Funk WD, Yoon JK.

651   Mouse R-spondin2 is required for apical ectodermal ridge maintenance in the hindlimb. Dev

652   Biol. 2007.

653   43. Neufeld S, Rosin JM, Ambasta A, Hui K, Shaneman V, Crowder R, Vickerman L, Cobb J.

654   A conditional allele of Rspo3 reveals redundant function of R-spondins during mouse limb

655   development. Genesis. 2012.

656   44. Szenker-Ravi E, Altunoglu U, Leushacke M, Bosso-Lefèvre C, Khatoo M, Thi Tran H, et

657   al. RSPO2 inhibition of RNF43 and ZNRF3 governs limb development independently of

658   LGR4/5/6. Nature. 2018.

659   45. Gómez-Picos P, Eames BF. On the evolutionary relationship between chondrocytes and

660   osteoblasts. Frontiers in Genetics. 2015.

661   46. Ferguson GB, Van Handel B, Bay M, Fiziev P, Org T, Lee S, et al. Mapping molecular

662   landmarks of human skeletal ontogeny and pluripotent stem cell-derived articular

663   chondrocytes. Nat Commun. 2018.

664   47. Fabre PJ, Leleu M, Mascrez B, Giudice Q Lo, Cobb J, Duboule D. Single-cell mRNA

665   profiling reveals heterogeneous combinatorial expression of Hoxd genes during limb

666   development. BMC Biol. 2018;:327619.

667   48. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S,

668   Christiansen L, Steemers FJ, Trapnell C, Shendure J. The single-cell transcriptional landscape

669   of mammalian organogenesis. Nature. 2019.

670   49. Hartmann C, Tabin CJ. Wnt-14 plays a pivotal role in inducing synovial joint formation in

671   the developing appendicular skeleton. Cell. 2001.

672   50. Akiyama H, Kim J-E, Nakashima K, Balmes G, Iwai N, Deng JM, Zhang Z, Martin JF,

673   Behringer RR, Nakamura T, de Crombrugghe B. Osteo-chondroprogenitor cells are derived

674   from Sox9 expressing precursors. Proc Natl Acad Sci. 2005.

675   51. Kozhemyakina E, Lassar AB, Zelzer E. A pathway to bone: signaling molecules and

676   transcription factors involved in chondrocyte development and maturation. Development.

677   2015.

678   52. Witte F, Chan D, Economides AN, Mundlos S, Stricker S. Receptor tyrosine kinase-like

679   orphan receptor 2 (ROR2) and Indian hedgehog regulate digit outgrowth mediated by the

680   phalanx-forming region. Proc Natl Acad Sci. 2010.

681   53. Hiscock TW, Tschopp P, Tabin CJ. On the Formation of Digits and Joints during Limb

682   Development. Developmental Cell. 2017.

683   54. Sanz-Ezquerro JJ, Tickle C. Fgf Signaling Controls the Number of Phalanges and Tip

684   Formation in Developing Digits. Curr Biol. 2003.

685   55. Gros J, Hu JKH, Vinegoni C, Feruglio PF, Weissleder R, Tabin CJ. WNT5A/JNK and

686   FGF/MAPK pathways regulate the cellular events shaping the vertebrate limb bud. Curr Biol.

687   2010.

688   56. Seki R, Kitajima K, Matsubara H, Suzuki T, Saito D, Yokoyama H, Tamura K. AP-2β is a

689   transcriptional regulator for determination of digit length in tetrapods. Dev Biol. 2015.

690   57. Jen Y, Manova K, Benezra R. Expression patterns of Id1, Id2, and Id3 are highly related

691   but distinct from that of Id4 during mouse embryogenesis. Dev Dyn. 1996.

692   58. Ros MA, Sefton M, Nieto MA. Slug, a zinc finger gene previously implicated in the early

693   patterning of the mesoderm and the neural crest, is also involved in chick limb development.

694   Development. 1997.

695   59. Nieto MA. The snail superfamily of zinc-finger transcription factors. Nature Reviews

696   Molecular Cell Biology. 2002.

697   60. Zuzarte-Luís V, Hurlé JM. Programmed cell death in the developing limb. Int J Dev Biol.

698   2002.

699   61. Lorda-Diez CI, Torre-Pérez N, García-Porrero JA, Hurle JM, Montero JA. Expression of

700   Id2 in the developing limb is associated with zones of active BMP signaling and marks the

701   regions of growth and differentiation of the developing digits. Int J Dev Biol. 2009.

702   62. Pignatti E, Zeller R, Zuniga A. To BMP or not to BMP during vertebrate limb bud

703   development. Seminars in Cell and Developmental Biology. 2014.

704   63. Cunningham TJ, Chatzi C, Sandell LL, Trainor PA, Duester G. Rdh10 mutants deficient

705   in limb field retinoic acid signaling exhibit normal limb patterning but display interdigital

706   webbing. Dev Dyn. 2011.

707   64. Stewart TA, Liang C, Cotney J, Noonan JP, Sanger T, Wagner G. Evidence against

708   tetrapod-wide digit identities and for a limited frame shift in bird wings. bioRxiv.

709   2018;:224147.

710   65. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018.

711   Nucleic Acids Res. 2017; December 2017:1–8.

712   66. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell

713   transcriptomic data across different conditions, technologies, and species. Nat Biotechnol.

714   2018.

715   67. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of

716   single-cell RNA-seq data with Bioconductor. F1000Research. 2016;5:2122.

717   68. Scialdone A, Natarajan KN, Saraiva LR, Proserpio V, Teichmann SA, Stegle O, Marioni

718   JC, Buettner F. Computational assignment of cell-cycle stage from single-cell transcriptome

719   data. Methods. 2015;85:54–61.

720   69. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of

721   genomic datasets with the R/ Bioconductor package biomaRt. Nat Protoc. 2009;4:1184–91.

722   70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for

723   RNA-seq data with DESeq2. Genome Biol. 2014.

724   71. Zhao S, Guo Y, Sheng Q, Shyr Y. Heatmap3: an improved heatmap package with more

725   powerful and convenient features. BMC Bioinformatics. 2014.

726   72. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski

727   B, Ideker T. Cytoscape: A software Environment for integrated models of biomolecular

728   interaction networks. Genome Res. 2003.

729   73. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers

730   differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids

731   Res. 2015;43:e47.

732   74. Rozen S, Skaletsky H. Primer3 on the WWW for General Users and for Biologist

733   Programmers In: Krawetz S, Misener S (eds) Methods and protocols: methods in molecular

734   biology. Totowa, NJ Humana Press. 2000.

735   75. McGlinn E, Mansfield JH. Detection of Gene Expression in Mouse Embryos and Tissue

736   Sections. In: Pelegri FJ, editor. Vertebrate Embryogenesis: Embryological, Cellular, and

737   Genetic Methods. Totowa, NJ: Humana Press; 2011. p. 259–92.

Feregrino et al.

Figures
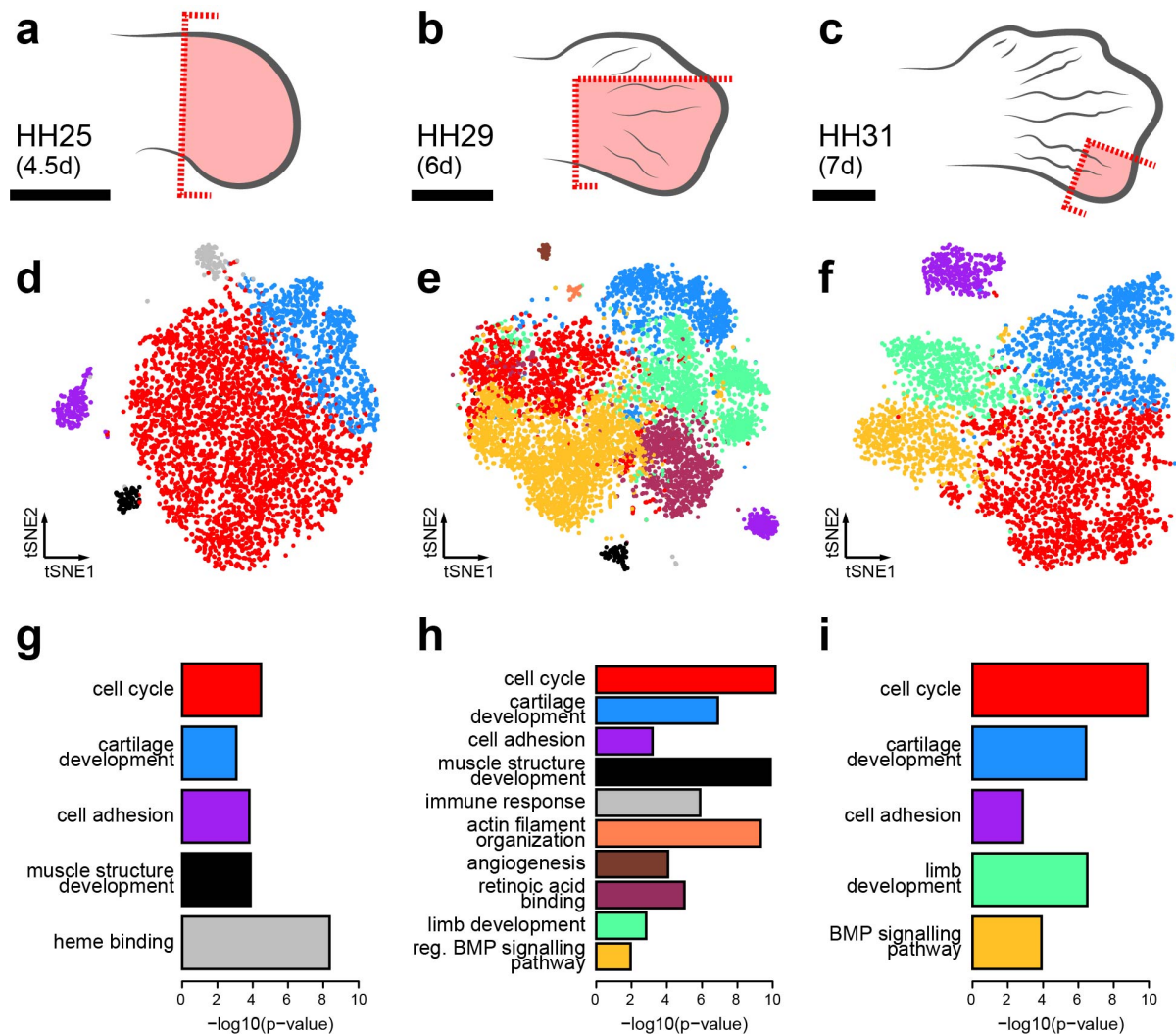
Fig. 1

Sampling strategy and tissue composition of the developing chicken autopod. (a-c) Dissection schemes, highlighted in red, for sampling the different stages of hindlimb development (scale bar ~1mm). **(d-f)** tSNE representation of the three datasets, representing 5,982 (HH25), 6,823 (HH29) and 4,823 (HH31) according to their transcriptome similarities. Cellular color codes reflect unsupervised graph-based clustering results. Comparable cell populations identified in multiple samples are visualized using the same color. **(g-i)** Select overrepresented GO-terms, from analysis of the overexpressed genes, for each cluster at stages **(g)** HH25, **(h)** HH29 and **(i)** HH31.

Feregrino et al.
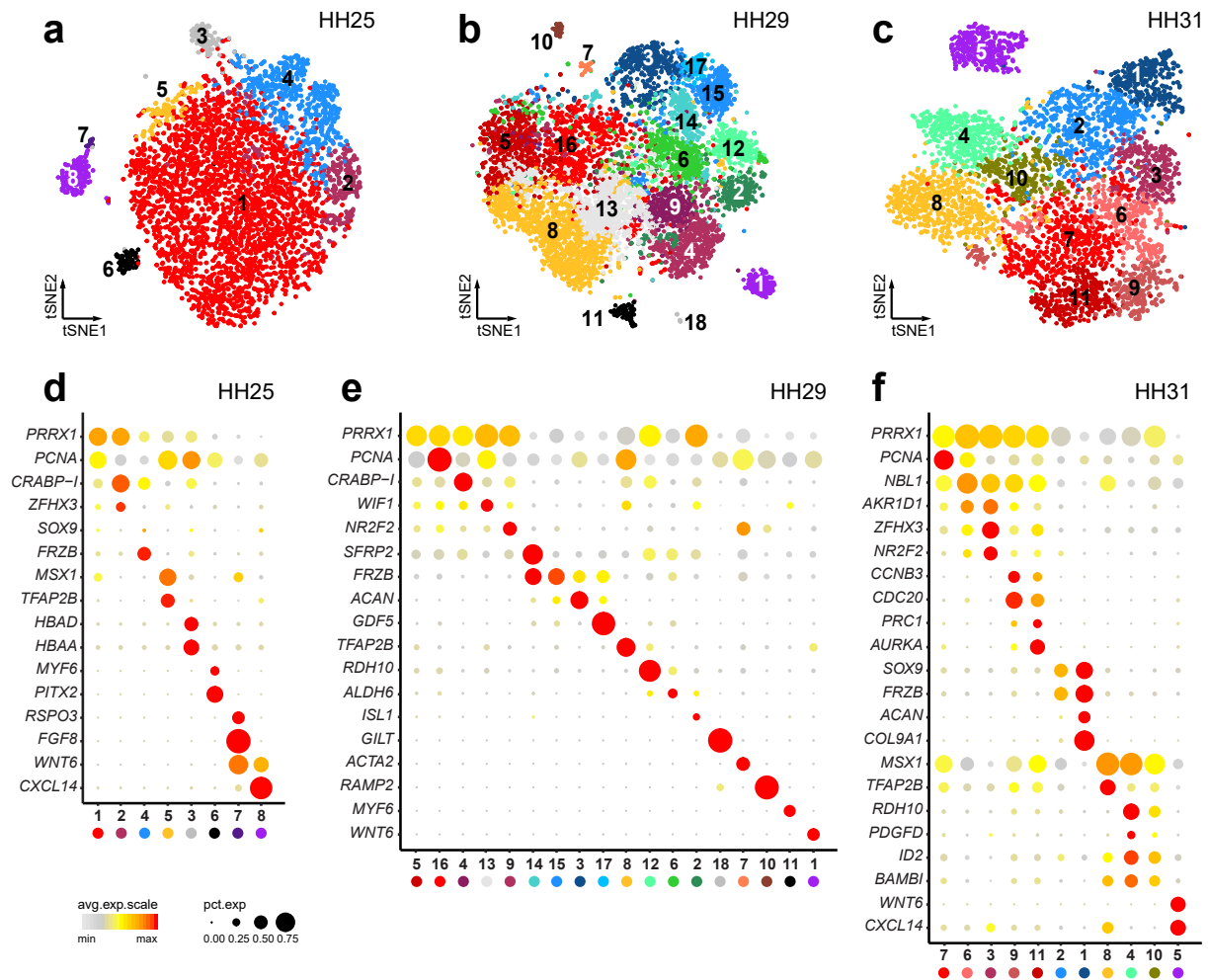
Fig. 2

Cell population sub-structure and marker gene expression (a-c) tSNE plots of the three datasets. Colors now represent fine-tuned unsupervised graph-based clustering, with similar colorations relating to the results of the first clustering step. Comparable cell populations identified in multiple samples are visualized using the same color. For reference, sub-cluster numbers are added. **(d-f)** Dot plots of sub-cluster marker gene expression. Averaged expression level (heatmap) and percentage of cells showing >0 expression (dot size) is visualized across all samples, for all identified sub-clusters. Same color-coding for sub-clusters identification is used as in **(a-c)**.

27

Feregrino et al.
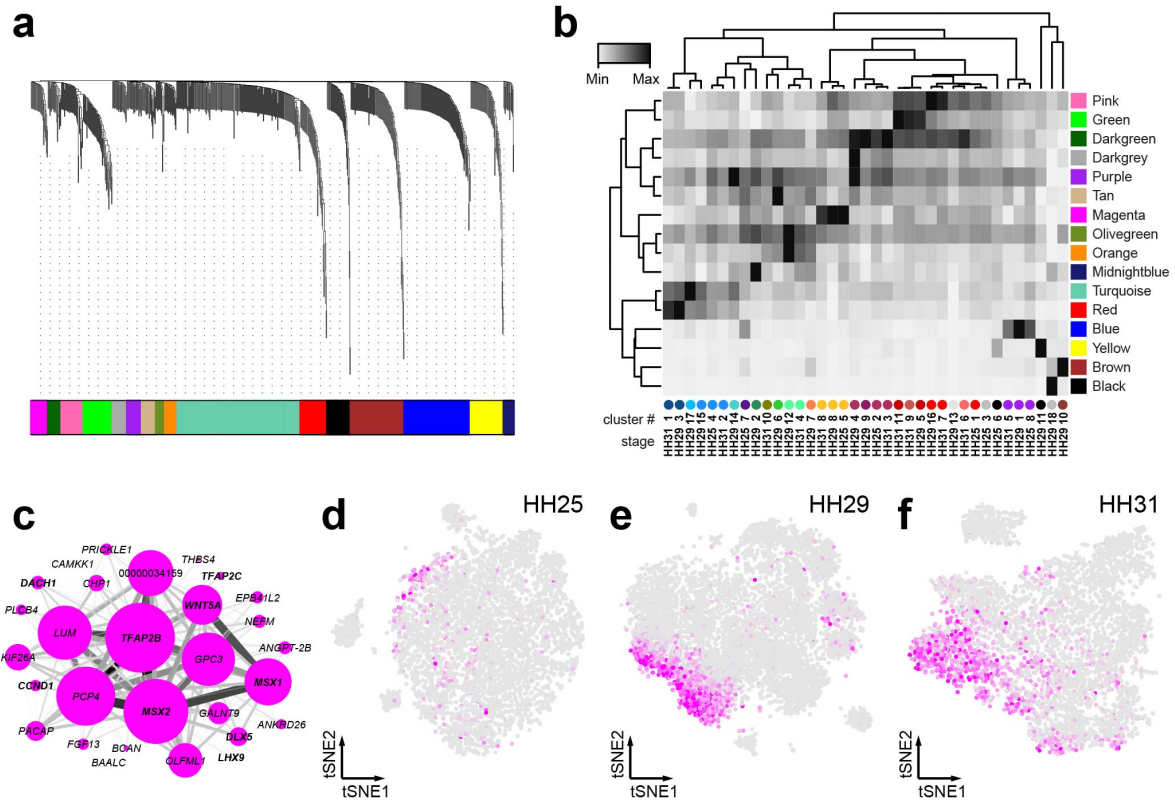
Fig. 3

Weighted correlation network analysis and gene co-expression modules. (a) WGCNA gene hierarchical clustering dendrogram and modules of co-expression. A total of 16 distinct co-expression modules are identified, visualized by colored bars at the bottom of the dendrogram (color scheme unrelated to previous cell clustering). **(b)** Heatmap of mean expression values per co-expression module, calculated across distinct cell sub-clusters and developmental stages. Ordering based on hierarchical clustering of averaged co-expression module activities and sub-clusters. Sub-clusters identification at bottom (number and color code) corresponds to Fig. 2a-c. **(c)** *Cytoscape* visualization of co-expression module Magenta. Node size is proportional to module membership of each gene, edge thickness represents correlation of pair-wise gene co-expression. **(d-f)** Heatmap representing the averaged cellular activity of the Magenta module, plotted on tSNE representations of the different samples. Color intensity is proportional to the mean expression of the module in each cell.

Fig. 4
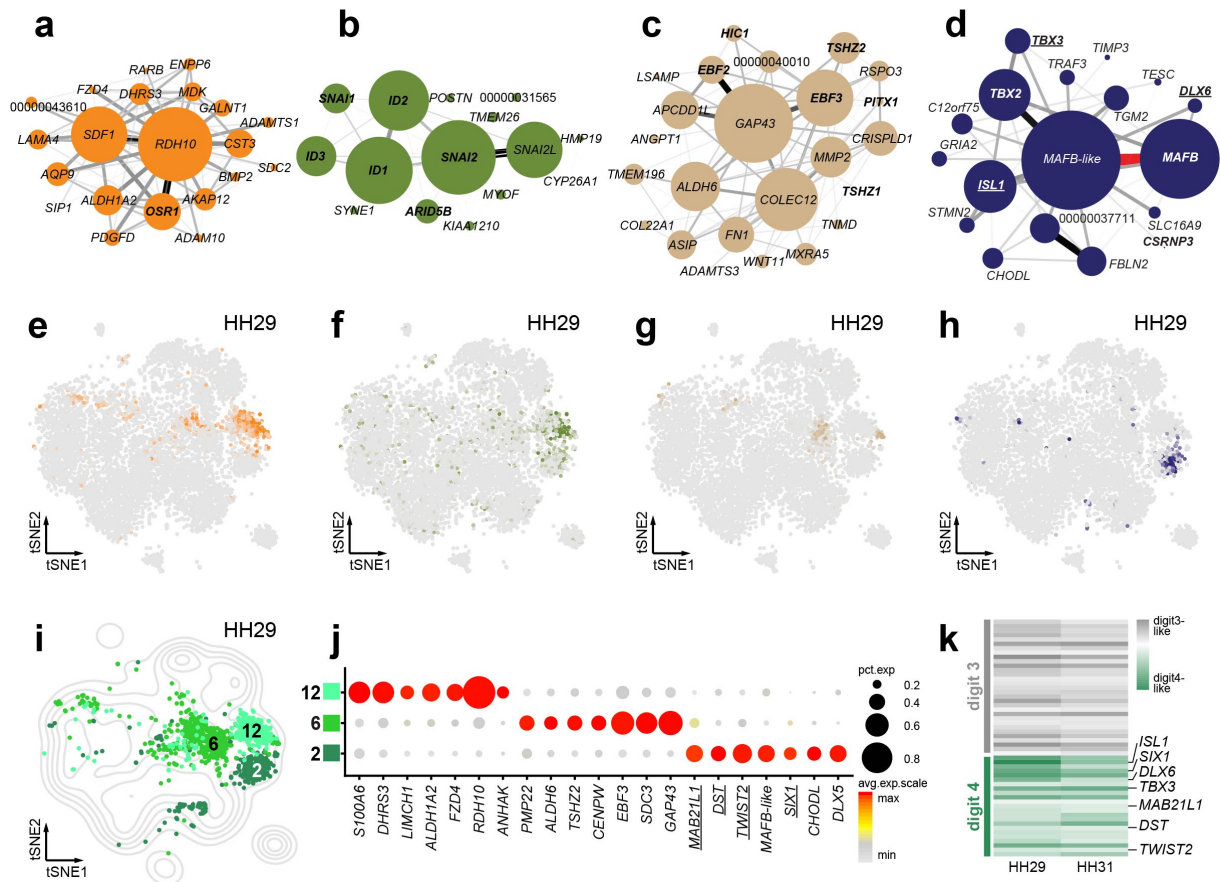
Molecular and spatial heterogeneity in the interdigit mesenchyme. (a-g) Interdigit-associated co-expression modules **(a)** Orange, **(b)** Olivegreen, **(c)** Tan, and **(d)** Midnightblue. Node size represents gene module membership, edge thickness gene pair-wise correlation. Gene names in bold are classified as transcription factors, uncharacterized genes show only Ensembl numbers following the "ENSGALG" gene code. **(e-h)** Heatmaps of averaged activity levels of the corresponding modules, visualized on top of a tSNE plot for sample HH29. **(i)** Contour density plot of the tSNE projection for sample HH29, to delineate overall cell distribution. Partial tSNE plot on top, to visualize only cells belonging to interdigit-like sub-clusters (Color-coding and numbering according to Fig. 2b). **(j)** Expression dot plot of differentially expressed genes between the three interdigit sub-clusters at stage HH29. **(k)** Heatmap visualization of "digit3-like" and "digit4-like" gene sets at stages HH29 and HH31, based on differential expression analysis of digit-specific bulk RNA-seq data by Wang et al., 2011. Underlined gene names in **(d,j)** denote membership to the "digit IV-like" gene set.

29

Feregrino et al.
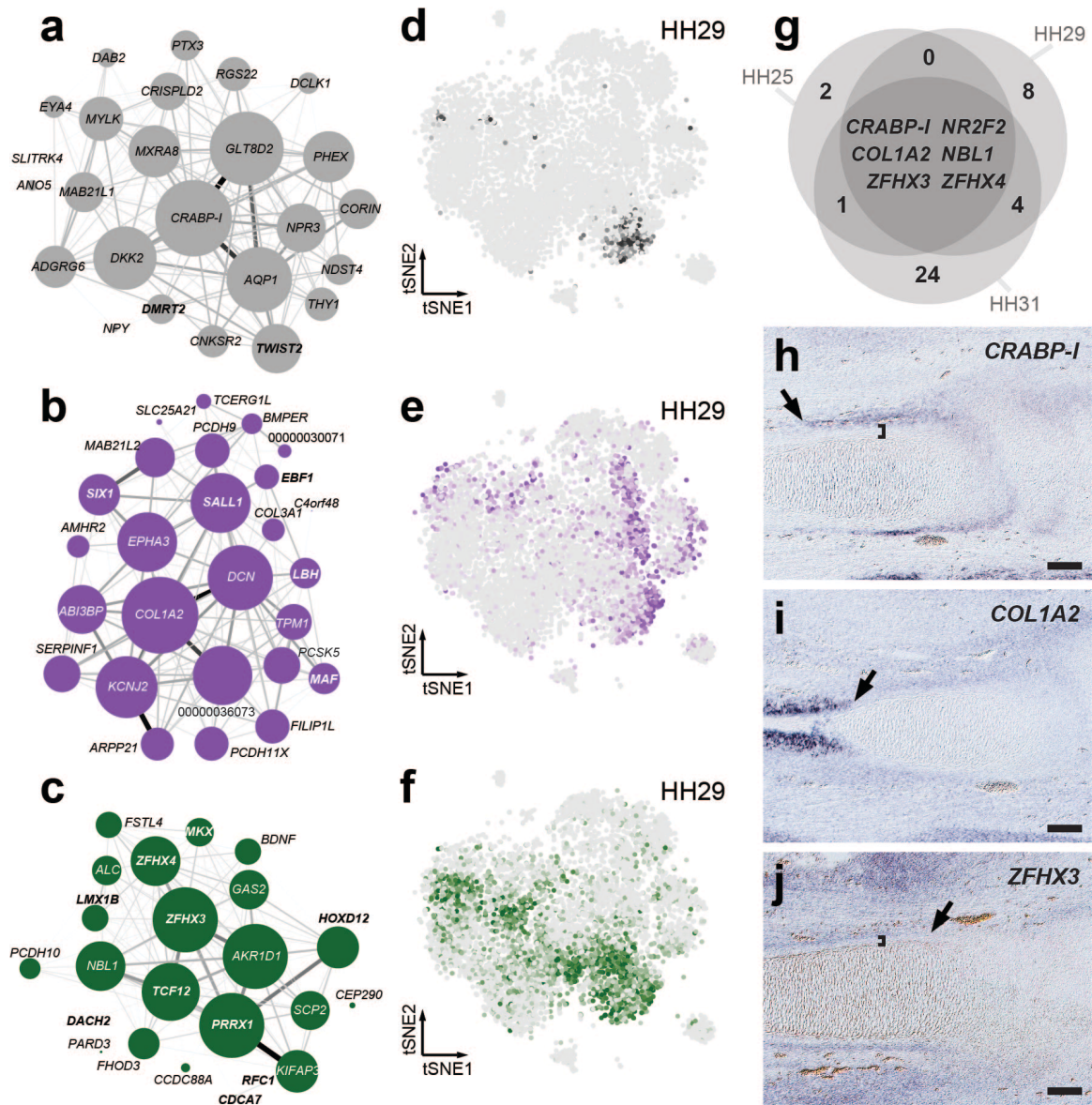
Fig. 5

Transcriptional modules in the non-skeletal connective tissue (nsCT). (a-c) Gene co-expression modules **(a)** Darkgray, **(b)** Purple and **(c)** Darkgreen enriched for peri-skeletal genes. Gene names in bold are classified as transcription factors, uncharacterized genes show only Ensembl numbers following the "ENSGALG" gene code. **(d-f)** Corresponding averaged module activities visualized as heatmaps on stage HH29 tSNE plots. **(d)** Venn diagram of shared overexpressed genes in the nsCT populations of the three samples. **(h-i)** Section *in situ* hybridization on stage HH31 chicken hindlimbs for three shared nsCT marker genes, *CRABP-I*, *COL1A2* and *ZFHX3*. Arrows denote extent of expression along the long bone axis, while brackets indicate separation from the forming skeletal element (scale bar=100mm).

30
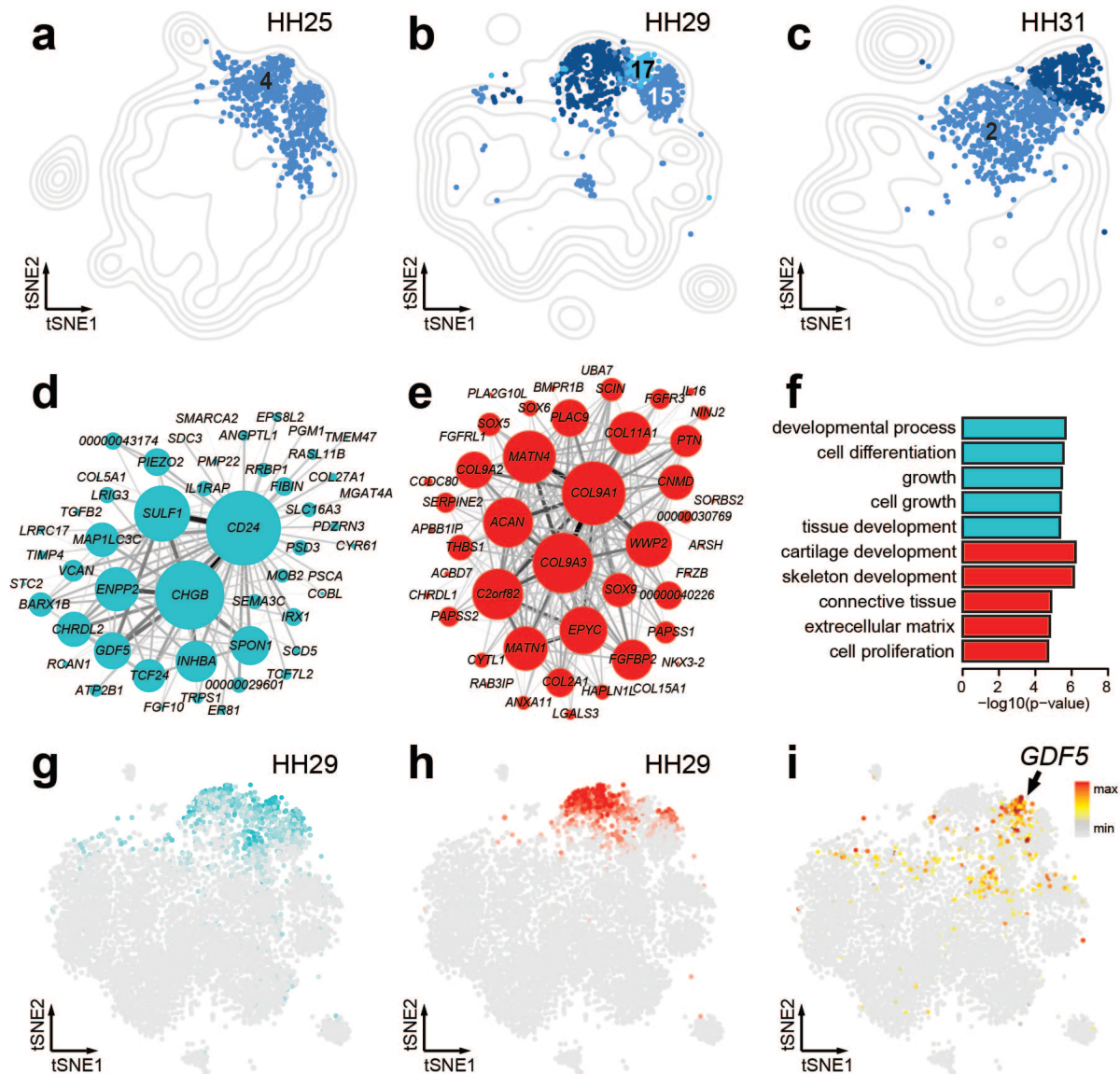
800

801   **Fig. 6**

802   **Transcriptional modules and sub-populations in skeletogenic cells (a-c)** Contour density

803   plot of tSNE projection for each sample. Partial tSNE plot on top, to visualize only cells

804   belonging to skeletogenic sub-clusters (Color-coding and numbering according to Fig. 2b).

805   Same color / shade across samples indicates comparable cell populations. **(d-e)** Gene co-

806   expression modules **(d)** Turquoise and **(e)** Red. Representation of the Turquoise module only

807   shows the 50 genes with the top membership, of a total of 215. **(f)** Top 5 GO-terms, from

808   analysis of the genes member of modules Turquoise and Red. **(g-h)** Averaged module

809   activities visualized as heatmaps on stage HH29 tSNE plots corresponding to the modules

810   Turquoise and Red. **(i)** Expression heatmap of *GDF5* visualized on stage HH29 tSNE.
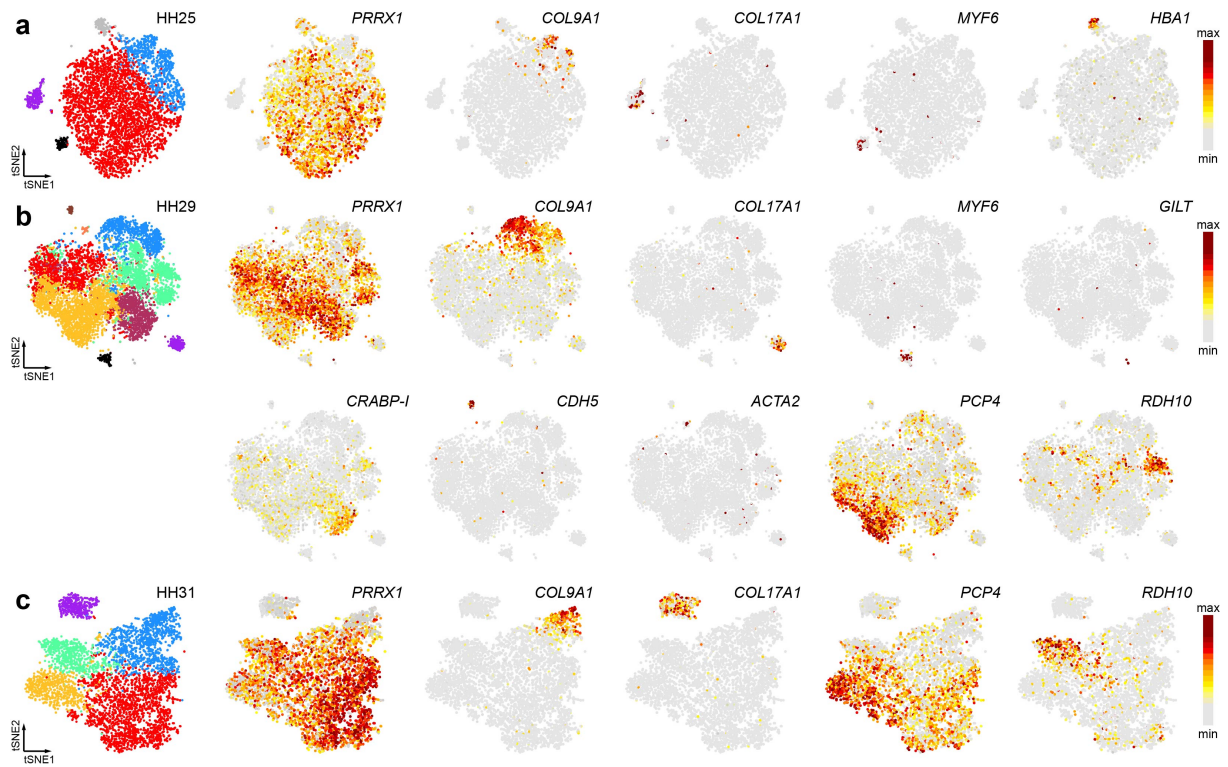
811   **Additional figures**

812
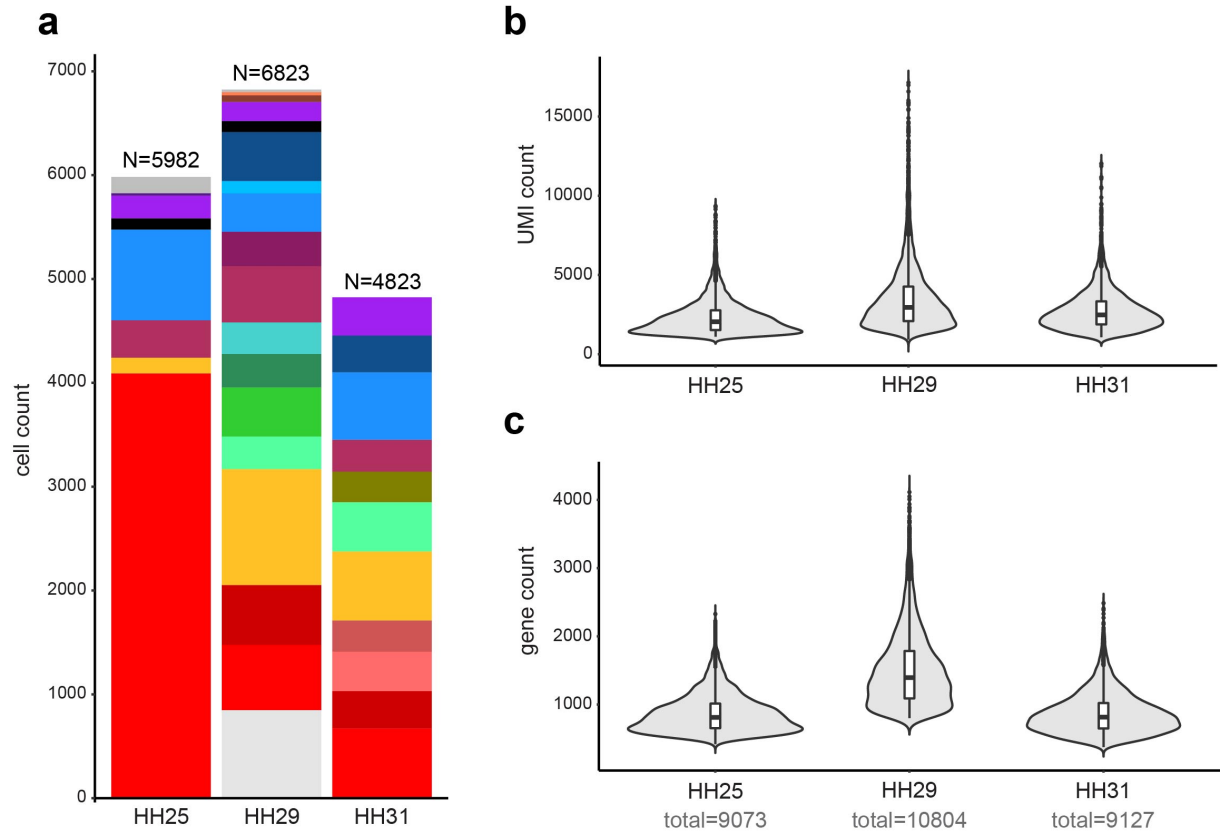
814   **Fig. S1 Sample compositions and data statistics. (a)** Cellular composition of the samples

815   and datasets, color code corresponds to Fig. 2a-c. **(b)** UMI count distributions across the

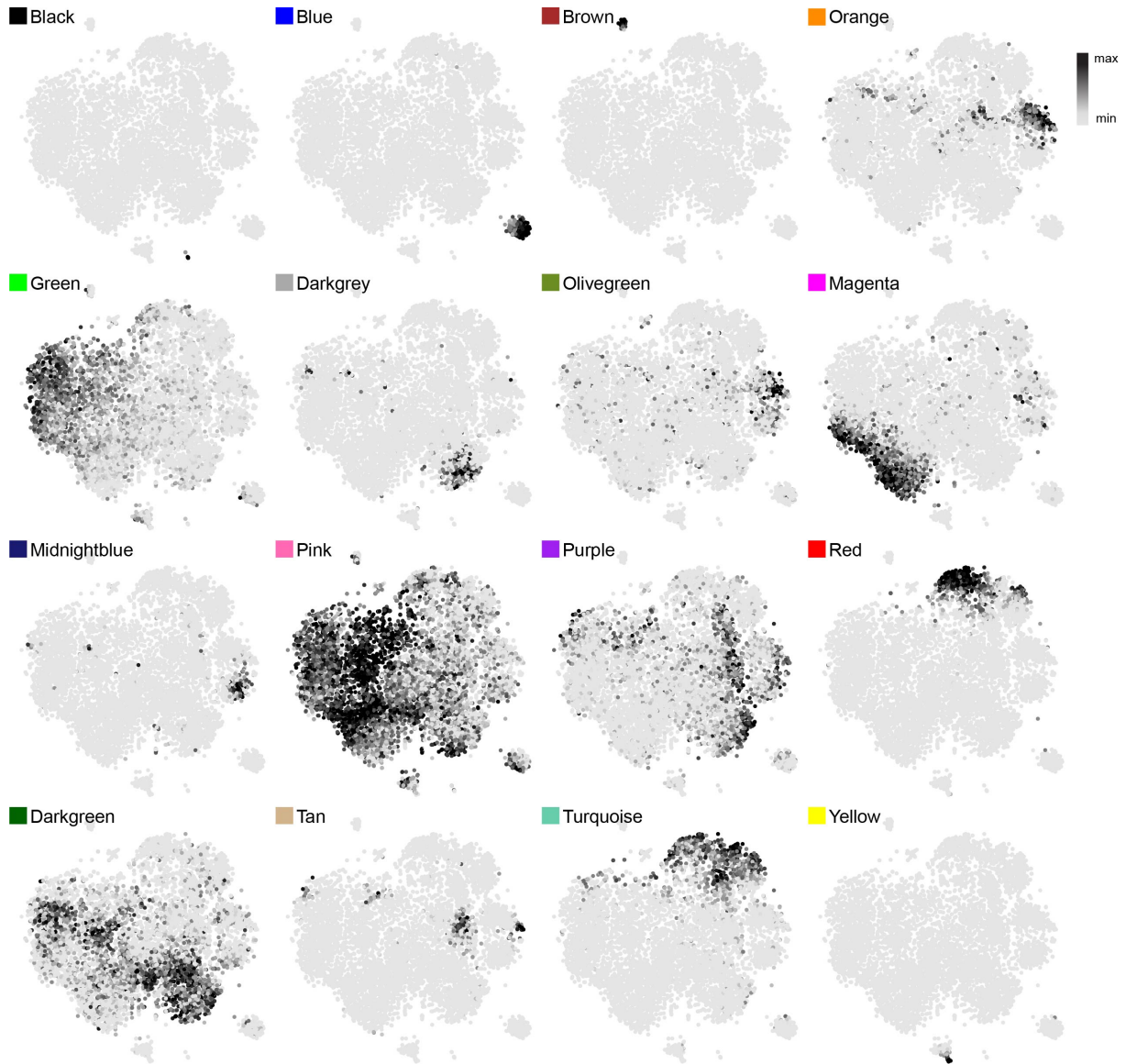816   samples. **(c)** Gene count distributions across the samples.

Feregrino et al.

817

Fig. S2 Expression patterns of marker genes. Related to Fig. 1. Normalized expression patterns of selected genes to identify the different cell populations in our broad clustering, plotted on the tSNEs from sample **(a)** HH25, **(b)** HH29 and **(c)** HH31.

33

Feregrino et al.

821

Fig. S3 Co-expression modules expression patterns. Related to Fig. 3. Average expression
of each WGCNA co-expression module on the tSNE of sample HH29.

34

824 **Additional file 1**

825 **XLSX**

826 **Genes with enriched expression per cell population in sample HH25**

827 Genes enriched in the different cell clusters, calculated to be differentially expressed between

828 each cell cluster and the rest of the cells in the sample. **p_val**: originally calculated p value;

829 **avg_logFC**: average log fold-change relative to the rest of the cells; **pct.x**: percentage of cells

830 in the focus cluster expressing the gene; **pct.rest**: percentage of cells in the rest of the clusters

831 expressing the gene; **p_val_adj**: p value adjusted for multiple testing; cluster: cluster number

832 in the main text and figures; **gene**: ENSEMBL gene identifier; **name**: gene symbol, or name

833 when available; **enrichment**: ratio of pct.x : pct.rest.

834 **Additional file 3**

835 **XLSX**

836 **Genes with enriched expression per cell population in sample HH29**

837 Genes enriched in the different cell clusters, calculated to be differentially expressed between

838 each cell cluster and the rest of the cells in the sample. **p_val**: originally calculated p value;

839 **avg_logFC**: average log fold-change relative to the rest of the cells; **pct.x**: percentage of cells

840 in the focus cluster expressing the gene; **pct.rest**: percentage of cells in the rest of the clusters

841 expressing the gene; **p_val_adj**: p value adjusted for multiple testing; cluster: cluster number

842 in the main text and figures; **gene**: ENSEMBL gene identifier; **name**: gene symbol, or name

843 when available; **enrichment**: ratio of pct.x : pct.rest.

844 **Additional file 4**

845 **XLSX**

846 **Genes with enriched expression per cell population in sample HH31**

847 Genes enriched in the different cell clusters, calculated to be differentially expressed between

848 each cell cluster and the rest of the cells in the sample. **p_val**: originally calculated p value;

849 **avg_logFC**: average log fold-change relative to the rest of the cells; **pct.x**: percentage of cells

850 in the focus cluster expressing the gene; **pct.rest**: percentage of cells in the rest of the clusters

851 expressing the gene; **p_val_adj**: p value adjusted for multiple testing; cluster: cluster number

852 in the main text and figures; **gene**: ENSEMBL gene identifier; **name**: gene symbol, or name

853 when available; **enrichment**: ratio of pct.x : pct.rest.

854 **Additional file 5**

855 **XLSX**

856 **Co-expression modules and their genes**

Feregrino et al.

857   Genes part of the different co-expression modules. **nodeName**: ENSMBL identifier of the

858   genes part of the module; **altName**: gene symbol, or name when available; **membership**:

859   membership to the module.