1 **Genome-wide strategies identify molecular niches regulated by connective tissue-**

2 **associated transcription factors**

3 Mickael Orgeur[1,2,3], Marvin Martens[3], Georgeta Leonte[2,4], Sonya Nassari[3], Marie-Ange Bonnin[3],

4 Stefan T. Börno[2], Bernd Timmermann[2], Jochen Hecht[2,5,6,7], Delphine Duprez[3,*,#] and Sigmar

5 Stricker[1,2,*,#].

6 [1] Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany

7 [2] Max Planck Institute for Molecular Genetics, Berlin, Germany

8 [3] Sorbonne Universités, UPMC Univ. Paris 06, CNRS UMR 7622, Inserm U1156, IBPS-

9 Developmental Biology Laboratory, 75005 Paris, France

10 [4] Institute of Biology, Freie Universität Berlin, Berlin, Germany

11 [5] Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité Universitatsmedizin,

12 Berlin, Germany

13 [6] Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr.

14 Aiguader 88, 08003 Barcelona, Spain

15 [7] Universitat Pompeu Fabra (UPF), Barcelona, Spain

16 [*] Equal contribution as senior authors

17 [#] Corresponding authors: Sigmar Stricker, sigmar.stricker@fu-berlin.de; Delphine Duprez,

18 delphine.duprez@upmc.fr

Keywords

20   Chick, limb development, connective tissue differentiation, transcription factors, ChIP-seq, RNA-

21   seq, transcriptional regulatory network, signalling pathways, extracellular matrix.

22

List of abbreviations

24   3F, triple-FLAG; chMM, chick micromass; CDS, coding sequence; ChIP-seq, chromatin

25   immunoprecipitation followed by massively parallel DNA sequencing; CT, connective tissue; DE,

26   differentially expressed; ECM, extracellular matrix; FDR, false-discovery rate; GO, gene ontology;

27   IDR, irreproducibility discovery rate; padj, Benjamini-Hochberg adjusted p-value; PCA, principal

28   components analysis; rlog; regularized logarithm; RNA-seq, whole-transcriptome sequencing; TF,

29   transcription factor; TFBS, transcription factor binding sites; TPM, transcripts per million; TSS,

30   transcriptional start site.

Abstract

Background

Connective tissues support, connect and separate tissues and organs, playing crucial roles in development, homeostasis and fibrosis. Cell specification and differentiation is triggered by the activity of specific transcription factors. While key transcription factors have been identified for differentiation processes of most tissues, connective tissue differentiation remains largely unstudied.

Results

To gain insight into the regulatory cascades involved in connective tissue differentiation, we selected five zinc finger transcription factors - OSR1, OSR2, EGR1, KLF2 and KLF4 - based on their expression patterns and/or known involvement in the differentiation of mesenchymal cells into connective tissue subtypes. We combined RNA-seq with ChIP-seq profiling in chick limb cells following overexpression of individual transcription factors. We identified a set of common genes regulated by all five transcription factors, which constitutes a connective tissue core expression set. This common core was enriched in genes associated with axon guidance and myofibroblast signature. In addition, each of the transcription factors regulated a different set of extracellular matrix components and signalling molecules, which define local molecular niches important for connective tissue development and function.

Conclusions

The established regulatory network identifies common and distinct molecular signatures downstream of five connective tissue-associated transcription factors and provides insight into the signalling pathways governing limb connective tissue differentiation. It also suggests a concept whereby local molecular niches can be created via the expression of specific transcription factors impinging on the specification of microenvironments.

Background

All cells in a multicellular eukaryotic organism share the same genetic information. However, the choice of specification, differentiation and function taken by a progenitor cell encompasses dramatic transcriptional and finally phenotypic changes that differ for each cell/tissue type. Lineage-specific genetic programs consist in a fine-tuning between the repression and the expression of a given set of genes in response to extrinsic and intrinsic signals at a specific location and/or at a precise time [1]. Progenitor cells are specified and induced to differentiate along a certain lineage upon activation of lineage-specific key transcription factors (TFs) that drive specific transcriptional programs [2]. As a consequence, the differentiating progenitor cells express lineage-specific genes that reinforce lineage commitment, as well as providing unique characteristics to the specific cell type and the tissue it gives rise to.

Connective tissue (CT) is one of the main components of the body supporting tissues and organs. The term of CT gathers together an ensemble of tissues such as specialized CT (cartilage and bone), soft CT (adipose tissue and vasculature) and dense CT. Dense CT can be divided into regular CT (tendon and ligament) and irregular CT (loose CT surrounding or within organs such as muscle CT) [3]. Regular and irregular CTs are part of the musculoskeletal system and are associated with cartilage/bone and skeletal muscle. Dysregulation of CT homeostasis leads to fibrosis, which is observed during pathological tissue repair or healing processes and in cancer [4]. Although fibrosis is a common research subject, normal CT formation during development remains to date poorly investigated.

The appendage of vertebrate embryos is an excellent model system for analysing tissue differentiation and cellular interactions during development. In limbs, cells forming the skeleton, as well as regular and irregular CTs, are derived from the lateral plate mesoderm, while myogenic cells originate from the somites [5,6]. Classical embryological experiments have shown that limb patterning is dependent on lateral plate-derived CTs that provide instructive cues to guarantee correct muscle, nerve and vessel formation [7–10]. This indicates that CT cells are key players creating local

4

80   microenvironments that contain permissive and/or instructive cues for organ patterning. While master

81   TFs governing cell-type specific gene expression programs have been identified for cartilage (SOX5,

82   SOX6, SOX9), bone (OSX, RUNX2) and muscle (MYF5, MYF6, MYOD) development [11,12],

83   knowledge is sparse for dense CT. CT is mostly identified by gene/protein expression associated with

84   CT function. Irregular CT is associated with type-III and -VI collagens, while regular CT

85   (tendon/ligament) is characterized by the expression of structural and functional components such as

86   type-I and -XII collagens and Tenomodulin (TNMD) [7,13]. Few TFs specific for CT lineages have

87   been identified. Scleraxis (SCX) is to date the unique marker for tendon progenitors, however it is

88   not necessary for development of most tendons [14,15]. Early growth response 1 (EGR1) is also not

89   required for mouse tendon development but is involved in type-I collagen production in chick and

90   mouse developing tendons [16]. Moreover, EGR1 forced expression is sufficient to induce the

91   expression of tendon-associated genes in murine mesenchymal stem cells [17]. Limb irregular CT

92   associated with muscle is marked by the expression of TCF4 (TCF7l2), TBX4 or TBX5, but they

93   have no obvious role in CT differentiation [18,19]. In contrast, Odd-skipped related 1 and 2 (OSR1

94   and OSR2) are expressed and involved in loose irregular CT differentiation during chick and mouse

95   limb development [20,21,Vallecillo Garcia et al. in revision].

96   Here, we analysed the molecular mechanisms underlying CT differentiation and function during

97   chick limb development, in order to provide a framework for future analyses of CT development and

98   CT-muscle interconnectivity. To this end, we selected five zinc finger TFs involved or presumably

99   involved in CT differentiation. OSR1, OSR2 and EGR1 were chosen based on their demonstrated

100   contribution in irregular and regular CT development, respectively. KLF2 and KLF4 (Krüppel-like

101   factor 2 and 4) were chosen based on their expression patterns in CT associated with tendons,

102   although their role in limb development is presently not elucidated. We combined whole-

103   transcriptome sequencing (RNA-seq) and chromatin immunoprecipitation followed by massively

104   parallel DNA sequencing (ChIP-seq) to identify the gene regulatory programs downstream of each

105   TF. This allowed us to design a novel, unique and unexplored global regulatory network underlying

5

106   CT differentiation and to identify common and specific molecular niches that are shaped during this

107   process.

Results

Limb expression patterns of CT-associated TFs

We investigated gene expression patterns of the five selected CT-associated TFs in chick limbs during development. Consistent with previous observations [20,21], *OSR1* and *OSR2* were expressed in dorsal and ventral limb regions of E4.5 chick embryos (Fig. 1a, b). *OSR1* and *OSR2* expression domains overlapped with those of *SCX* and *MYOD*, which labelled tendon and myogenic cells, respectively (Fig. 1c, d). In contrast to both *OSR* transcripts, *EGR1*, *KLF2* and *KLF4* were not detected in E4.5 limb buds. At later stages of limb development, when the initial pattern of the musculoskeletal system is set, both *OSR1* and *OSR2* were not expressed in *SCX*-positive tendons, but rather expressed in muscle CT, interstitial to muscle fibres, while *OSR1* was also detected surrounding individual muscles (Fig. 1e-i). *EGR1* was expressed in tendons, close to muscle attachments (Fig 1j, k), as previously described [16], while *KLF2* and *KLF4* transcripts delineated *SCX*-positive tendons of the knee of E9.5 chick embryos (Fig. 1l-q). In summary, *OSR1* and *OSR2* label irregular CT, whereas *EGR1*, *KLF2* and *KLF4* are expressed in different regions of regular CT in developing chick limbs.

The selected TFs influence differentiation of limb mesenchymal cells

To analyse the functionality of the five TFs towards CT differentiation, we chose the chick micromass (chMM) explant model (Fig. 2a). In this three-dimensional culture model, primary limb bud cells behave close to the in vivo situation and differentiate into the mesenchymal lineages observed in native limb buds [22]. We tested the ability of the five TFs to shift cell differentiation via their retroviral overexpression in chMM cultures. While the overall morphology of the chMM cultures remained unchanged across all conditions, cartilage differentiation was affected upon TF overexpression (Fig. 2b, c). In agreement with previous observations [21], OSR1 and OSR2 overexpression reduced chondrogenic matrix production by 58% and 67%, respectively, compared

7

133 to control cultures (Fig. 2c, d). Similarly, KLF2 and KLF4 overexpression induced a reduction of

134 cartilage nodule formation, but to a lower extent compared to OSR1 and OSR2 (Fig. 2c, d). EGR1

135 was the only factor that increased chondrogenic matrix production within the chMM cultures (Fig.

136 2c, d). Quantitative RT-PCR analysis of transcript levels of cartilage-associated genes, *SOX9* and

137 *COL2A1*, confirmed the inhibitory effect of OSR1, OSR2, KLF2 and KLF4 overexpression, as well

138 as the positive effect of EGR1 overexpression on cartilage differentiation (Fig. 2e-g). EGR1

139 overexpression increased the expression of the tendon differentiation marker, *TNMD*, while not

140 affecting that of the irregular CT markers *COL3A1* and *COL6A1* (Fig. 2e). Overexpression of KLF2,

141 but not that of KLF4, increased the expression levels of the tendon markers *SCX* and *TNMD* (Fig.

142 2f). Both KLF factors also increased *COL6A1* expression (Fig. 2f). OSR2 overexpression increased

143 the expression of the CT markers *COL3A1* and *COL6A1*, while OSR1 overexpression only affected

144 *COL3A1* expression (Fig. 2g). In summary, the TFs had different outcome on CT differentiation.

145 OSR1 and OSR2 drove undifferentiated limb mesenchymal cells towards irregular CT differentiation

146 at the expense of cartilage differentiation. EGR1 induced tendon and cartilage marker expression,

147 while not affecting irregular CT marker expression. KLF2, but not KLF4, promoted the expression

148 of tendon markers, while both KLF factors increased *COL6A1* expression and decreased *COL2A1*

149 expression. In conclusion, all five TFs proved functional in this model towards an effect on CT cell

150 differentiation thus validating them for further analysis.

151

152 **Transcriptome analysis reveals similar regulatory functions between the five TFs**

153 To gain insight into regulatory functions of the five TFs, transcriptome analysis was performed by

154 RNA-seq. Total RNAs were extracted from two independent biological replicates of 5-day chMM

155 cultures overexpressing each of the TF and subjected to high-throughput sequencing. Principal

156 components analysis (PCA) and hierarchical clustering of the Euclidean distances on global gene

157 expression profiles depicted a separation between the TF-overexpressing chMM cultures (Fig. 3a;

158 Additional file 1: Fig. S1). Consistent with their similar expression domains in irregular CT, the gene

8

159   expression profiles induced upon OSR1 and OSR2 overexpression were grouped together. In contrast,

160   the gene expression profiles retrieved upon overexpression of the tendon-related TFs, EGR1, KLF2

161   and KLF4, were gathered together in a second group. However, the gene expression profiles upon

162   KLF2 and KLF4 overexpression were more similar to each other than that induced upon EGR1

163   overexpression, which was consistent with their distinct expression domains associated with tendons.

164   In summary, the gene expression profiles retrieved in the chMM cultures are in line with the limb

165   expression patterns of the five TFs in the different CT subtypes.

166   Following differential expression analysis, we identified between 1,369 and 2,907 differentially

167   expressed (DE) genes for each TF-overexpressing culture compared to control cultures, resulting in

168   a total of 10,712 DE genes for all TFs (Additional file 1: Fig. S2a-e). Interestingly, DE genes detected

169   across all chMM culture conditions corresponded to a list of 4,298 non-redundant genes, indicating

170   that the TFs shared common regulatory targets (Fig. 3b; Additional file 2: Table S1). While about

171   one third (1,487; 34.6%) of DE genes were specific to a single TF, 2,811 (65.4%) DE genes were

172   shared by at least two TFs (Fig. 3b). In addition, 726 (16.9%) DE genes were identified in all TF-

173   overexpressing culture conditions (Fig. 3b). This indicates that the five TFs share a core of common

174   regulatory targets, despite being expressed in distinct CT subtypes. In line with this, fold-change

175   comparison (whether the gene is upregulated or downregulated) of the DE genes shared by at least

176   two TFs revealed a high consistency among the TF regulatory patterns. Only 48 (1.7%) shared DE

177   genes were identified as being regulated in opposite directions between the subset of TFs

178   misregulating them (Additional file 1: Fig. S3). Therefore, the TFs did not have only similarities in

179   the genes they regulated, but also in the manner these genes were affected.

180   Given the high consistency in the TF regulatory patterns and the elevated number of shared regulated

181   genes, a gene clustering approach was performed on the 4,298 non-redundant DE genes by using *K*-

182   means. This approach partitioned the DE genes into 8 clusters (Fig. 3c). A gene ontology (GO)

183   analysis was then performed to identify potential biological processes enriched within each cluster

184   (Fig. 3d). Genes downregulated by all five TFs were mainly involved in protein localization, ion

9

185   transport and metabolic processes (Fig. 3c, d: cluster I). Genes upregulated by all five TFs were

186   related to metabolic processes, gene expression, cellular component organization and several GO

187   terms associated with the regulation of cell signalling and communication (Fig. 3c, d: cluster VIII).

188   The clusters II, III, IV and V corresponded to genes upregulated specifically by one TF or by two

189   closely related paralogous TFs (OSR1/OSR2 or KLF2/KLF4) (Fig. 3c). Genes in these clusters were

190   mainly enriched for cell differentiation, mesoderm development, cell signalling/communication and

191   biological/cell adhesion (Fig. 3d). Genes downregulated upon OSR1 and OSR2 overexpression were

192   enriched for biological processes related to chondrogenesis (Fig. 3c, d: cluster VI), which was

193   consistent with their anti-chondrogenic effect in chick cell cultures (Fig. 2c, d, g) [21]. Genes

194   downregulated upon KLF2 and KLF4 overexpression were related to cell signalling and adhesion

195   (Fig. 3c, d: cluster VII). In summary, the five CT-associated TFs differentially regulate the expression

196   of genes mainly related to cell differentiation, signalling and adhesion. Thereby, they show a

197   significant degree of overlapping regulatory function despite belonging to distinct CT subtypes.

198

199   **Molecular signatures downstream of the five TFs**

200   The gene clustering approach revealed that a high proportion of genes upregulated by the selected

201   CT-associated TFs was involved in signal transduction and biological adhesion. To further investigate

202   this feature, signalling pathway enrichment analysis was performed on the complete set of DE genes

203   identified for each TF. Of particular interest, signalling pathways related to extracellular matrix

204   (ECM) components, such as integrin and cadherin signalling pathways, Wnt signalling, CCKR

205   signalling and angiogenesis were enriched across all five TFs (Fig. 4a). Additional pathways were

206   specifically enriched by a subset of TFs. For instance, TGF-β signalling pathway was identified upon

207   overexpression of OSR1, OSR2 and KLF2, while Notch signalling pathway was enriched for both

208   KLF2 and KLF4 DE genes (Fig. 4a). In addition, "axon guidance mediated by netrin" and

209   "cytoskeletal regulation by Rho GTPase" pathways were enriched in both OSR1- and OSR2-

210   associated DE genes (Fig. 4a). When comparing the averaged fold change across all TFs for each DE

10

211 gene, it appeared that DE genes within each aforementioned pathway were significantly upregulated

212 (Fig. 4b: median log2 fold change close to 1; Wilcoxon rank-sum test, $P < 0.05$). This tendency was

213 not observed for the remaining non-DE genes associated with these signalling pathways (Fig. 4c:

214 median log2 fold change close to 0). Nevertheless, a proportion of DE genes appeared rather

215 downregulated for the integrin and TGF-β signalling pathways (Fig. 4b: lower whisker). This

216 corresponded to a set of genes mainly repressed by both OSR factors (Additional file 1: Fig. S4a, b).

217 Most of these genes encode collagens and BMP/GDF signalling molecules associated with cartilage

218 and bone development, consistent with the anti-chondrogenic function of OSR1 and OSR2 (Fig 2c,

219 d, g) [21]. In agreement with the signalling pathway enrichment analysis, overrepresentation test on

220 the 4,298 non-redundant DE genes highlighted ECM, membrane and cytoskeleton cellular

221 components (Fig. 4d). Altogether, gene expression profiling of chMM cultures overexpressing each

222 TF supports a core of common regulatory functions across all TFs related to cell signalling,

223 communication and adhesion. In addition, each TF (or paralogous TFs) also appear to be involved in

224 e.g. regulation of individual signalling pathways, which could contribute to create a local

225 microenvironment related to each CT subtype.

226

227 **Establishing genome-wide TF binding profiles**

228 To further clarify the molecular mechanisms downstream of each TF, we investigated their genome-

229 wide binding profile via ChIP-seq. Due to the absence of specific antibodies targeting each of the

230 selected chicken TFs, we used the triple-FLAG (3F) tag that was fused C-terminally to the coding

231 sequence (CDS) of each TF prior to their insertion into the RCAS-BP(A) vector. ChIP-seq was

232 performed in two independent biological replicates of 5-day chMM cultures overexpressing each of

233 the 3F-tagged TFs. Prior to ChIP, retroviral infection and overexpression of the recombinant TFs

234 were assessed by immunohistochemistry and Western blot analysis against the 3F tag (Additional file

235 1: Fig. S5a, b). Similarity across all ChIP-seq signal profiles was assessed genome-widely in 500-bp

236 non-overlapping windows by PCA. Comparison of the three first principal components partitioned

237 the TF signal profiles by biological replicates and TF subgroups (Fig. 5a), indicating that paralogous

238 TFs (OSR1/OSR2 and KLF2/KLF4) had a similar distribution across the genome.

239 Following peak calling, a total of 95,884 TF binding sites (TFBS) were identified (OSR1, 20,983;

240 OSR2, 22,403; EGR1, 16,627; KLF2, 21,352; and KLF4, 14,519), which was 10-fold higher than the

241 total number of DE genes. To potentially distinguish between functional and non-functional TF

242 binding events, we decided to build a comprehensive map of the chMM chromatin landscape by

243 characterizing promoter and enhancer regulatory domains. ChIP-seq experiments were performed

244 against five histone post-translational modifications in two independent biological replicates of 5-day

245 chMM cultures infected with retroviral particles overexpressing no recombinant protein,

246 corresponding to the control conditions used for RNA-seq (Additional file 1: Fig. S6a). This approach

247 was chosen to reveal the unbiased chromatin landscape in chMM cultures independently of TF

248 overexpression. Mono-, bi- and tri-methylation of H3K4 (H3K4me1/2/3) were assessed to identify

249 promoter and enhancer domains [23]. In addition, H3K27ac and H3K27me3 enrichments were

250 investigated to distinguish between regions of transcriptional activity and facultative

251 heterochromatin, respectively [23]. Based on this approach, we identified 20,427 promoters and

252 55,597 enhancers (Additional file 1: Fig. S6b, c). Surprisingly, we observed a decreased enrichment

253 for histone marks associated with transcriptional activity (H3K4me3, H3K27ac) compensated by an

254 increased signal of the repressive mark H3K27me3 at the transcriptional start site (TSS) positions of

255 DE genes, as compared to their genome-wide levels (Fig. 5b, c: y-axis scale). This tendency was

256 confirmed when investigating the TSS positions of a set of randomly selected genes of similar size

257 (Fig. 5d: y-axis scale). This suggested the existence of bivalent promoter domains, which are known

258 to be enriched in lineage-regulatory genes [24]. Consistently, DE genes were more associated with

259 bivalent promoter domains at their TSS positions than randomly selected genes, regardless of their

260 similar gene expression levels (Fig. 5e, f: left panels, white curve). By separating H3K4me3 (active

261 and bivalent promoters) and H3K27me3 (bivalent promoters only) signals, it appeared that H3K4me3

262 active mark was overall less enriched at the TSS positions of DE genes than randomly selected genes

12

263 (Fig. 5e, f: middle panels), whereas the H3K27me3 repressive mark displayed an opposite distribution

264 (Fig. 5e, f: right panels). Although we cannot exclude that the increased ratio between repressive and

265 active signal at promoters of genes affected by TF overexpression may reflect regulatory dynamics

266 in the different cell populations [25], bivalent promoter domains suggest that DE genes are overall

267 dynamically regulated and likely associated with CT differentiation and subtype-specific function, as

268 opposed to housekeeping and ubiquitous genes that would be active and expressed across all cell

269 types in limb cultures.

270

271 **Genome-wide patterns of TF occupancy**

272 In order to assess TFBS functionality, binding locations identified for each TF were intersected with

273 the regulatory domains. We focused on TFBS contained within promoters and enhancers, since we

274 considered that binding events located in these regulatory domains likely contributed to the regulation

275 of gene expression. Out of the 95,884 binding sites identified across all TFs, 31,289 (32.6%)

276 overlapped promoter and enhancer regions, corresponding to 3,819-9,291 (17.9%-55.9%) binding

277 events for each TF (Fig. 6a). De novo motif analysis was then performed on the 1,000 most significant

278 binding sites for each TF. Recognition motifs identified for OSR1 and OSR2 were very similar and

279 highly conserved with their known binding motifs in the fruit fly and the mouse (Fig. 6b) [26,27]. In

280 agreement with previous reports, EGR1 and KLF4 binding motifs were enriched in cytosine/guanine

281 (Fig. 6b) [26,28]. KLF2 recognition motif was highly consistent with the core binding sequence of

282 the KLF protein family and was similar to the KLF4 secondary motif (Fig. 6b) [29]. Both binding

283 motifs identified for KLF4 could contribute to its regulatory pattern observed in limb cell cultures,

284 considering the 767 DE genes specifically identified for KLF4 and the 1,866 DE genes shared

285 between KLF4 and KLF2 (Fig. 3b). Given the shared DE genes and the analogous recognition motifs,

286 we then wondered whether the TFs depicted similarities regarding their occupancy within promoters

287 and enhancers. Summit locations of all TFBS were retrieved and compared to each other. Binding

288 events located within 500 bp surrounding peak summits (± 250 bp) were considered as shared binding

13

289   locations. The 31,289 TFBS associated with promoters and enhancers corresponded to 17,714

290   binding regions. Out of these 17,714 regions, 10,979 (62.0%) were specific to a single TF, whereas

291   6,735 (38.0%) were shared by at least two TFs, including 1,045 (5.9%) that were common to the five

292   TFs (Additional file 1: Fig. S7a). Shared occupancy among the TFs was further assessed by

293   comparing the total number of shared binding sites that were identified between each TF pair.

294   Consistent with their similar binding motifs, OSR1 and OSR2 shared more binding locations with

295   each other than with any other TF (Additional file 1: Fig. S7b, c). Likewise, EGR1 and KLF4 tended

296   to bind preferentially at similar regions as compared to the other TFs (Additional file 1: Fig. S7d, e).

297   Surprisingly, we did not observe any preferential binding for KLF2 with any of the other TFs,

298   including KLF4 (Additional file 1: Fig. S7f). Altogether, the binding motifs of the TFs and their

299   genome-wide occupancy profiles appear consistent with the common and distinct regulatory patterns

300   observed at the gene expression level.

301

302   **Identification of functional TFBS confirms a common regulatory core and distinct functions**

303   **for the five TFs**

304   To finally distinguish between indirectly and directly regulated genes, the 17,714 non-redundant

305   binding regions identified across all TFs were intersected with the 4,298 DE genes identified from

306   the RNA-seq data. Regions spanning from 10 kb upstream of the gene TSS to 10 kb downstream of

307   the gene 3'-end were investigated for the presence of TFBS located within regulatory domains. This

308   resulted in the identification of 3,210 genes that were potentially directly regulated by the TFs

309   (Additional file 2: Table S1). The proportion of putative direct targets ranged from 20.9% (OSR1,

310   417; KLF2, 449) to 49.4% (EGR1, 677), depending on the TF (Fig. 6c). Consistent with the previous

311   observations that the TFs shared common regulatory patterns and binding events in addition to their

312   own specificity, the 3,210 genes considered as potential direct targets corresponded to 1,858 non-

313   redundant genes (Fig. 6d). While 1,076 (57.9%) genes were directly regulated by a single TF, 782

314   (42.1%) genes were shared by at least two TFs, including 77 (4.1%) genes common to all TFs (Fig.

14

315   6d). In particular, OSR1 and OSR2 shared 318 target genes, while KLF4 possessed 317 target genes

316   in common with EGR1 and 313 target genes in common with KLF2 (Fig. 6d). To further explore the

317   regulatory link existing between the TFs, we next investigated whether the TFs sharing the same

318   target genes tended to display similar binding locations. Consistent with the genome-wide occupancy

319   analysis, OSR1 and OSR2 shared more binding locations with each other than with any other TF,

320   albeit OSR2 displayed a higher binding specificity than OSR1 (Additional file 1: Fig. S8a, b).

321   Likewise, EGR1 and KLF4 tended to bind preferentially at similar regions as compared to the other

322   TFs, while also exhibiting their own binding specificity (Additional file 1: Fig. S8c, d). In contrast to

323   the genome-wide occupancy analysis that did not depict any preferential binding between KLF2 and

324   KLF4, we observed that both TFs tended to occupy similar locations in the vicinity of their common

325   target genes, as compared to the other TFs (Additional file 1: Fig. S8e, f). In conclusion, the binding

326   profiles of CT-associated TFs reflect their specificity and similarity in regards to their regulatory

327   patterns.

328

329   **Validation of selected target genes**

330   Coexpression of a TF and its putative target gene is a prerequisite for transcriptional regulation.

331   Therefore, we compared the expression domains of TFs with that of selected candidate genes

332   encoding signalling molecules in chick limbs. *NTN1* (netrin 1) was one of the genes that was

333   upregulated in limb cell cultures upon overexpression of each TF (Fig. 7a; Additional file 1: Fig. S9a;

334   Additional file 2: Table S1), through binding within an intronic enhancer (Fig. 7b). *NTN1* encodes a

335   laminin-related secreted protein involved in axon guidance [30,31]. In E5.5 chick embryos, *NTN1*

336   was expressed in both limb stylopod and zeugopod, displaying overlapping expression domains with

337   those of all five TFs (Fig. 7c-h). At E8, *NTN1* was expressed in tendons, overlapping with *EGR1*

338   expression domain close to muscle attachment (Fig. 7i, j). *NTN1*, *OSR1* and *OSR2* expression was

339   observed in muscle CT at E9.5 (Fig. 7k-m). In addition, *NTN1* transcripts were detected in tissues

340   delineating tendons at E9.5, similarly to *KLF2* and *KLF4* transcripts (Fig. 7n-p).

15

341   Given the similar regulatory profiles of OSR1 and OSR2, we also selected *WNT11*, a common target

342   gene of both TFs. *WNT11* encodes a secreted component of the non-canonical Wnt planar cell polarity

343   pathway [32]. Both OSR factors increased *WNT11* expression in chMM cultures and bound at the

344   same location within an intronic enhancer (Fig. 8a, b; Additional file 1: Fig. S9b; Additional file 2:

345   Table S1). In E5.5 chick embryos, *WNT11* was expressed in limb mesenchyme, consistent with *OSR1*

346   and *OSR2* expression patterns (Additional file 1: Fig. S9c-e). At E8 and E9.5, *WNT11* transcripts

347   were detected in irregular CT within and surrounding muscles, overlapping with *OSR1* and *OSR2*

348   transcripts (Fig. 8d-i). *WNT11* expression in CT downstream of OSR1 and OSR2 is consistent with

349   the identified role of WNT11 in regulating muscle fibre type and orientation [33]. An additional target

350   gene, *GDF6*, which encodes a secreted signalling factor of the TGF-β superfamily [34], was

351   upregulated upon overexpression of OSR1 and OSR2 in chMM cultures (Fig. 8a; Additional file 1:

352   Fig. S9b; Additional file 2: Table S1). However, only a binding site for OSR2 was detected in the

353   vicinity of *GDF6*, indicating that OSR1 was not directly involved in the regulation of *GDF6*

354   expression (Fig. 8c). In limbs of E5.5 and E8 chick embryos, *GDF6* expression domains overlapped

355   with those of *OSR2* (Fig. 8j, k; Additional file 1: Fig. S9d, f).

356   As a specific target gene of EGR1, we selected *WNT4*, which encodes a secreted member of the

357   canonical Wnt signalling pathway [35]. *WNT4* was upregulated upon EGR1 overexpression in chMM

358   cultures and was associated with an EGR1 binding site in its promoter region (Fig. 8l, m; Additional

359   file 1: Fig. S9g; Additional file 2: Table S1). This is consistent with previous findings where EGR1

360   has been shown to bind upstream of *Wnt4* gene in the uterine endometrium during mouse pregnancy

361   [36]. *EGR1* and *WNT4* displayed overlapping expression domains in proximal regions of forelimbs

362   of E5.5 chick embryos (Additional file 1: Fig. S9h, i) and in tendons, close to muscle attachment in

363   E8 limbs (Fig. 8n-p).

364   Given the common regulatory patterns of both KLF factors, we selected *FZD1*, which encodes a

365   frizzled class receptor of Wnt signalling proteins [37]. *FZD1* was upregulated upon overexpression

366   of KLF2 and KLF4 in chMM cultures (Fig. 8q; Additional file 1: Fig. S9j; Additional file 2: Table

367 S1), and harboured a binding site for both KLF factors within its promoter region (Fig. 8r). In E8

368 chick embryos, *FZD1* was expressed in tissues delineating tendons, overlapping with the expression

369 domains of *KLF2* and *KLF4* (Fig. 8t-v). Considering that KLF4 also displayed a distinct regulatory

370 profile (Fig. 3c, 6d), we selected *INHBA*, which encodes the inhibin beta A subunit, a member of the

371 TGF-β signalling pathway [38]. *INHBA* was upregulated upon KLF4 overexpression in chMM

372 cultures (Fig. 8q; Additional file 1: Fig. S9j; Additional file 2: Table S1). In addition, a KLF4 binding

373 site was located within an enhancer upstream of the TSS position of *INHBA* (Fig. 8s). In E5.5 and E8

374 chick limbs, *INHBA* and *KLF4* displayed overlapping expression domains in tissues delineating

375 tendons (Fig. 8u, w; Additional file 1: Fig. S9k-m). Altogether, the selected target genes and their

376 related CT-associated TFs exhibited overlapping expression domains in chick limbs.

377

378 **Common and divergent signalling/ECM signatures regulated by the CT-associated TFs**

379 CT cells shape their microenvironment mainly by production of signalling/ECM molecules and/or

380 via remodelling of the ECM. To explore this feature, we built a regulatory network on the 189 DE

381 genes that were associated with seven of the previously identified signalling pathways (Fig. 4a, b;

382 Additional file 2: Table S1). The resulting transcriptional network was composed of 513 interactions

383 divided between 175 (34.1%) direct and 338 (65.9%) indirect connections (Fig. 9a). This network

384 highlighted common and unique features for the CT-associated TFs. 38 (20.1%) genes were regulated

385 by all five TFs, revealing a CT-typical signalling signature, whereas 47 (24.9%) genes were

386 exclusively shared by paralogous TFs (OSR1/OSR2 and KLF2/KLF4) and 45 (23.8%) genes were

387 specific to a single TF (Fig. 9a; Additional file 2: Table S1). The regulatory network was then

388 subdivided for each individual TF to visualize the molecular interplay of each TF on selected

389 signalling pathways (Additional file 1: Fig. S10-14). For instance, the Wnt signalling pathway was

390 differently affected depending on the TF, with different sets of *WNT* ligand and *FZD* receptor genes

391 regulated by each TF (Additional file 1: Fig. S10-14).

392   To reduce complexity, we then focused on the Notch, TGF-β and Wnt signalling pathways (Fig. 9b).

393   Key components of these signalling pathways, *NOTCH1* (Notch receptor) and *SMAD4* (TGF-β

394   signalling transducing protein) were regulated by all five TFs (Fig. 9b; Additional file 2: Table S1).

395   By contrast, other genes were regulated by a subset of TFs. For instance, *WNT11* and *BMPR1B* (BMP

396   receptor) were specific to OSR1 and OSR2, whereas *FZD1* and *PRKCQ* (protein kinase C theta) were

397   regulated exclusively by both KLF factors (Fig. 9b; Additional file 2: Table S1). Lastly, we found

398   genes that were specific to a single TF. This is the case for *GDF6* and *SMAD9* directly upregulated

399   by OSR2, *WNT4* and *TCF7L1* (*TCF3*) directly upregulated by EGR1, and *INHBA* and *BMP8B* (BMP

400   secreted ligand) directly upregulated by KLF4 (Fig. 9b; Additional file 2: Table S1). Altogether, this

401   regulatory network identifies signalling genes that likely contribute to the biological function of all

402   CTs or CT subtypes.

403   By finally focusing on direct target genes associated with the ECM, we found that the TFs regulated

404   distinct, nevertheless partly overlapping molecular niches (Fig. 9c; Additional file 2: Table S1).

405   *ADAMTS15* was directly upregulated by the five TFs and *ADAMTS8* was specific to EGR1, whereas

406   *ADAMTS18* was directly downregulated by OSR2 and KLF2, (Fig. 9c; Additional file 2: Table S1).

407   ADAMTS proteins are secreted metalloproteases with thrombospondin type-I motif that are involved

408   in procollagen processing [39]. While ADAMTS15 and ADAMTS8 have proteoglycanolytic activity

409   [39], mutations in *ADAMTS18* have been associated with bone disorders [40]. In addition, CT-

410   associated TFs appeared to mediate collagen deposition by directly regulating genes encoding

411   collagen α-chains. *COL4A1* and *COL4A2* were directly upregulated by OSR1, OSR2 and KLF2 (Fig.

412   9c; Additional file 2: Table S1). Type-IV collagen contributes to the assembly of basal lamina by

413   binding to laminins [41]. By contrast, *COL9A1*, *COL9A2* and *COL11A1* were directly downregulated

414   by OSR1 and OSR2 (Fig. 9c; Additional file 2: Table S1). Type-IX and XI collagens are known to

415   form a network with type-II collagen in cartilaginous ECM [42]. KLF4 directly promoted the

416   expression of *COL1A1* (Fig. 9c; Additional file 2: Table S1). Type-I collagen fibrils are the main

417   component of tendons [7]. The ECM also acts as a source of developmental signals by sequestering

18

418   and diffusing paracrine factors. The TFs appeared to directly mediate the positive expression of genes

419   encoding laminin-related secreted netrins, such as *NTN1* (all five TFs), *NTN3* (EGR1 and KLF4)

420   *NTN4* (OSR1 and EGR1) and *NTNG1* (OSR2) (Fig. 9c; Additional file 2: Table S1). In conclusion,

421   the CT-associated TFs contribute to provide local patterning cues by mediating the expression of

422   precise environmental molecules.

Discussion

In this study, we have designed a transcriptional network downstream of five zinc finger TFs involved in CT differentiation. Two TFs (OSR1 and OSR2) were related to irregular CT, one TF (EGR1) to regular CT (tendon), and two TFs (KLF2 and KLF4) showed expression in CT delineating tendons. By combining different genome-wide strategies, we identified common and specific molecular signatures involved in limb CT differentiation. TF overexpression led to transcriptional activity changes in limb cells impacting on numerous cellular processes, including cell communication, migration and cell-cell/cell-matrix adhesion processes. Consistently, genes encoding signalling molecules, ECM components and cytoskeletal proteins appeared as regulated by the five TFs.

Core molecular network downstream of the five TFs

We found 4,298 non-redundant DE genes upon overexpression of the five TFs. 65.4% of these genes were shared by at least two TFs, while 16.9% were common to all TFs. When direct regulation as judged by TF binding was considered, 77 genes were shared between the five TFs (Additional file 2: Table S1). We performed a conservative analysis and also restricted regulatory elements to a distance of 20 kb. It is well established that enhancer elements can be located further away, however identification of these regulatory interactions would include analysis of the 3D chromatin structure [43]. Altogether the number of common targets we identified is likely to be imperfect. Our data nevertheless show that the five TFs display common target genes despite their expression in different subcompartments of limb musculoskeletal system. This indicates that irrespective of CT type, whether it is specialized, dense regular or dense irregular, key molecular features are shared during the differentiation process of CT types, suggesting an archetypical CT signature.

One example for a common and directly regulated gene downstream of the five TFs is *NTN1*. Netrin 1 is a secreted ligand involved in axon guidance and developmental angiogenesis, in addition to preventing apoptosis triggered by one of its receptors, DCC (deleted in colorectal cancer) [44]. With

20

448   the exception of EGR1, which only activates the expression of netrin ligands (*NTN1*, *NTN2* and

449   *NTN3*), the other TFs positively regulate the expression of netrin receptors, *UNC5A* and/or *UNC5B*

450   (Additional file 2: Table S1), known to mediate the netrin 1-induced axon chemorepulsion [44]. Our

451   data suggests that NTN1 is an unexpected actor involved in migration and/or survival of CT cells

452   during limb development. The molecular core downstream of the five TFs also includes a

453   myofibroblast signature with *SRF*, *TAGLN* (*SM22*, transgelin), *TAGLN2* (transgelin 2), *CNN1*

454   (calponin 1) and *ACTG2* (actin gamma 2) genes, which are positively activated by the five TFs,

455   although not involving systematically direct binding sites (Additional file 2: Table S1). The

456   myofibroblast signature is also confirmed by the presence of *SMAD4* in the list of the 77 common

457   and directly regulated genes (Additional file 2: Table S1). *SMAD4* is a well-known profibrotic factor

458   downstream of TGF-β1 [45]. The myofibroblast signature upon TF overexpression indicates that

459   developmental CT differentiation shares molecular mechanisms with myofibroblast activation during

460   fibrosis. In this context, it is noteworthy that *NOTCH1*, a component of a developmental signalling

461   pathway described to be also involved in adult fibrosis [46], is a common and directly regulated gene

462   downstream of the five TFs (Additional file 2: Table S1). The upregulation of Notch pathway

463   components by each of the TFs suggests an unexpected involvement of the Notch signalling in limb

464   CT formation during development.

465

466   **Specific regulatory patterns downstream of the five TFs**

467   In addition to sharing a common molecular core, each TF displayed a specific regulatory pattern,

468   albeit convergence was observed between related TFs, i.e. between OSR1 and OSR2 associated with

469   irregular CT, and between EGR1, KLF2 and KLF4 mainly associated with regular CT. OSR1 and

470   OSR2 are two markers of irregular CT and overexpression of each factor promotes the expression of

471   irregular CT markers, such as *COL3A1*, while inhibiting that of cartilage markers in chick limb cells,

472   as previously observed [21]. The list of OSR1 and OSR2 target genes identifies a molecular signature

473   of irregular CT downstream of OSR factors in chick limb cells and potentially increases the limited

21

474 number of irregular CT markers that are currently available. In addition, OSR1 and OSR2 represses

475 the expression of genes related to cartilage lineage, which is consistent with the upregulation of

476 cartilage-associated genes observed in *Osr1*-positive cells extracted from limbs of *Osr1* null mouse

477 embryos [Vallecillo Garcia et al. in revision]. The master regulator of cartilage *SOX9* appears to be a

478 direct target of OSR2, and the accessory regulatory factors *SOX5* and *SOX6* appear to be direct targets

479 of both OSR1 and OSR2 in chMM cultures (Additional file 2: Table S1). OSR1 and OSR2 share 318

480 common target genes, but also display their own specificity. While only 42 target genes are unique

481 to OSR1, 250 target genes are specifically regulated by OSR2 (Additional file 2: Table S1). The BMP

482 ligand *GDF6* is one of the OSR2 specific target genes. GDF6 is known to play a role in establishing

483 boundaries between skeletal elements during limb development, since inactivation of the *Gdf6* gene

484 causes defects in joint, ligament and cartilage formation in mice [34]. This is in line with *Osr2/OSR2*

485 expression in joint interzones in mouse [20] and chick embryos (Fig. 8j), as well as with the joint

486 fusion defects observed in *Osr1/Osr2* double mouse mutants [47].

487 Although not being specific to tendons, EGR1 overexpression is sufficient to drive tendon cell

488 differentiation in mouse mesenchymal stem cells [17]. Over 100 genes upregulated upon EGR1

489 overexpression were listed as being enriched in the transcriptome of *Scx*-positive cells isolated from

490 limbs of mouse embryos [48]. This list includes *ADAMTS8*, *ADAMTS15*, *TAGLN*, *TAGLN2*, *FZD5*

491 and *WNT4*, among others (Additional file 2: Table S1). *BMP4*, known to be expressed in chick limb

492 tendons [49], was also positively regulated by EGR1 (although not directly) in our data (Additional

493 file 2: Table S1). EGR1 is characterised as a fibrosis-promoting factor in many organs [50]. EGR1

494 has been also shown to directly regulate *Tgfb2* transcription in adult mouse tendons [17]. Here, we

495 show that *TGFB1* (coding for the main fibrotic factor) is positively regulated by EGR1 (albeit not

496 directly) in chick limb cells (Additional file 2: Table S1).

497 The exact function of KLF2 and KLF4 in limb musculoskeletal system formation is currently not

498 known. However, we show here that both TFs display a striking expression delineating *SCX*-positive

499 tendon/ligaments. In addition to the clear adhesion/migration signature downstream of both KLF

22

500 factors in chick limb cells, KLF2 and KLF4 activates cell cycle genes such as *CDKN1A* (*P21*) and

501 pluripotency genes (*SOX7*, *DKK1*, among others). Combined with the KLF2 and KLF4 well-known

502 function in somatic cell reprogramming and pluripotency [51], this leads to the interesting idea that

503 cells surrounding *SCX*-positive expression domains could be a source of tendon progenitors during

504 development. Consistent with this idea, different tenogenic properties have been described for

505 peritenon cells and tendon proper cells [52]. Beyond the 313 target genes that are shared between

506 both KLFs and their similar binding occupancy in the vicinity of these target genes, KLF4 possesses

507 439 specific target genes. KLF4 specificity is corroborated with its primary binding motif that differs

508 from the KLF core binding sequence identified in KLF2 and KLF4 ChIP-seq data. Indeed, the KLF4

509 binding site identified upstream of *INHBA* encompasses its primary recognition motif, whereas KLF2

510 and secondary KLF4 binding motifs are detected in the promoter region of their common target gene

511 *FZD1*.

512

513 **A versatile molecular toolkit for shaping local niches**

514 A significant proportion of directly or indirectly regulated DE genes comprises genes encoding

515 signalling-associated molecules or ECM components and cell-matrix attachment molecules. The

516 ECM is a three-dimensional insoluble network composed of secreted macromolecules, which

517 provides positional and physical cues to influence cell position and migration [53,54]. Moreover, the

518 ECM is a storage space for diverse growth factors that can be released upon e.g. proteolytic cleavage

519 or mechanical stimulation. In this view, the ECM is both a scaffold structure and an integral part of

520 cell-cell signalling mechanisms. Our data provide evidence that regional subspecification of limb bud

521 mesenchymal tissues by TFs may be concomitant to local changes in the extracellular milieu.

522 Importantly, we show that a common ECM signature activated by all five TFs exists for limb CT,

523 while in addition each TF drives specific ECM and signalling factor genes. Individual TF or

524 combinatorial TFs will impinge on the production of a particular ECM with specific growth factor

525 decoration. This likely influences the behaviour of neighbouring tissues to create niches for invading

23

526   cells. This is in line with the recognized importance of cell-ECM interactions for skeletal muscle,

527   nerve and blood vessel development [55,56]. Thus, each TF or coordinated expression of a

528   combination of TFs may be an elegant and adaptable way to achieve tissue

529   (sub)compartmentalization in development.

530 **Conclusions**

531 The transcriptional network presented here brings new insights into the molecular mechanisms

532 orchestrating chick limb CT differentiation and function. The activity of CT-associated TFs impinges

533 on overall common mechanisms including cell adhesion, migration and signalling. However, in

534 addition to a common gene regulatory core downstream of the five TFs, each TF drives a specific

535 program. These common and specific programs are likely to be at the root of tissue subspecification

536 and local compartmentalization in developing limbs leading to the creation of local niches supporting

537 organogenesis. This regulatory network offers valuable resources and opens new roads to better

538 understand CT formation and function during limb development. In addition, such adaptable local

539 transcriptional programs may apply to diverse contexts and might be a general principle of functional

540 modulation.

Methods

Experimental procedures

Chick embryos

Fertilized eggs used for in situ hybridization were provided by the Institut de Sélection Animale (JA 57 strain, Lyon, France). Fertilized eggs used to prepare chMM cultures were obtained from VALO BioMedia (Lohmann Selected Leghorn strain, Osterholz-Scharmbeck, Germany). Embryos were staged according to the number of days in ovo at 37.5°C.

Molecular cloning of the transcription factors

The CDSs of the chicken TFs OSR1, OSR2, EGR1, KLF2 and KLF4 were amplified by PCR by using the primers listed in (Additional file 3: Table S2). Cloning of the TF CDSs was performed by using a modified version of the pSlax-13 vector and the RCAS-BP(A) vector as previously described [57], with the exception that the 3F tag was fused C-terminally to each CDS.

Chick micromass cultures

chMM cultures were prepared as previously described [58]. Briefly, limb buds were extracted from E4.5 chick embryos, ectoderm was dissociated by using a Dispase solution (Gibco) at 3 mg/mL, and limb mesenchyme was digested by using a solution composed of 0.1% Collagenase type Ia (Sigma-Aldrich), 0.1% Trypsin (Gibco) and 5% FBS (Biochrom) in 1X DPBS (Gibco). Prior to seeding, mesenchymal cells were mixed with retroviruses (1:1) and maintained in culture for 5 days at 37°C in DMEM/Ham's F-12 (1:1) medium (Biochrom) supplemented with 10% FBS, 0.2% chicken serum (Sigma-Aldrich), 1% L-glutamine (Lonza) and 1% penicillin/streptomycin (Lonza). To assess cartilage differentiation, chMM cultures were fixed for 30 min with Kahle's fixation solution (1% formalin, 30% ethanol and 4% acetic acid) and stained overnight at 4°C in 1% Alcian blue (Sigma-Aldrich) in 0.1 M HCl. Chondrogenic matrix areas were measured by using ImageJ [59]. For Eosin staining, chMM cultures were fixed overnight with 4% PFA in 1X PBS at 4°C and incubated for 2 min with 2.5 g/L of Eosin (Sigma-Aldrich) in 80% ethanol and 0.5% acetic acid. Viral 3F-tagged TF

26

566 expression was monitored by using a mouse antibody directed against the 3F tag (Sigma-Aldrich,

567 F1804; 1:500). Immunohistological staining was performed by using the Vectastain Elite ABC and

568 the DAB Peroxidase Substrate kits (Vector Laboratories).

RNA sequencing

570 Two biological replicates of chMM cultures were prepared from two independent pools of E4.5 limb

571 buds and infected for 5 days with RCAS-BP(A) retroviruses overexpressing each of the TFs or no

572 recombinant protein as control. For both replicates, RNA extracts were obtained by harvesting 6

573 chMM cultures with RLT buffer (Qiagen). Total RNAs were purified by using the RNeasy mini kit

574 (Qiagen) in combination to a DNase I (Qiagen) treatment to prevent genomic DNA contamination.

575 RNA libraries were prepared by using the TruSeq Stranded mRNA Library Preparation kit (Illumina),

576 which enables to preserve the RNA strand orientation. Strand-specific 50-bp paired-end reads were

577 generated by using a HiSeq 2500 sequencer (Illumina) with a mean insert size of 150 bp (Additional

578 file 3: Table S3).

ChIP sequencing

580 Harvesting of chMM cultures and ChIP experiments were performed as previously described [57].

581 Histone modification occupancy was investigated in two independent biological replicates of chMM

582 cultures infected with RCAS-BP(A) retroviruses overexpressing no recombinant protein. 10 µg (8

583 chMM cultures) of chromatin extracts were incubated overnight at 4°C with gentle rocking with the

584 following antibodies: 4 µg of mouse anti-H3K4me1 (Abcam, ab8895); 8 µL of mouse anti-H3K4me2

585 (Abcam, ab32356); 4 µL of mouse anti-H3K4me3 (Millipore, 07-473); 4 µg of mouse anti-H3K27ac

586 (Abcam, ab4729); and 4 µg of mouse anti-H3K27me3 (Millipore, 07-449). TF binding profiles were

587 investigated in two independent biological replicates of chMM cultures infected with RCAS-BP(A)

588 retroviruses overexpressing each of the TFs. 30 µg (24 chMM cultures) of chromatin extracts were

589 incubated overnight at 4°C with gentle rocking with 10 µg of mouse anti-FLAG (Sigma-Aldrich,

590 F1804). Antibody-TF/Histone-DNA complexes were pulled down by using 40 µL of magnetic beads

591 (Dynabeads protein G; Thermo Fischer). Ethanol-precipitated ChIP samples were resuspended in 46

27

592   μL of ddH20. Libraries were prepared by using the NEBNext Ultra DNA Library Preparation kit for

593   Illumina (New England Biolabs). 50-bp single-end reads were generated by using a HiSeq 1500

594   sequencer (Illumina) (Additional file 3: Tables S4, 5). As input control, sonicated DNA from the

595   nuclear fraction of each sample used for the ChIP procedures was also sequenced.

596   **In situ hybridization**

597   Endogenous expression of the TFs was assessed by in situ hybridization on paraffin-embedded tissue

598   sections. Chick embryo limbs were fixed overnight at 4°C in 60% ethanol, 30% formaldehyde at 37%

599   and 10% acetic acid, and further processed as previously described [60]. For whole-mount in situ

600   hybridization, chick embryos were fixed overnight at 4°C with 4% formaldehyde in 1X PBS and

601   processed as previously described [61]. The following probes were used: *cOSR1* and *cOSR2* [20];

602   *cEGR1* [16]; *cKLF2* and *cKLF4* [62]. Expression of tendon and myogenic markers were assessed

603   with the following probes: *cSCX* [15]; *cMYOD* [63]. Primers listed in (Additional file 3: Table S2)

604   were used to generate probes detecting the following genes: *cFZD1*; *cGDF6*; *cINHBA*; *cNTN1* [64];

605   *cWNT4* (GEISHA ID, WNT4.UApcr); *cWNT11* (GEISHA ID, WNT11.UApcr).

606   **Quantitative RT-PCR analysis**

607   Total RNAs were isolated from independent biological replicates of chMM cultures infected for 5

608   days with RCAS-BP(A) retroviruses overexpressing each of the TFs or no recombinant protein as

609   control. RNA extracts were obtained as described for RNA sequencing. 500 ng of RNA extracts were

610   used as template for cDNA synthesis using the High-Capacity cDNA Reverse Transcription kit

611   (Applied Biosystems). Quantitative RT-PCR was performed by using the SYBR Green PCR Master

612   mix (Applied Biosystems) in duplicates. Relative mRNA levels were calculated according to the 2-

613   ΔΔCt method [65]. ΔCts were obtained from Ct normalized with chick *RPS17* (*S17*) and *GAPDH*.

614   For each investigated gene, the mRNA levels of control chMM cultures were normalized to 1.

615   Statistical analysis was performed by using Mann-Whitney U test with the GraphPad Prism V6

616   software. Primers used for quantitative RT-PCR are listed in (Additional file 3: Table S2).

617

Computational analysis

Gene expression profiles

620 RNA-seq strand-specific read pairs were mapped against the chicken genome galGal4 [66] by using

621 TopHat2 v0.14 [67] (parameters: -r 150; -N 3; --read-edit-dist 3; --library-type fr-firststrand; -i 50; -

622 G) and the gene annotation model previously generated [68]. Alignment maps were split by strand

623 by using SAMtools v1.2 [69] according to their FLAG field (strand plus: -f 128 -F 16, -f 80; strand

624 minus: -f 144, -f 64 -F 16). Fragments (both reads of a pair) mapped on gene features were counted

625 by using featureCounts v1.4.6-p3 [70] (parameters: -p; -s 2; --ignoreDup; -B; -R). Chimeric fragments

626 aligned on different chromosomes were taken into consideration to overcome the gene fragmentation

627 due to the location of gene parts on multiple chromosome contigs [68]. Fragment counts were then

628 normalized by using DESeq2 v1.8.1 [71] and transcript abundances were calculated as transcripts per

629 million (TPM) values according to the formula described in [72]. To evaluate the discrepancy among

630 biological replicates and conditions, a regularized-logarithm (rlog) transformation was applied to

631 normalized fragment counts followed by PCA analysis and hierarchical clustering of the Euclidean

632 distances [71]. Differential expression analysis was finally carried out by using DESeq2 and a false-

633 discovery rate (FDR; alpha) of 0.01. Genes with an absolute fold change of at least 2 and a Benjamini-

634 Hochberg adjusted p-value (padj) below 0.01 were considered as being differentially expressed

635 (Additional file 2: Table S1). Heat maps were generated by using the function heatmap.2 from the R

636 package gplots. For given gene lists, rlog transformed fragment counts were used as input and

637 hierarchical clustering was performed according to the one minus Pearson correlation.

638 *K*-means gene clustering

639 *K*-means clustering was performed on the normalized fragment counts of the DE genes by using

640 GENE-E [73] with a row distance metric set at 1 minus Pearson correlation and 2,000 iterations. The

29

641 number of *K* clusters was defined at 8 because lower values did not separate distinct gene clusters

642 and higher values subdivided meaningful gene clusters.

643 **Gene ontology analysis**

644 GO analyses were performed for given gene lists by using the PANTHER statistical

645 overrepresentation test r20160321 [74] and the Bonferroni correction for multiple testing. The

646 following annotations were interrogated: PANTHER version 10.0 released on 2015-05-15 for GO-

647 slim biological process, cellular component and pathways; GO ontology database released on 2016-

648 04-23 for GO biological process complete.

649 **ChIP-seq coverage profiles**

650 50-bp single-end reads generated for each ChIP and input fractions were first filtered on their quality

651 by using the FASTX-Toolkit v0.0.13 [75]. Reads with a median quality value of minimum 28 were

652 retrieved and mapped against the chicken genome galGal4 [66] by using BWA v0.5.9 [76] (default

653 parameters). Uniquely mapped reads were then extracted and duplicated reads were finally removed

654 by using the tool rmdup from SAMtools v1.2 [69]. Histone mark and TF coverage profiles were

655 generated by using the tool bdgcmp from MACS2 v2.1.0.20140616 [77]. ChIP-seq signal was

656 normalized independently for each biological replicate against the pooled input controls of both

657 replicates according to the negative log10 of the Poisson p-value (-m ppois). Similarity between the

658 TF binding profiles was assessed genome-widely in 500-bp non-overlapping windows by using PCA

659 analysis with the R function prcomp (parameters: center TRUE; scale. TRUE).

660 **Histone modification peak calling**

661 Peak calling for the histone ChIP-seq was performed as suggested by the ENCODE consortium [78].

662 For each histone modification, peaks were called independently for each biological replicate and for

663 the pooled biological replicates, each time against the merged input control of both replicates, by

664 using MACS2 v2.1.0.20140616 [77] (parameters: --bw 400, according to the sonicated DNA size; -

665 g 1.0e9; --to-large). Except for the H3K27me3 mark, peak calling was performed twice for each

666   replicate and pooled replicate: (i) narrow peaks passing a p-value (-p) of 0.01; and (ii) broad peaks

667   passing an additional broad-peak p-value (-p 0.01; --broad; --broad-cutoff) of 0.1. Only broad peaks

668   were called for the H3K27me3 ChIP-seq due to its diffused signal. Broad peaks detected for each

669   replicate and pooled replicate that contain at least one narrow peak were extracted by using BEDtools

670   intersect v2.24.0 [79]. Final sets of peaks for each histone modification were obtained by filtering

671   broad peaks called for the pooled replicates that are shared between both biological replicates

672   independently.

673   **Identification of regulatory domains**

674   Regulatory domains were defined according to the combination of the different histone modification

675   profiles obtained by ChIP-seq, independently of the gene annotation model and TSS positions given

676   the fragmentation of the chicken genome [68]. Domains were divided into three categories: (1)

677   promoters; (2) enhancers; and (3) repression islands. (1) Promoters were defined according to the

678   presence of H3K4me3 signal. (2) Enhancers corresponded to regions enriched for H3K4me1 and

679   devoid of H3K4me3 signal. (3) Repression islands were distinguished by the unique presence of

680   H3K27me3 signal. Additional regions enriched for H3K4me2 but with no detectable H3K4me1

681   signal were classified as promoters, whereas regions containing both H3K4me1/2 marks were defined

682   as enhancers. Promoter and enhancer domains were further subcategorised into four distinct states

683   according to the active marks H3K4me3 and H3K27ac, and the repressive mark H3K27me3: (i)

684   inactive, no active and repressive signal detected (H3K4me3^{-}, H3K27ac^{-}, H3K27me3^{-}); (ii) poised,

685   no active mark but repressive signal detected (H3K4me3^{-}, H3K27ac^{-}, H3K27me3^{+}); (iii) active, only

686   active mark detected (H3K4me3^{+} and/or H3K27ac^{+}, H3K27me3^{-}); and (iv) bivalent, both active and

687   repressive marks detected (H3K4me3^{+} and/or H3K27ac^{+}, H3K27me3^{+}).

688   **Chromatin landscape at TSS positions**

689   Normalized ChIP-seq signal was averaged for each histone modification from -2.5 to +2.5 kb

690   surrounding the TSS of all genes, DE genes and randomly selected genes. To further investigate the

691   increased enrichment of H3K27me3 mark, the 4,298 DE genes were filtered based on three criteria:

692   (i) gene located on one single chromosome with a minimum size of 20 kb; (ii) gene body length of at

693   least 1 kb; and (iii) -10/+10-kb regions around TSS within the chromosome borders. The resulting

694   list was composed of 3,070 DE genes. The same criteria were applied to the randomly selected genes

695   giving rise to a set of 3,080 random genes. 10-kb regions surrounding each TSS were retrieved and

696   split into 100 intervals of 200 bp. For the genes having multiple transcripts with distinct TSS

697   positions, the most upstream TSS was selected. Regulatory domains contained in each 200-bp interval

698   were recovered in order to identify the most dominant domain per interval. Intervals marked with

699   active and bivalent promoters were plotted in blue and red, respectively.

Transcription factor peak calling

701   Quality of the TF ChIP-seq data was evaluated following the ENCODE consortium guidelines and

702   metrics (Additional file 3: Table S5) [80]. Peaks were called by using MACS2 v2.1.0.20140616 [77]

703   with low-stringency parameters to obtain a significant list of peaks (--bw 130/135, as determined by

704   the cross-correlation analysis; -g 1.0e9; --to-large; -p 0.025). Irreproducible discovery rate (IDR)

705   analysis was performed on the top 125,000 peaks according to their p-value [80] (parameters:

706   peak.half.width -1; min.overlap.ratio 0; is.broadpeak F; ranking.measure p.value). The final set of

707   TFBS was determined by selecting the number of peaks with an IDR threshold below 0.01 obtained

708   from the pooled-replicate consistency analysis.

Transcription factor occupancy

710   TFBS locations were intersected with regulatory domains and gene features by using BEDtools

711   intersect v2.24.0 [79]. Summits of TFBS located within promoters and enhancers were retrieved and

712   extended ± 250 bp. Extended summits were then merged by using BEDtools merge v2.24.0 [79].

713   Merged regions bound by at least two different TFs were further investigated to measure the genome-

714   wide shared occupancy level between each TF pair. For each TF, the number of binding locations

715   that intersected each of the four remaining TFs were counted separately and compared to its total

716   number of shared binding locations. A similar approach was applied when analysing the shared

717   occupancy level of the TFs in the vicinity of their target genes, albeit only the binding locations of

718   the TFs directly regulating them were considered.

719   **Binding motif analysis**

720   Motif analysis was performed by using DREME v4.11.2 [81] (default parameters) on the 150-bp

721   sequences surrounding the summits (± 75 bp) of the 1,000 most significant TF peaks that overlapped

722   with promoters and enhancers. Recognition motifs thus identified were then compared against motif

723   databases by using Tomtom v4.11.2 [82] (default parameters).

724   **Transcriptional regulatory network**

725   The regulatory network was built on the 189 DE genes that were regulated by at least one TF and

726   associated with the selected signalling pathways using Cytoscape v3.4.0 and the edge-weighted

727   spring-embedded layout [83]. The five TFs were determined as source nodes, while the DE genes

728   were defined as target nodes. Interactions between each source node and its target nodes were marked

729   as direct or indirect whether the differential expression was associated with a functional TFBS or not,

730   respectively.

Declarations

Ethics approval and consent to participate: not applicable.

Consent for publication: not applicable.

Availability of data and material: Sequencing data have been deposited on the Gene Expression Omnibus (GEO) database under the SuperSeries accession number GSE100517.

Competing interests: The authors declare that they have no competing interests.

Authors' contributions: MO, DD and SS designed and conceived the study. MO, MM, GL, SN and MAB performed the experiments and collected the data. MO, MM, GL, SN, MAB, DD and SS analysed the data. STB and BT performed and supervised the RNA-seq procedure. JH performed and supervised the ChIP-seq procedure. MO, DD and SS interpreted the data. MO, DD and SS wrote the manuscript, with comments and approval from all authors.

755   providing access to Galaxy web servers. We thank Sophie Gournet (Institut de Biologie Paris-Seine,

756   Paris, France) for illustrations.

References

757

758 1. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific

759 enhancers. Nat Rev Mol Cell Biol. 2015;16:144–54.

760 2. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control.

761 Nat. Rev. Genet. 2012;13:613–26.

762 3. Nassari S, Duprez D, Fournier-Thibault C. Non-myogenic contribution to muscle development

763 and homeostasis: the role of connective tissues. Front. Cell Dev. Biol. 2017;5:22.

764 4. Kalluri R. The biology and function of fibroblasts in cancer. Nat. Rev. Cancer. 2016;16:582–98.

765 5. Chevallier A, Kieny M, Mauger A. Limb-somite relationship: origin of the limb musculature. J.

766 Embryol. Exp. Morphol. 1977;41:245–58.

767 6. Christ B, Jacob HJ, Jacob M. Experimental analysis of the origin of the wing musculature in

768 avian embryos. Anat. Embryol. (Berl). 1977;150:171–86.

769 7. Gaut L, Duprez D. Tendon development and diseases. Wiley Interdiscip. Rev. Dev. Biol.

770 2016;5:5–23.

771 8. Kardon G. Muscle and tendon morphogenesis in the avian hind limb. Development.

772 1998;125:4019–32.

773 9. Lance-Jones C, Dias M. The influence of presumptive limb connective tissue on motoneuron

774 axon guidance. Dev. Biol. 1991;143:93–110.

775 10. Michaud JL, Lapointe F, Le Douarin NM. The dorsoventral polarity of the presumptive limb is

776 determined by signals produced by the somites and by the lateral somatopleure. Development.

777 1997;124:1453–63.

778 11. Kronenberg HM. Developmental regulation of the growth plate. Nature. 2003;423:332–6.

779 12. Braun T, Gautel M. Transcriptional mechanisms regulating skeletal muscle differentiation,

780 growth and homeostasis. Nat. Rev. Mol. Cell Biol. 2011;12:349–61.

781   13. Huang AH, Lu HH, Schweitzer R. Molecular regulation of tendon cell fate during development.

782   J. Orthop. Res. 2015;33:800–12.

783   14. Murchison ND, Price BA, Conner DA, Keene DR, Olson EN, Tabin CJ, et al. Regulation of

784   tendon differentiation by scleraxis distinguishes force-transmitting tendons from muscle-anchoring

785   tendons. Development. 2007;134:2697–708.

786   15. Schweitzer R, Chyung JH, Murtaugh LC, Brent AE, Rosen V, Olson EN, et al. Analysis of the

787   tendon cell fate using Scleraxis, a specific marker for tendons and ligaments. Development.

788   2001;128:3855–66.

789   16. Lejard V, Blais F, Guerquin MJ, Bonnet A, Bonnin MA, Havis E, et al. EGR1 and EGR2

790   involvement in vertebrate tendon differentiation. J. Biol. Chem. 2011;286:5855–67.

791   17. Guerquin MJ, Charvet B, Nourissat G, Havis E, Ronsin O, Bonnin MA, et al. Transcription

792   factor EGR1 directs tendon differentiation and promotes tendon repair. J. Clin. Invest.

793   2013;123:3564–76.

794   18. Hasson P, DeLaurier A, Bennett M, Grigorieva E, Naiche LA, Papaioannou VE, et al. Tbx4 and

795   Tbx5 acting in connective tissue are required for limb muscle and tendon patterning. Dev. Cell.

796   2010;18:148–56.

797   19. Kardon G, Harfe BD, Tabin CJ. A Tcf4-positive mesodermal population provides a prepattern

798   for vertebrate limb muscle patterning. Dev. Cell. 2003;5:937–44.

799   20. Stricker S, Brieske N, Haupt J, Mundlos S. Comparative expression pattern of Odd-skipped

800   related genes Osr1 and Osr2 in chick embryonic development. Gene Expr. Patterns. 2006;6:826–34.

801   21. Stricker S, Mathia S, Haupt J, Seemann P, Meier J, Mundlos S. Odd-skipped related genes

802   regulate differentiation of embryonic limb mesenchyme and bone marrow mesenchymal stromal

803   cells. Stem Cells Dev. 2012;21:623–33.

804   22. Ahrens PB, Solursh M, Reiter RS, Singley CT. Position-related capacity for differentiation of

37

805   limb mesenchyme in cell culture. Dev. Biol. 1979;69:436–50.

806   23. Kim TK, Shiekhattar R. Architectural and functional commonalities between enhancers and

807   promoters. Cell. 2015;162:948–59.

808   24. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide

809   maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60.

810   25. Hong SH, Rampalli S, Lee JB, McNicol J, Collins T, Draper JS, et al. Cell fate potential of

811   human pluripotent stem cells is encoded by histone modifications. Cell Stem Cell. 2011;9:24–36.

812   26. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, et al. Diversity and

813   complexity in DNA recognition by transcription factors. Science. 2009;324:1720–3.

814   27. Meng X, Brodsky MH, Wolfe SA. A bacterial one-hybrid system for determining the DNA-

815   binding specificity of transcription factors. Nat. Biotechnol. 2005;23:988–94.

816   28. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, et al. Integration of external signaling

817   pathways with the core transcriptional network in embryonic stem cells. Cell. 2008;133:1106–17.

818   29. Sunadome K, Yamamoto T, Ebisuya M, Kondoh K, Sehara-Fujisawa A, Nishida E. ERK5

819   regulates muscle cell fusion through Klf transcription factors. Dev. Cell. 2011;20:192–205.

820   30. Serafini T, Colamarino SA, Leonardo ED, Wang H, Beddington R, Skarnes WC, et al. Netrin-1

821   is required for commissural axon guidance in the developing vertebrate nervous system. Cell.

822   1996;87:1001–14.

823   31. Dominici C, Moreno-Bravo JA, Puiggros SR, Rappeneau Q, Rama N, Vieugue P, et al. Floor-

824   plate-derived netrin-1 is dispensable for commissural axon guidance. Nature. 2017;545:350–4.

825   32. Gao B. Wnt regulation of planar cell polarity (PCP). Curr. Top. Dev. Biol. 2012;101:263–95.

826   33. Gros J, Serralbo O, Marcelle C. WNT11 acts as a directional cue to organize the elongation of

827   early muscle fibres. Nature. 2009;457:589–93.

828   34. Settle SH, Rountree RB, Sinha A, Thacker A, Higgins K, Kingsley DM. Multiple joint and

829   skeletal patterning defects caused by single and double mutations in the mouse Gdf6 and Gdf5

830   genes. Dev. Biol. 2003;254:116–30.

831   35. DiRocco DP, Kobayashi A, Taketo MM, McMahon AP, Humphreys BD. Wnt4/β-catenin

832   signaling in medullary kidney myofibroblasts. J. Am. Soc. Nephrol. 2013;24:1399–412.

833   36. Liang XH, Deng WB, Li M, Zhao ZA, Wang TS, Feng XH, et al. Egr1 protein acts downstream

834   of estrogen-leukemia inhibitory factor (LIF)-STAT3 pathway and plays a role during implantation

835   through targeting Wnt4. J. Biol. Chem. 2014;289:23534–45.

836   37. Laeremans H, Rensen SS, Ottenheijm HCJ, Smits JFM, Blankesteijn WM. Wnt/frizzled

837   signalling modulates the migration and differentiation of immortalized cardiac fibroblasts.

838   Cardiovasc. Res. 2010;87:514–23.

839   38. Howley BV, Hussey GS, Link LA, Howe PH. Translational regulation of inhibin βA by TGFβ

840   via the RNA-binding protein hnRNP E1 enhances the invasiveness of epithelial-to-mesenchymal

841   transitioned cells. Oncogene. 2016;35:1725–35.

842   39. Apte SS. A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1

843   motif (ADAMTS) superfamily: functions and mechanisms. J. Biol. Chem. 2009;284:31493–7.

844   40. Wei J, Liu C, Li Z. ADAMTS-18: a metalloproteinase with multiple functions. Front. Biosci.

845   (Landmark Ed). 2014;19:1456–67.

846   41. Mouw JK, Ou G, Weaver VM. Extracellular matrix assembly: a multiscale deconstruction. Nat.

847   Rev. Mol. Cell Biol. 2014;15:771–85.

848   42. Fernandes RJ, Schmid TM, Eyre DR. Assembly of collagen types II, IX and XI into nascent

849   hetero-fibrils by a rat chondrocyte cell line. Eur. J. Biochem. 2003;270:3243–50.

850   43. de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory

851   landscapes. Nature. 2013;502:499–506.

852   44. Cirulli V, Yebra M. Netrins: beyond the brain. Nat. Rev. Mol. Cell Biol. 2007;8:296–306.

853   45. Xu F, Liu C, Zhou D, Zhang L. TGF-β/SMAD pathway and its regulation in hepatic fibrosis. J.

854   Histochem. Cytochem. 2016;64:157–67.

855   46. Hu B, Phan SH. Notch in fibrosis and as a target of anti-fibrotic therapy. Pharmacol. Res.

856   2016;108:57–64.

857   47. Gao Y, Lan Y, Liu H, Jiang R. The zinc finger transcription factors Osr1 and Osr2 control

858   synovial joint formation. Dev. Biol. 2011;352:83–91.

859   48. Havis E, Bonnin MA, Olivera-Martinez I, Nazaret N, Ruggiu M, Weibel J, et al. Transcriptomic

860   analysis of mouse limb tendon cells during development. Development. 2014;141:3683–96.

861   49. Wang H, Noulet F, Edom-Vovard F, Le Grand F, Duprez D. Bmp signaling at the tips of

862   skeletal muscles regulates the number of fetal muscle progenitors and satellite cells during

863   development. Dev. Cell. 2010;18:643–54.

864   50. Ghosh AK, Quaggin SE, Vaughan DE. Molecular basis of organ fibrosis: potential therapeutic

865   approaches. Exp. Biol. Med. 2013;238:461–81.

866   51. Jiang J, Chan YS, Loh YH, Cai J, Tong GQ, Lim CA, et al. A core Klf circuitry regulates self-

867   renewal of embryonic stem cells. Nat. Cell Biol. 2008;10:353–60.

868   52. Mienaltowski MJ, Adams SM, Birk DE. Tendon proper- and peritenon-derived progenitor cells

869   have unique tenogenic properties. Stem Cell Res. Ther. 2014;5:86.

870   53. Charras G, Sahai E. Physical influences of the extracellular environment on cell migration. Nat.

871   Rev. Mol. Cell Biol. 2014;15:813–24.

872   54. Rozario T, DeSimone DW. The extracellular matrix in development and morphogenesis: A

873   dynamic view. Dev. Biol. 2010;341:126–40.

874   55. Eichmann A, Le Noble F, Autiero M, Carmeliet P. Guidance of vascular and neural network

875   formation. Curr. Opin. Neurobiol. 2005;15:108–15.

876   56. Thorsteinsdóttir S, Deries M, Cachaço AS, Bajanca F. The extracellular matrix dimension of

877 skeletal muscle development. Dev. Biol. 2011;354:191–207.

878 57. Ibrahim DM, Hansen P, Rödelsperger C, Stiege AC, Doelken SC, Horn D, et al. Distinct global

879 shifts in genomic binding profiles of limb malformation-associated HOXD13 mutations. Genome

880 Res. 2013;23:2091–102.

881 58. Solursh M, Ahrens PB, Reiter RS. A tissue culture analysis of the steps in limb chondrogenesis.

882 In Vitro. 1978;14:51–61.

883 59. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis.

884 Nat. Methods. 2012;9:671–5.

885 60. Wilkinson DG, Bailes JA, Champion JE, McMahon AP. A molecular analysis of mouse

886 development from 8 to 10 days post coitum detects changes only in embryonic globin expression.

887 Development. 1987;99:493–500.

888 61. Henrique D, Adam J, Myat A, Chitnis A, Lewis J, Ish-Horowicz D. Expression of a Delta

889 homologue in prospective neurons in the chick. Nature. 1995;375:787–90.

890 62. Antin PB, Pier M, Sesepasara T, Yatskievych TA, Darnell DK. Embryonic expression of the

891 chicken Krüppel-like (KLF) transcription factor gene family. Dev. Dyn. 2010;239:1879–87.

892 63. Pourquié O, Fan CM, Coltey M, Hirsinger E, Watanabe Y, Bréant C, et al. Lateral and axial

893 signals involved in avian somite patterning: a role for BMP4. Cell. 1996;84:461–71.

894 64. Murakami S, Ohki-Hamazaki H, Watanabe K, Ikenaka K, Ono K. Netrin 1 provides a

895 chemoattractive cue for the ventral migration of GnRH neurons in the chick forebrain. J. Comp.

896 Neurol. 2010;518:2019–34.

897 65. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative

898 PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25:402–8.

899 66. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and

900 comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.

901   Nature. 2004;432:695–716.

902   67. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment

903   of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol.

904   2013;14:R36.

905   68. Orgeur M, Martens M, Börno ST, Timmermann B, Duprez D, Stricker S. A dual transcript-

906   discovery approach to improve the delimitation of gene features from RNA-seq data in the chicken

907   model. bioRxiv. 2017; doi:https://doi.org/10.1101/156406.

908   69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence

909   Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

910   70. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning

911   sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

912   71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-

913   seq data with DESeq2. Genome Biol. 2014;15:550.

914   72. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM

915   measure is inconsistent among samples. Theory Biosci. 2012;131:281–5.

916   73. GENE-E. https://software.broadinstitute.org/GENE-E/. Accessed 28 Jun 2017.

917   74. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7:

918   improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium.

919   Nucleic Acids Res. 2010;38:D204-10.

920   75. FASTX-Toolkit: FASTQ/A short-reads pre-processing tools.

921   http://hannonlab.cshl.edu/fastx_toolkit/. Accessed 23 Dec 2016.

922   76. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform.

923   Bioinformatics. 2009;25:1754–60.

924   77. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based

925   analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.

926   78. Kellis M, Wold B, Snyder MP, Bernstein BE, Kundaje A, Marinov GK, et al. Defining

927   functional DNA elements in the human genome. Proc. Natl. Acad. Sci. U. S. A. 2014;111:6131–8.

928   79. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.

929   Bioinformatics. 2010;26:841–2.

930   80. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq

931   guidelines and practices of the ENCODE and modENCODE consortia. Genome Res.

932   2012;22:1813–31.

933   81. Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics.

934   2011;27:1653–9.

935   82. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble W, Maniatis T, Goodbourn S, et al.

936   Quantifying similarity between motifs. Genome Biol. 2007;8:R24.

937   83. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software

938   environment for integrated models of biomolecular interaction networks. Genome Res.

939   2003;13:2498–504.

940   **Figures**
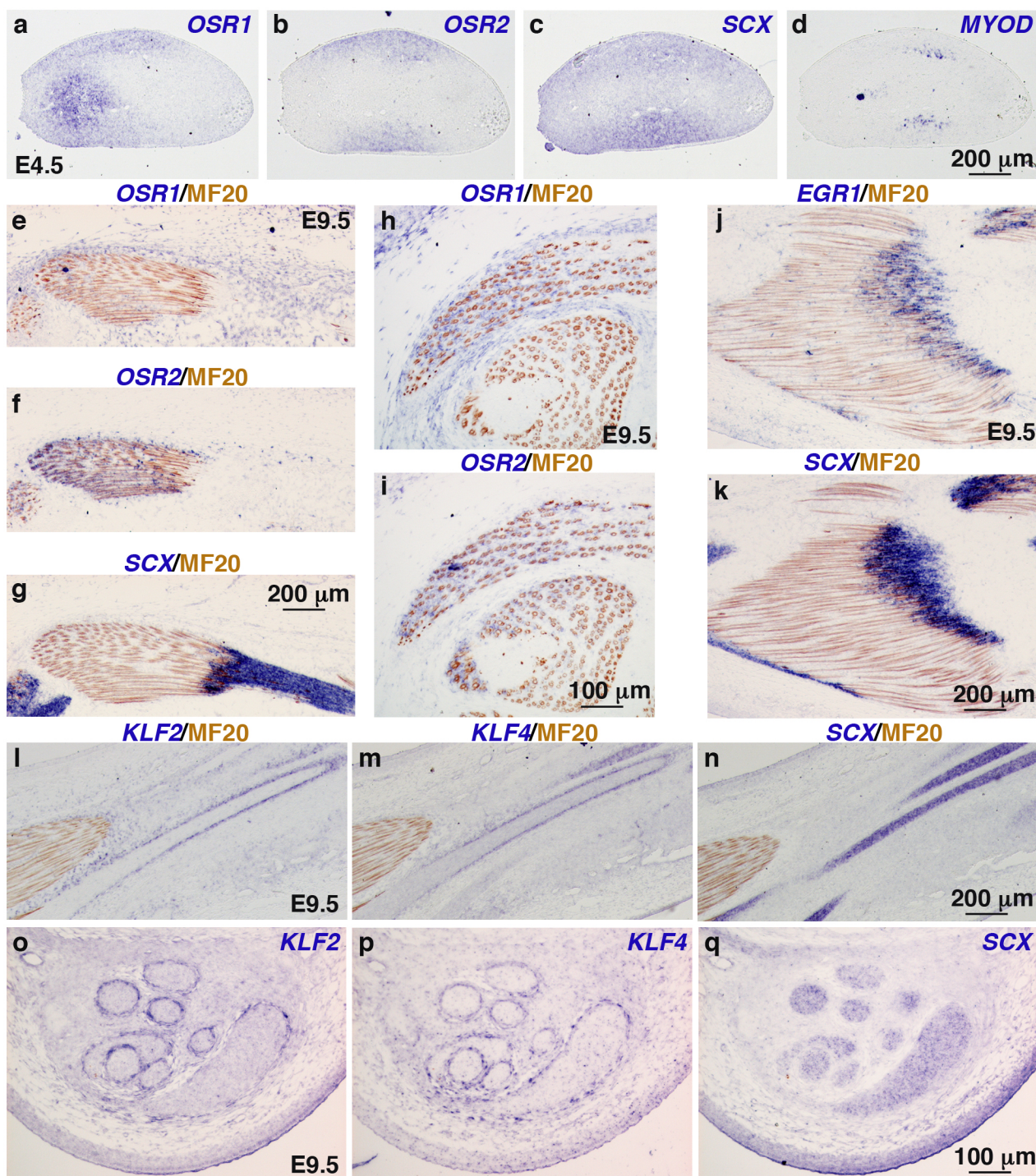
941

942   **Fig. 1** Endogenous expression of CT-associated TFs in hindlimbs of chick embryos. (**a-d**) In situ

943   hybridization to hindlimbs of E4.5 chick embryos. Adjacent and transverse limb sections were

944   hybridized with *OSR1* (**a**), *OSR2* (**b**), *SCX* (**c**) and *MYOD* (**d**) probes (blue). (**e-q**) In situ hybridization

945   to hindlimbs of E9.5 chick embryos followed by immunohistochemistry with the MF20 antibody

946   (brown), which recognizes skeletal muscle myosins. (**e-g**) Adjacent and longitudinal limb sections

44

947   were hybridized with *OSR1* (**e**), *OSR2* (**f**) and *SCX* (**g**) probes (blue). (**h**, **i**) Adjacent and transverse

948   limb sections were hybridized with *OSR1* (**h**) and *OSR2* (**i**) probes (blue). (**j**, **k**) Adjacent and

949   longitudinal limb sections were hybridized with *EGR1* (**j**) and *SCX* (**k**) probes (blue). (**l-n**) Adjacent

950   and longitudinal limb sections were hybridized with *KLF2* (**l**), *KLF4* (**m**) and *SCX* (**n**) probes (blue).

951   (**o-q**) Adjacent and transverse limb sections were hybridized with *KLF2* (**o**), *KLF4* (**p**) and *SCX* (**q**)

952   probes (blue).

953
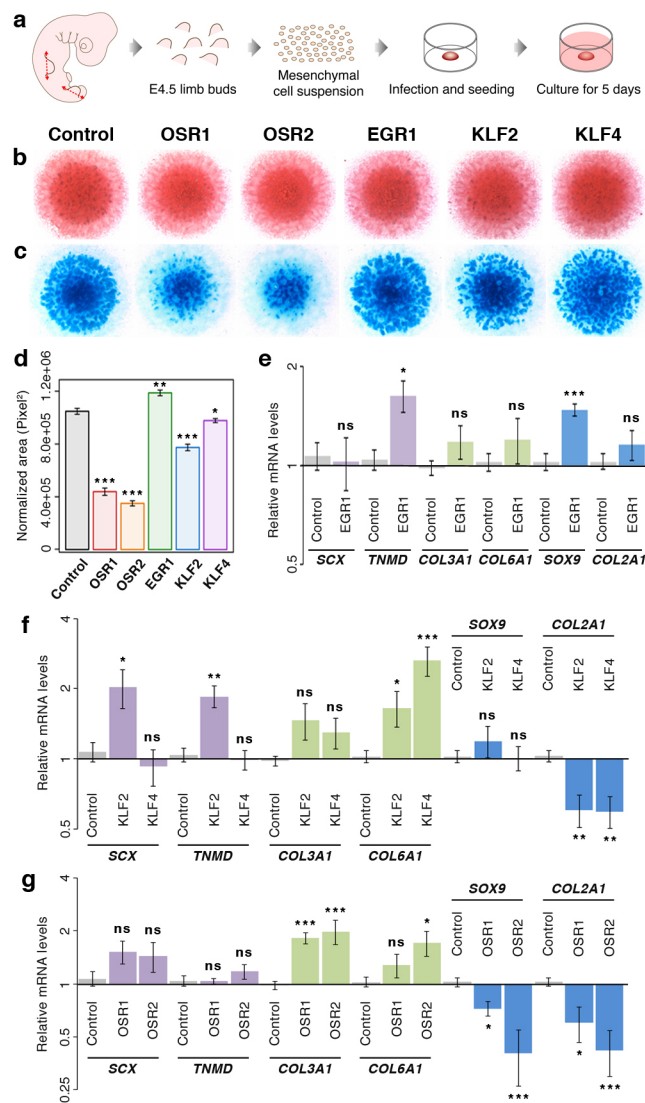
Fig. 2 Differentiation of limb mesenchymal cells following TF overexpression. (**a**) Chick mesenchymal cells were isolated from E4.5 limb buds and cultured in high density for five days. (**b**) Eosin staining of TF-overexpressing chMM cultures. (**c**) Alcian blue staining of cartilage nodules formed in TF-overexpressing chMM cultures. (**d**) Quantification of chondrogenic matrix production: mean ± SEM; paired Student's *t*-test: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (**e-g**) Quantitative RT-PCR analysis of CT marker gene expression upon overexpression of EGR1 (**e**), KLF2/KLF4 (**f**) and OSR1/OSR2 (**g**) in chMM cultures. Graphs depict relative mRNA levels of *SCX* and *TNMD* (tendon markers), *COL3A1* and *COL6A1* (irregular CT markers), and *SOX9* and *COL2A1* (cartilage markers): mean ± SEM; two-tailed Mann-Whitney U test: ns, non-significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.
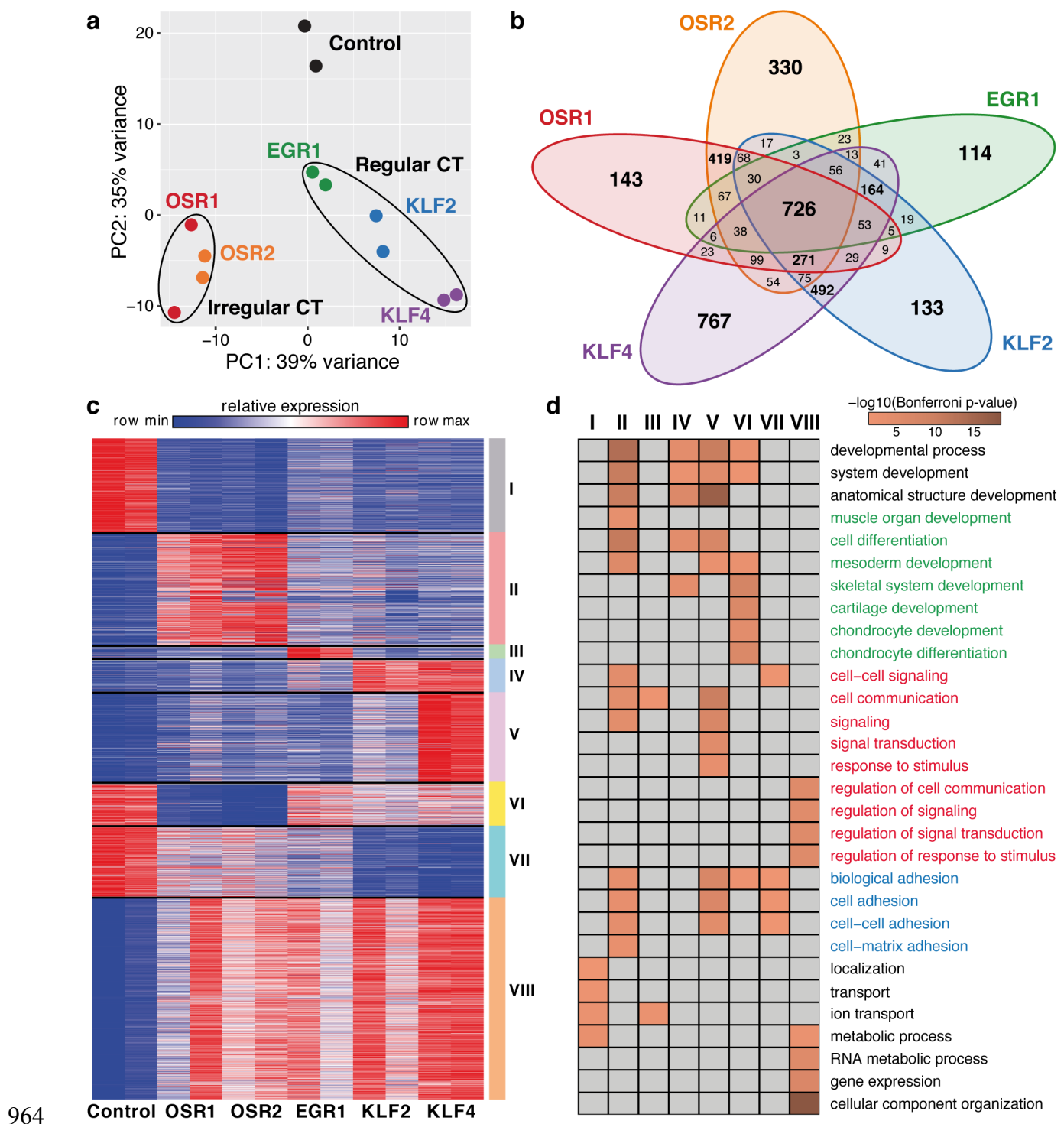
Fig. 3 Gene expression profiles in chMM cultures upon overexpression of CT-associated TFs. (**a**) PCA analysis on global gene expression profiles of TF-overexpressing chMM cultures. (**b**) Venn diagram of the 4,298 non-redundant DE genes detected across all TF-overexpressing chMM cultures. (**c**) Gene clusters identified by *K*-means partitioning on the 4,298 non-redundant DE genes. (**d**) GO analysis for biological processes of the DE genes belonging to each *K*-means cluster. GO terms related to cell differentiation and development are depicted in green, cell signalling and

47

971   communication in red, biological and cell adhesion in blue. Clusters having no significant enrichment

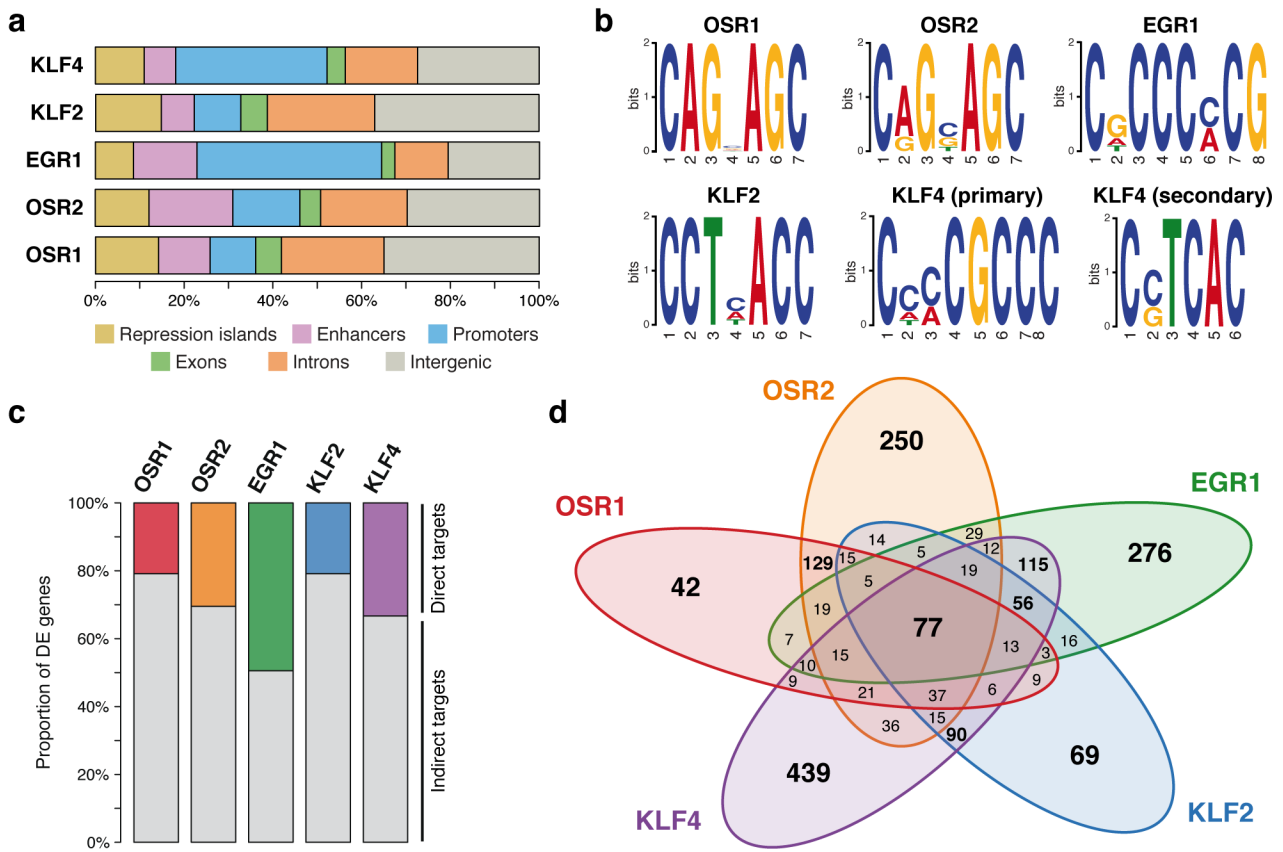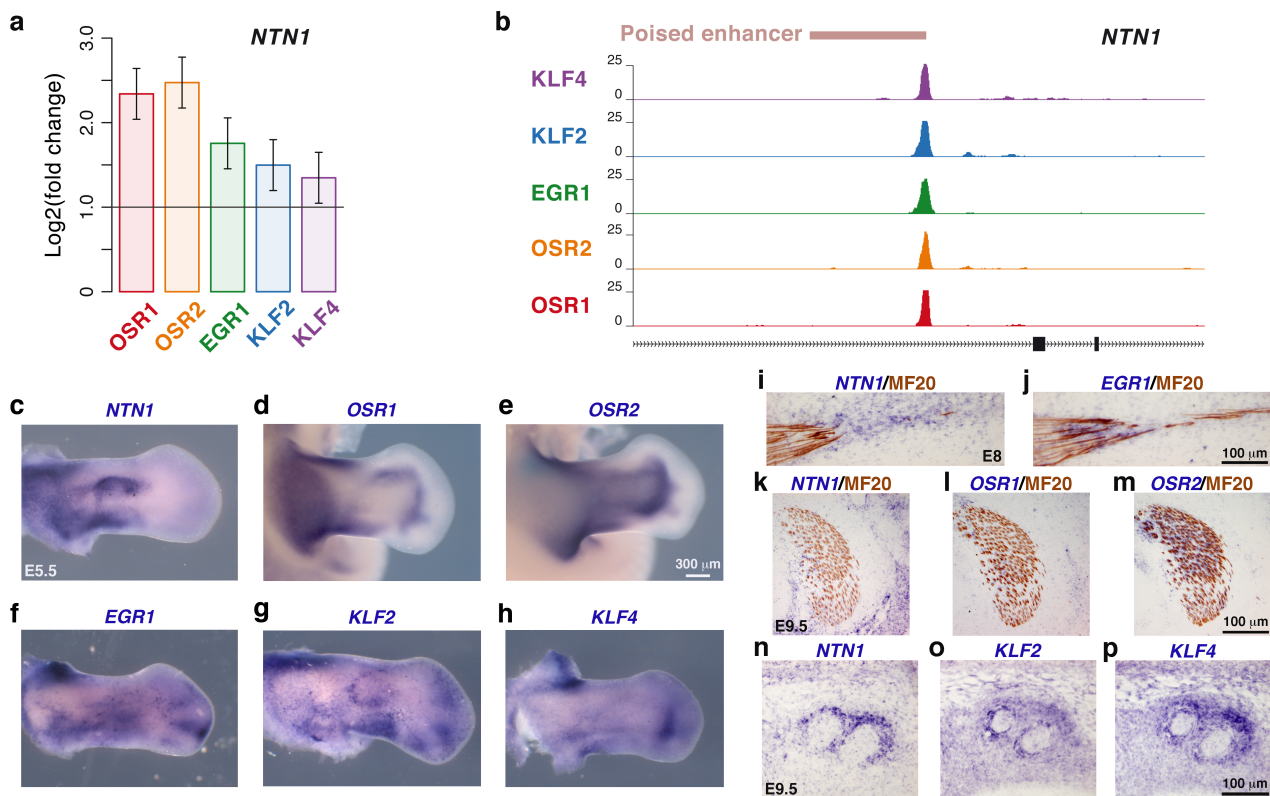972   for the specified GO terms are depicted in grey.

973
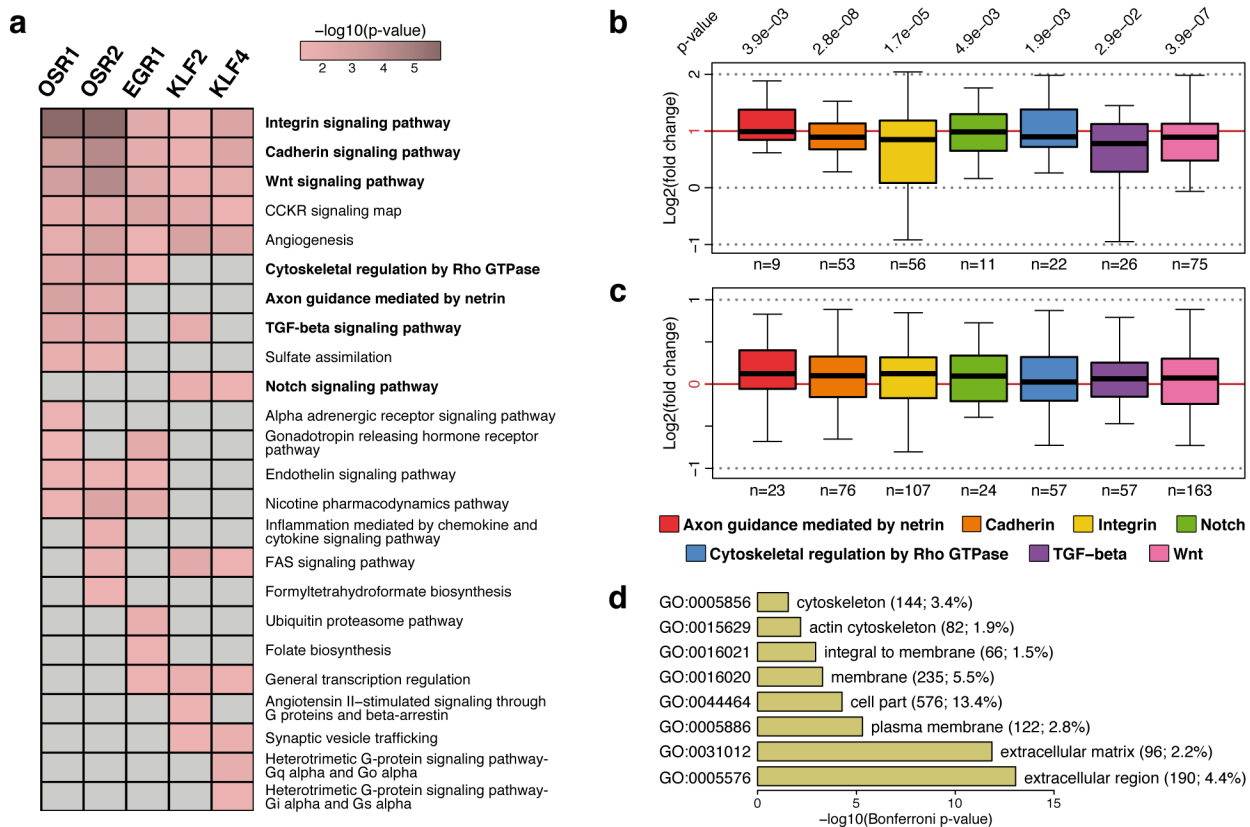
Fig. 4 Signalling signature enrichment analysis of DE genes. (**a**) Panther pathways overrepresented within the DE genes detected upon overexpression of each TF in chMM cultures. DE genes having no enrichment for the specified Panther pathway are depicted in grey. (**b**, **c**) Global expression levels of DE genes (**b**) and non-DE genes (**c**) belonging to the selected Panther pathways. Log2 fold changes of each gene were averaged across all chMM culture conditions and replicates. Number of genes (n) in each Panther pathway is indicated at the bottom of each box. Paired Wilcoxon rank-sum test: p-values are indicated at the top of each box. (**d**) Cellular component GO analysis of the 4,298 non-redundant DE genes.

49
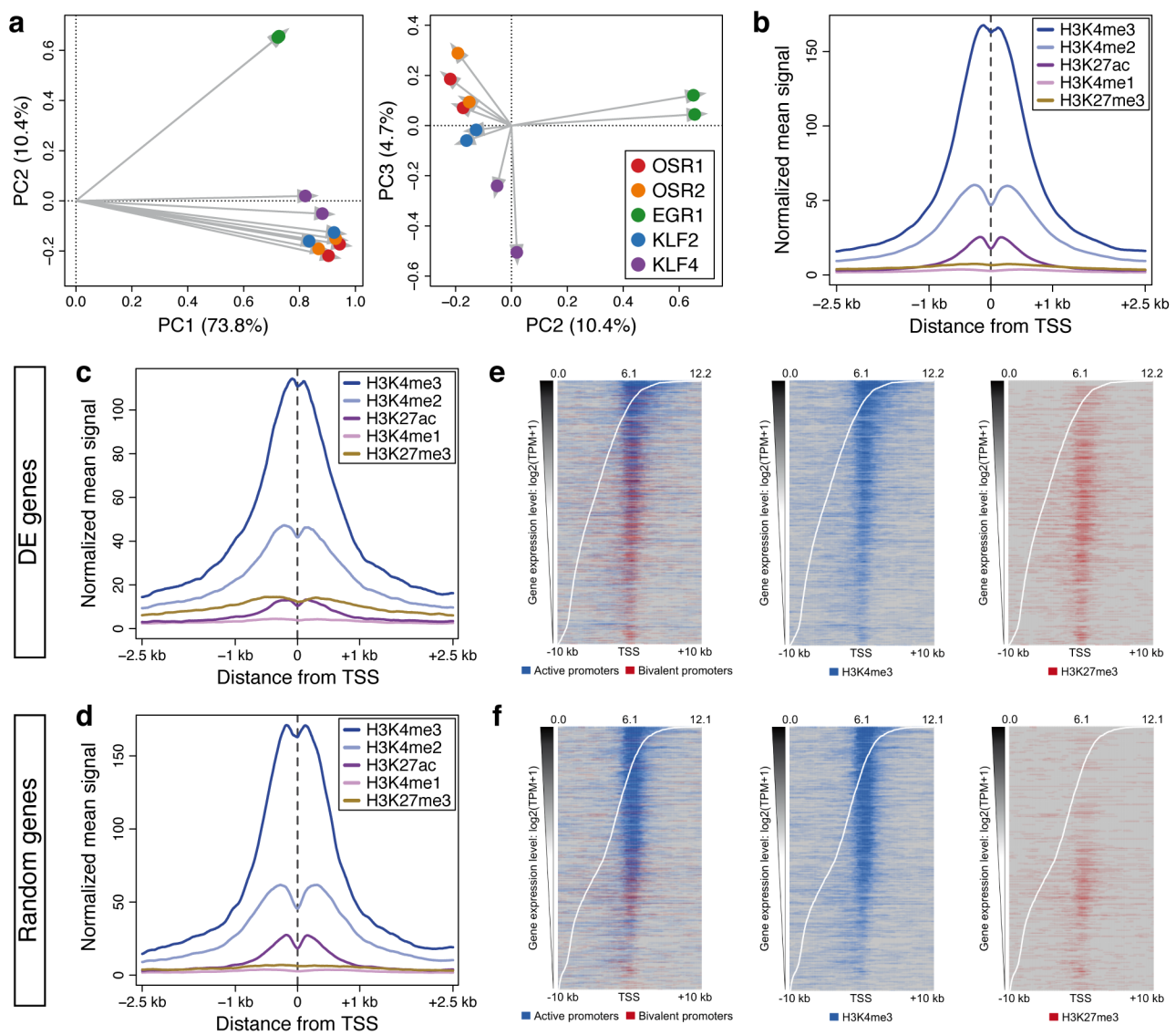
Fig. 5 TF binding profiles and chromatin landscape in chMM cultures. (**a**) PCA analysis on the normalized ChIP-seq signal profiles of all TFs and biological replicates. (**b-d**) Normalized mean histone ChIP-seq signal surrounding the TSS of all genes (**b**), DE genes (**c**) and randomly selected genes (**d**). (**e**, **f**) Distribution of active (blue) and bivalent (red) promoter domains at the TSS of DE genes (**e**) and randomly selected genes (**f**). H3K4me3 (blue) signal is present in active and bivalent promoter domains, whereas H3K4me3 (red) signal is only detected in bivalent promoter domains. Intervals with a main regulatory domain being different from active and bivalent promoter are depicted in grey. Genes were ordered according to their expression levels (white curve).
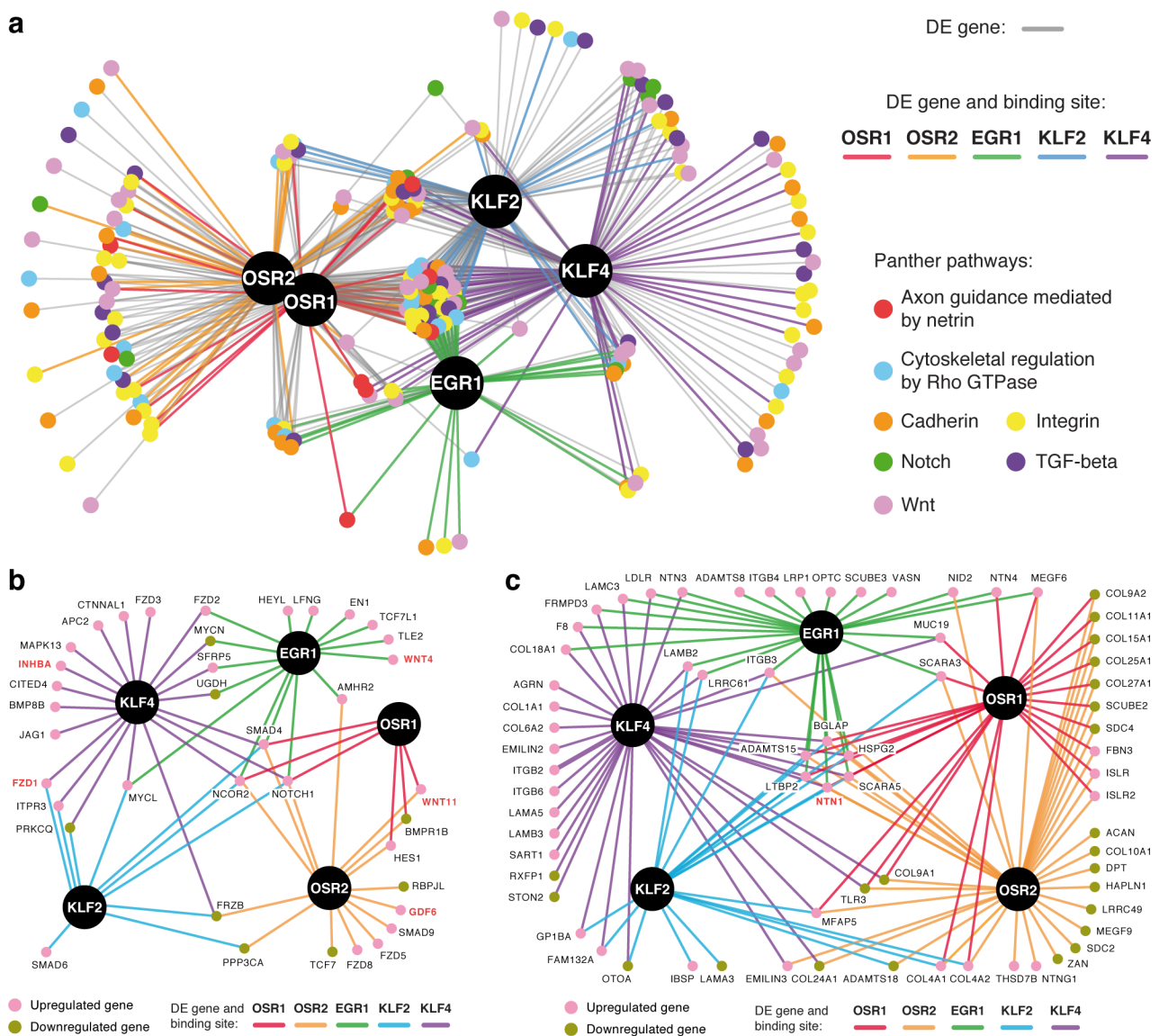
Fig. 6 Regulatory patterns of CT-associated TFs. (**a**) Proportion of TFBS in chromatin domains and gene features. (**b**) TF recognition motifs. (**c**) Proportion of direct and indirect target genes of each TF in chMM cultures. (**d**) Venn diagram of the 1,858 non-redundant direct target genes detected across all TF-overexpressing chMM cultures.

Fig. 7 *NTN1* is a common target gene to the five CT-associated TFs. (**a**) *NTN1* expression levels in TF-overexpressing chMM cultures determined by RNA-seq. (**b**) Binding site within an intronic enhancer of *NTN1* gene for the five TFs identified by ChIP-seq. (**c-h**) Whole-mount in situ hybridization to hindlimbs of E5.5 chick embryos with *NTN1* (**c**), *OSR1* (**d**), *OSR2* (**e**), *EGR1* (**f**), *KLF2* (**g**) and *KLF4* (**h**) probes (blue). (**i, p**) In situ hybridization to forelimbs of E8 (**i, j**) and E9.5 (**k-p**) chick embryos followed by immunohistochemistry with the MF20 antibody (brown). Adjacent and longitudinal limb sections were hybridized with *NTN1* (**i**) and *EGR1* (**j**) probes (blue). Adjacent and transverse limb sections were hybridized with *NTN1* (**k, n**), *OSR1* (**l**), *OSR2* (**m**), *KLF2* (**o**) and *KLF4* (**p**) probes (blue).
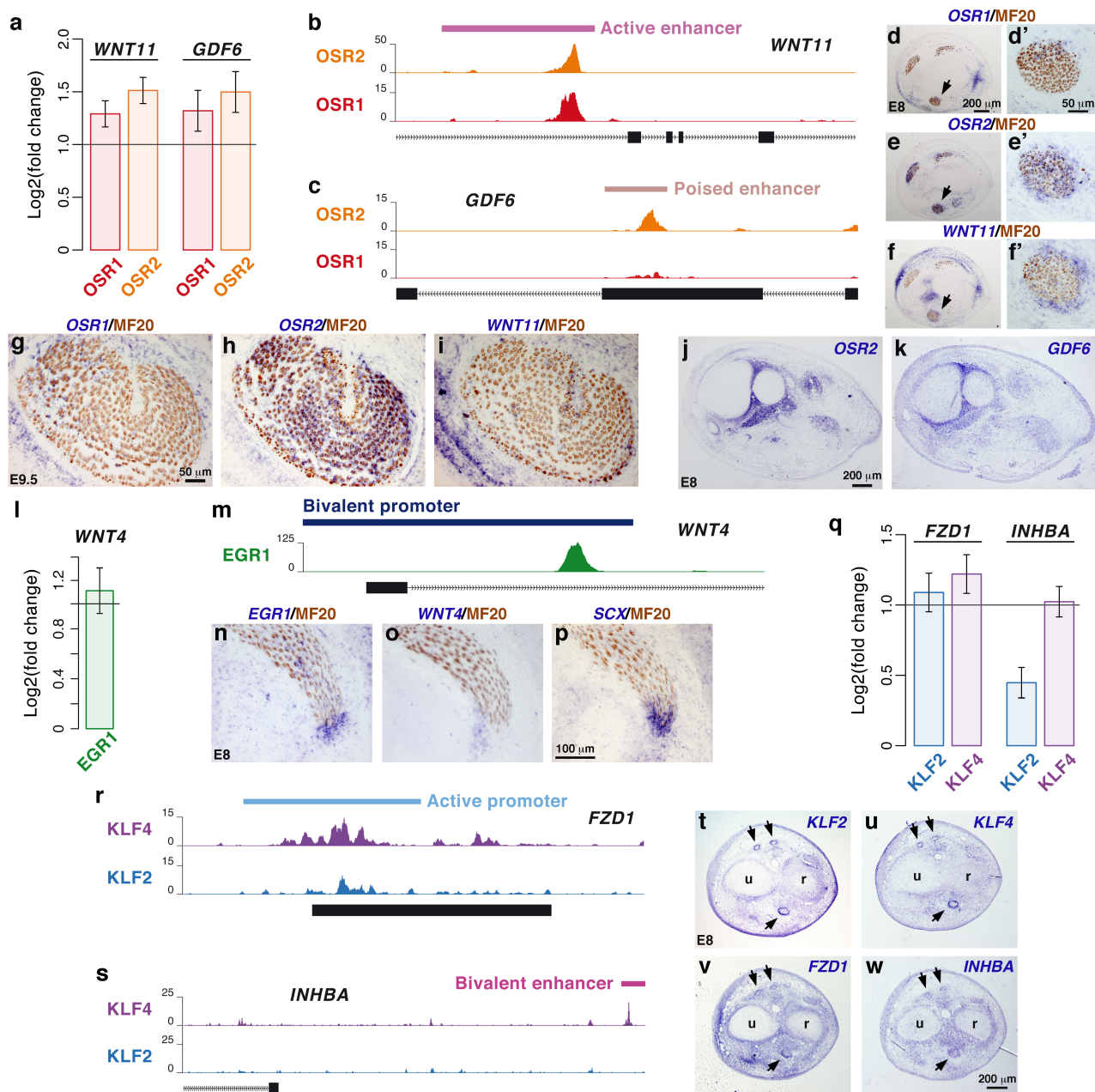
Fig. 8 Selection of target genes encoding signalling molecules downstream of the five CT-associated TFs. (**a**) *WNT11* and *GDF6* expression levels in OSR1- and OSR2-overexpressing chMM cultures determined by RNA-seq. (**b**) Binding site within an intronic enhancer of *WNT11* gene for OSR1 and OSR2 identified by ChIP-seq. (**c**) Binding site within an exonic enhancer of *GDF6* gene for OSR2 identified by ChIP-seq. (**d-i**) In situ hybridization to forelimbs of E8 (**d-f**) and E9.5 (**g-i**) chick embryos followed by immunohistochemistry with the MF20 antibody (brown). Adjacent and transverse limb sections were hybridized with *OSR1* (**d**, **g**), *OSR2* (**e**, **h**) and *WNT11* (**f**, **i**) probes (blue). (**d'**, **e'**, **f'**) are high magnifications of the arrowed muscle in (**d**, **e**, **f**). (**j**, **k**) In situ hybridization

53

1015   to adjacent and transverse forelimb sections of E8 chick embryos with *OSR2* (**j**) and *GDF6* (**k**) probes

1016   (blue). (**l**) *WNT4* expression levels in EGR1-overexpressing chMM cultures determined by RNA-seq.

1017   (**m**) Binding site within the promoter of *WNT4* gene for EGR1 identified by ChIP-seq. (**n-p**) In situ

1018   hybridization to forelimbs of E8 chick embryos followed by immunohistochemistry with the MF20

1019   antibody (brown). Adjacent and transverse limb sections were hybridized with *EGR1* (**n**), *WNT4* (**o**)

1020   and *SCX* (**p**) probes (blue). (**q**) *FZD1* and *INHBA* expression levels in KLF2- and KLF4-

1021   overexpressing chMM cultures determined by RNA-seq. (**r**) Binding site within the promoter of

1022   *FZD1* gene for KLF2 and KLF4 identified by ChIP-seq. (**s**) Binding site within an enhancer located

1023   upstream of *INHBA* gene for KLF4 identified by ChIP-seq. (**t-w**) In situ hybridization to adjacent and

1024   transverse forelimb sections of E8 chick embryos with *KLF2* (**t**), *KLF4* (**u**), *FZD1* (**v**) and *INHBA*

1025   (**w**) probes (blue).

Fig. 9 Regulatory networks of CT-associated TFs. (**a**) Transcriptional regulatory network of the CT-associated TFs and their target genes related to the indicated signalling pathways. Coloured connections correspond to direct interactions between the TFs and their target genes (DE gene and TFBS), while indirect interactions are depicted in grey (DE gene only). (**b**) Network representation of the 38 target genes associated with the Notch, TGF-β and Wnt signalling pathways that are directly regulated by the TFs. (**c**) Network representation of the 70 target genes associated with the ECM that are directly regulated by the TFs. Genes depicted in red correspond to the selected target genes investigated by in situ hybridization.

55

Additional files

1036   **Additional file 1:** Supplementary figures S1-S14. (PDF)

1037   **Additional file 2:** Supplementary table S1 combining RNA-seq and ChIP-seq data. (XLSX)

1038   **Additional file 3:** Supplementary tables S2-S5. (XLSX)