Gene regulatory dynamics during craniofacial development in a carnivorous marsupial

Laura E. Cook[1,2*], Charles Y. Feigin[3], John D. Hills[1], Davide M. Vespasiani[4], Andrew J. Pask[1†], Irene Gallego Romero[5,6,7†]

1. School of BioSciences, University of Melbourne, Royal Parade, Parkville, VIC, Australia
2. Environmental Genomics & Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
3. Department of Environment and Genetics, La Trobe University, Victoria, Australia
4. Epigenetics and Development, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia
5. Human Genomics and Evolution, St Vincent's Institute of Medical Research, Fitzroy, VIC, Australia
6. School of Medicine, University of Melbourne, Royal Parade, Parkville, VIC, Australia
7. Institute of Genomics, University of Tartu, Tartu, Estonia

* To whom correspondence should be addressed: lecook@lbl.gov, irene.gallego@svi.edu.au
† These authors contributed equally

Abstract

Marsupials and placental mammals exhibit significant differences in reproductive and life history strategies. Marsupials are born highly underdeveloped after an extremely short period of gestation, leading to prioritization of the development of structures critical for post-birth survival in the pouch. Critically, they must undergo accelerated development of the oro-facial region compared to placentals. Previously we described the accelerated development of the oro-facial region in the carnivorous Australian marsupial, the fat-tailed dunnart *Sminthopsis crassicaudata* that has one of the shortest gestations of any mammal. By combining genome comparisons of the mouse and dunnart with functional data for the enhancer-associated chromatin modifications, H3K4me3 and H3K27ac, we investigated divergence of craniofacial regulatory landscapes between these species. This is the first description of genome-wide face regulatory elements in a marsupial, with 60,626 putative enhancers and 12,295 putative promoters described. We also generated craniofacial RNA-seq data for the dunnart to investigate expression dynamics of genes near predicted active regulatory elements. While genes involved in regulating facial development were largely conserved in mouse and dunnart, the regulatory landscape varied significantly. Additionally, a subset of dunnart-specific enhancers were associated with genes highly expressed only in dunnart relating to cranial neural crest proliferation, embryonic myogenesis and epidermis development. Comparative RNA-seq analyses of facial tissue revealed dunnart-specific expression of genes involved in the development of the mechanosensory system. Accelerated development of the dunnart sensory system likely relates to the sensory cues received by the nasal-oral region during the postnatal journey to the pouch. Together these data suggest that accelerated face development in the dunnart may be driven by dunnart-specific enhancer activity. Our study highlights the power of marsupial-placental comparative genomics for understanding the role of enhancers in driving temporal shifts in development.

Introduction

The vertebrate head is a highly complex region of the body that plays a key role in an organism's ecology by centralizing numerous structures involved in diet, sensory perception and behavior[1]. Consequently, evolution has modified craniofacial development across lineages, producing a wide array of head morphologies concomitant with the diverse niches that vertebrates occupy. Craniofacial diversity among the major mammalian lineages in particular has long been of great interest, due to the striking differences in their developmental ontology.

Placental mammals are characterized by a long gestation with a high maternal investment during pregnancy with a considerable degree of both orofacial and neurocranial development occurring in the embryo *in utero.* As a result, the placental young experiences little functional constraint during early developmental stages. By contrast, marsupials have a short gestation and give birth to highly altricial young that must crawl to the teat, typically located within the maternal pouch, where they complete the remainder of their development *ex utero*. This unique reproductive method is thought to have imposed strong pressures on the evolution and development of the limbs and head. In particular, marsupials show accelerated development of the nasal cavity, tongue, oral bones and musculature relative to the development of the posterior end of the body[2–5], and generally when compared to placental embryonic development[2,3,6]. Additionally,aspects of the peripheral nervous system appear to be accelerated in the face. We and others have observed large medial nasal swellings early in marsupial development that are innervated and proposed to be necessary for the sensory needs of newborn young in their journey to the pouch[7–9]. Comparative morphometric studies have provided a wealth of evidence that this stark difference in craniofacial development has imposed different regimes of constraint on marsupial and placental mammals[2,3,5,6,10–12] with marsupials in particular showing significantly less interspecies variation in orofacial structures and nasal morphology than placental mammals[12–14]. In spite of these observations, the molecular mechanisms that underlie these differences in early craniofacial development between marsupial and placental remain poorly understood.

Cis-acting regulatory regions have been proposed to play a significant role in morphological divergence in the face, with a number of well-described enhancers that fine-tune face shape in mammals[15]. There is also some evidence of a role for regulatory regions in craniofacial heterochrony in marsupials. One recent study found a marsupial-specific region within a *Sox9* enhancer that drives early and broad expression in pre-migratory neural crest cell domains contributing to early migration of cranial neural crest cells relative to the mouse[16,17]. However, no study has thus far attempted to compare the overall regulatory landscape between marsupials and placentals at developmentally comparable stages. Such surveys have the potential to provide functional insights into the loci controlling craniofacial heterochrony in mammals and consequently the causative evolutionary changes in the genome that have driven the divergent ontogenies of marsupials and placentals.

In recent years, the fat-tailed dunnart (*Sminthopsis crassicaudata*, hereafter referred to as the dunnart) has emerged as a tractable marsupial model species[5,18]. Dunnarts are born after 13.5

days of gestation and craniofacial heterochrony in line with what has been reported in other marsupials is readily observable[2,3,5,6,10–12], making this species an excellent system for comparative studies with placental models. Being one of the most altricial marsupials, the dunnart provides an ideal model in which to study the regulatory landscape during craniofacial development in marsupials, as they rely entirely on the advanced chondrocranium with bony elements of the skeleton not present until approximately 24 hours after birth[5]. To investigate a potential role for regulatory elements in this heterochrony, we used ChIP-sequencing and RNA-sequencing on craniofacial tissue (fronto-nasal, mandibular and maxillary prominences) collected from new-born dunnart pouch young to perform a detailed characterization of chromatin marks during early craniofacial development and then to facilitate comparative analyses with the placental laboratory mouse. Our work provides valuable insights into genomic regions associated with regulatory elements regulating craniofacial development in marsupials and their potential role in craniofacial heterochrony.

Results

Defining craniofacial putative enhancer- and promoter regions in the dunnart

After validating the ability of our antibodies to enrich for dunnart chromatin marks (SupNote1), ChIP-seq libraries were sequenced to average depth of 57 million reads and mapped to a de novo assembly of the dunnart genome generated for this study (Methods for more details). Peak calling with MACS2 ($q < 0.05$) identified 80,989 regions reproducibly enriched for H3K4me3 and 121,281 regions reproducibly enriched for H3K27ac in dunnart facial prominence tissue. As this is the first epigenomic profiling for this species we performed extensive data quality control to ensure the robustness of the data. Similar to previous studies[19–23], we found that H3K4me3 often (62% of all H3K4me3 peaks) co-occupied the genome with H3K27ac (Fig1A) while 50% of H3K27ac-enriched regions were only associated with this mark (Fig1A). Active enhancers are generally enriched for H3K27ac[19,24] while sites of transcription initiation (active promoters) can be identified as being marked by both H3K27ac and H3K4me3[22,23].

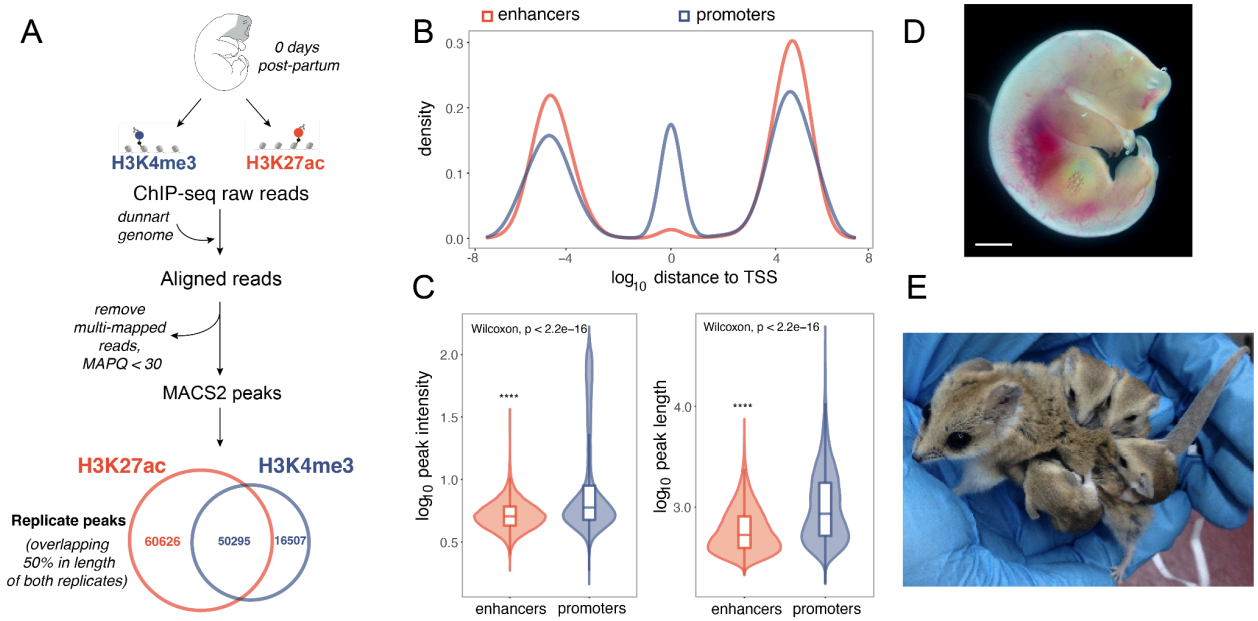


Figure 1. Analysis workflow and quality control of H3K4me3 and H3K27ac peaks in the fat-tailed dunnart (*Sminthopsis crassicaudata*). (A) Drawing of a D0 dunnart pouch young with dissected orofacial tissue shown in gray. Short-read alignment and peak calling workflow and numbers of reproducible peaks identified for H3K27ac (orange), H3K4me3 (blue) for craniofacial tissue. (B) Log_{10} distance to the nearest TSS for putative enhancer (orange) and promoters (blue). (C) Log_{10} of peak intensity and peak length are represented as boxplots and violin plots for putative enhancers (orange) and promoters (blue). Peak intensities correspond to average fold enrichment values over total input DNA across biological replicates. (D) Dunnart pouch young on the day of birth. Scale bar = 1mm. (E) Adult female dunnart carrying four young.

We initially defined promoters as those marked by only H3K4me3 or with >50% of reciprocal peak length for H3K27ac and H3K4me3 peaks, and enhancers as those marked only by H3K27ac, identifying 66,802 promoters and 60,626 enhancers. Enhancers were located on average 77 kb from TSS, while promoters were located on average 106 kb from the nearest TSS, despite there being a greater number of peaks located <1 kb from the TSS (1,008 enhancers versus 9,023 promoters; Fig1B). This was an unexpected finding as a large fraction (0.41) of promoters were located >3 kb from an annotated TSS (see SupNote2). H3K4me3 activity at enhancers is well-established[25,26], however, compared to H3K4me3 activity at promoters, H3K4me3 levels at enhancers are low[27].This is in line with our observations that H3K4me3 levels at enhancers are nearly 7 times lower than those observed at promoter regions (see SupNote2). Distance from TSS is frequently used to filter putative promoters from other elements. Therefore, we grouped peaks characterized as promoters based on their distance to the nearest TSS (see Methods), resulting in 12,295 high-confidence promoters for all of the following analyses (see SupNote3).

Candidate dunnart regulatory elements are associated with highly expressed genes involved in muscle, skin and bone development

Next, we asked what biological processes are associated with active regulatory regions in the dunnart face. To accomplish this, we linked peaks to genes in order to associate functional annotations of coding genes with the candidate regulatory elements that likely regulate their

expression. To make use of resources available in model organisms such as GO databases, we converted all dunnart gene IDs to mouse orthologous genes for downstream applications. This reduced the dataset to 35,677 putative enhancers and 8,589 promoters near (within 1Mbp) genes with a one-to-one ortholog in mouse (SupTable1).

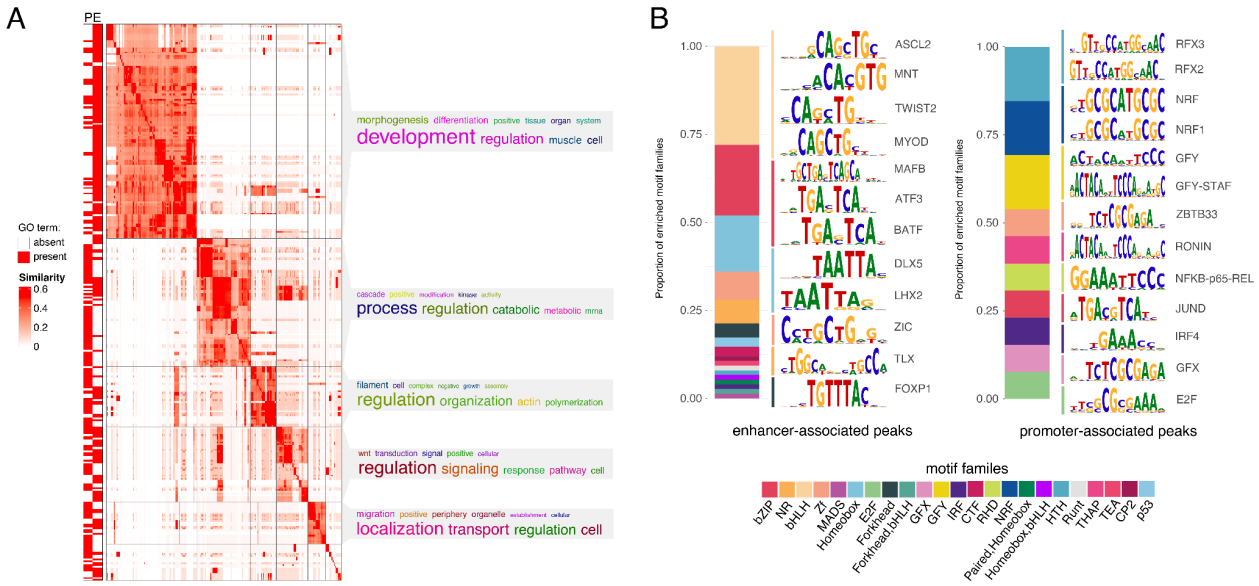


Figure 2. Predicted functional enrichment for dunnart peaks. (A) 304 significantly enriched GO terms clustered based on similarity of the terms. The function of the terms in each group are summarized by word clouds of the keywords. Rows marked by P were driven by genes linked to putative promoters, rows marked by E were driven by genes linked to putative enhancers. (B) Enriched TF motifs for transcription factor families (HOMER). PWM logos for preferred binding motifs of TFs are shown. The letter size indicates the probability of a TF binding given the nucleotide composition.

We found that gene annotations for both promoters and enhancers were enriched for 23% of the same GO terms, including cellular processes (protein localisation to plasma membrane, protein localisation to cell periphery, regulation of cell morphogenesis, positive regulation of cell migration) and development (axon development, camera-type eye development, muscle tissue development, striated muscle development). By contrast, 44% of GO terms were uniquely enriched amongst genes assigned to promoters and were related to mRNA processing, transcription, mRNA stability, cell cycle, and mRNA degradation (Fig2A, SupTable2) anduniquely enriched GO terms for genes assigned to candidate enhancers corresponded to processes indicative of early embryonic development (Fig2A, SupTable3).

Terms related to facial skeleton development were enriched amongst genes assigned to putative dunnart enhancers, including bone cell development, muscle cell development, secondary palate development, roof of mouth development and mesenchyme development, consistent with dunnart craniofacial morphology[5]. Enhancers active near important palate genes[28,29] such as *SHH*, *SATB2*, *MEF2C*, *SNAI2* and *IRF6* in the dunnart at birth may highlight potential regulatory mechanisms driving early palatal closure. In addition, terms related to development of the circulatory system, including regulation of vasculature development, circulatory system process and blood circulation were enriched amongst genes linked to

predicted enhancer regions (for example, *ACE*, *PDGFB*, *GATA4*, *GATA6*, *VEGFA*). This is consistent with observations that show the oral region of newborn dunnarts is highly vascularised, with blood vessels visible through their translucent skin at birth[5].

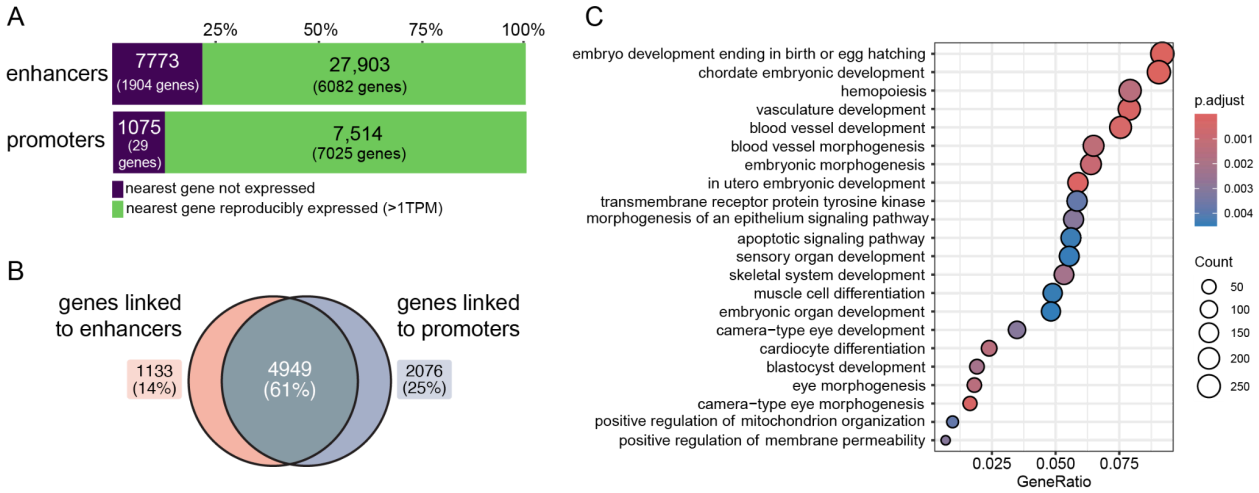


Figure 3. Genes linked to craniofacial enhancers and promoters in the dunnart are reproducibly expressed and involved in embryonic vasculature, muscle, skin and sensory system development. (A) The majority of nearest genes assigned to candidate enhancers and promoters were reproducibly expressed in dunnart face tissue, and (B) reproducibly expressed genes in dunnart were associated with both a promoter and enhancer region. (C) Biological processes enriched for genes medium to highly expressed (>10TPM) and linked to both a promoter and enhancer region (FDR-corrected, $p < 0.01$).

To gain further insight into dunnart gene regulation at this developmental stage, we scanned putative enhancers and promoters for 440 known Homer vertebrate motifs and tested for enriched TFs[30]. Enhancers were significantly enriched for 170 TFs relative to a background set of random GC- and length matched sequences (FDR-corrected, $p < 0.01$), including those with known roles in differentiation of cranial neural crest cells (TWIST, HOXA2), skeletal morphogenesis (DLX5, CREB5, HOXA2), bone development (ATF3, RUNX), cranial nerve development (ATOH1) and/or facial mesenchyme development (LHX2, FOXP1, MAFB; Fig2B, SupFig1). Consistent with the GO enrichment, TFBS in promoter sequences were dominated by transcriptional initiation regulatory sequences, with significant enrichment for 13 TFs (FDR-corrected, $p < 0.01$) including RFX3, RFX2, NRF, NRF1, GRY, ZBTB33, RONIN, JUND and GFX (Fig2B, SupFig1).

Next, we assessed predicted target gene expression by performing bulk RNA-seq in dunnart face tissue collected on the day of birth. There were 12,153 genes reproducibly expressed at a level > 1 TPM across three biological replicates, with the majority of genes expressed (67%; 8158/12153) associated with an active enhancer and/or promoter peak. Most enhancers (78% of all enhancers, SupTable4) or promoter (87% of all promoters, SupTable5) were linked to a reproducibly expressed gene in the dunnart (Fig3A) Additionally, the majority of reproducibly expressed genes near a regulatory peak (61%) were associated with both an enhancer and promoter region (Fig3B), highlighting the correspondence in active regulatory elements and

expressed genes at this time point in the dunnart face. Genes with a medium-to-high expression level at this stage (>10 TPM) and associated with at least one putative enhancer and promoter were enriched for biological processes including "*in utero* embryonic development", "skeletal system development", "muscle tissue development", "skin development", "vasculature development" and "sensory organ development" (Fig3C, SupTable6). Enrichment for the term "*in utero* embryonic development" is indicative of the altricial nature of the dunnart neonate. In a previous study, we showed that in the dunnart ossification begins post-birth (day of birth young corresponding approximately to embryonic day (E) 12.5 in mouse) and that the dunnart neonate instead likely relies on the well-developed cranial muscles and an extremely large chondrocranium for structural head and feeding support during early pouch life[5,31]. Consistent with this, we observed high expression (>20 TPM) of the key head myogenesis genes[32–34], *MYOD1*, *MYF6, MEF2C, PAX3, MYL1 and MYOG*, essential genes regulating chondrogenesis[35,36], *SOX9, COL2A1* and *FGFR1* and genes that act upstream of osteoblast differentiation[37,38], *MSX1, MSX2, CEBPA/G, ALPL, DLX3, DLX5, FGFR1, FGFR2* (SupTable4 and SupTable5).

Comparative analyses of regulatory elements in mouse and dunnart reveal conserved and dunnart-specific enhancers during craniofacial development

Previously, we defined the postnatal craniofacial development of the dunnart and characterized the developmental differences between dunnart and mouse[5]. Marsupials, including dunnarts, have accelerated development of the cranial bones, musculature and peripheral nerves when compared to placental mammals such as the mouse[2–9]. Having now characterized the chromatin landscape of the dunnart's craniofacial region, we next sought to compare it to that of a placental mammal. From our morphological study of dunnart craniofacial bone development, we estimated that craniofacial development in the newborn dunnart is approximately equivalent to E12.5 in mouse[5]. However, we compared the dunnart to all available stages in the mouse to also provide insights into the regulation of additional craniofacial features (e.g. muscle, cranial nerves, sensory system, skin). To do this, we took advantage of publicly available craniofacial ChIP-seq data for H3K4me3 and H3K27ac generated by the mouse ENCODE consortium[39–41] spanning multiple developmental time-points (E10.5-E15.5).

After applying consistent peak calling and filtering parameters to the dunnart ChIP-seq data we found ~17K enhancers and ~10K promoters per stage in the mouse (Fig4A, see SupNote4). The features of mouse predicted enhancer and promoters (percentage CpG, GC content, distance from TSS, peak length and peak intensity) were consistent with the observations in the dunnart (Fig4A, SupFig2B-E). After building dunnart-mm10 liftover chains (see Methods and SupNote5) we compared mouse and dunnart regulatory elements. The alignability (conserved sequence) for dunnart enhancers to the mouse genome was ~13% for 100bp regions (Supplementary Table 22). Between 0.74-6.77% of enhancer regions out of all alignable enhancers were present in both mouse and dunnart (SupTable7B). In contrast, between 45-57% of alignable promoter regions were present in mouse and dunnart (SupTable7A). Although this is a small fraction of the total peaks (~8% for promoters and ~0.5% of enhancers (SupTable7), it suggests that, consistent with the literature[19], promoter regions are more stable over large

evolutionary distances and that shifts in developmental timing of craniofacial marsupial may be more likely to be driven by recently evolved enhancer regions in marsupials.

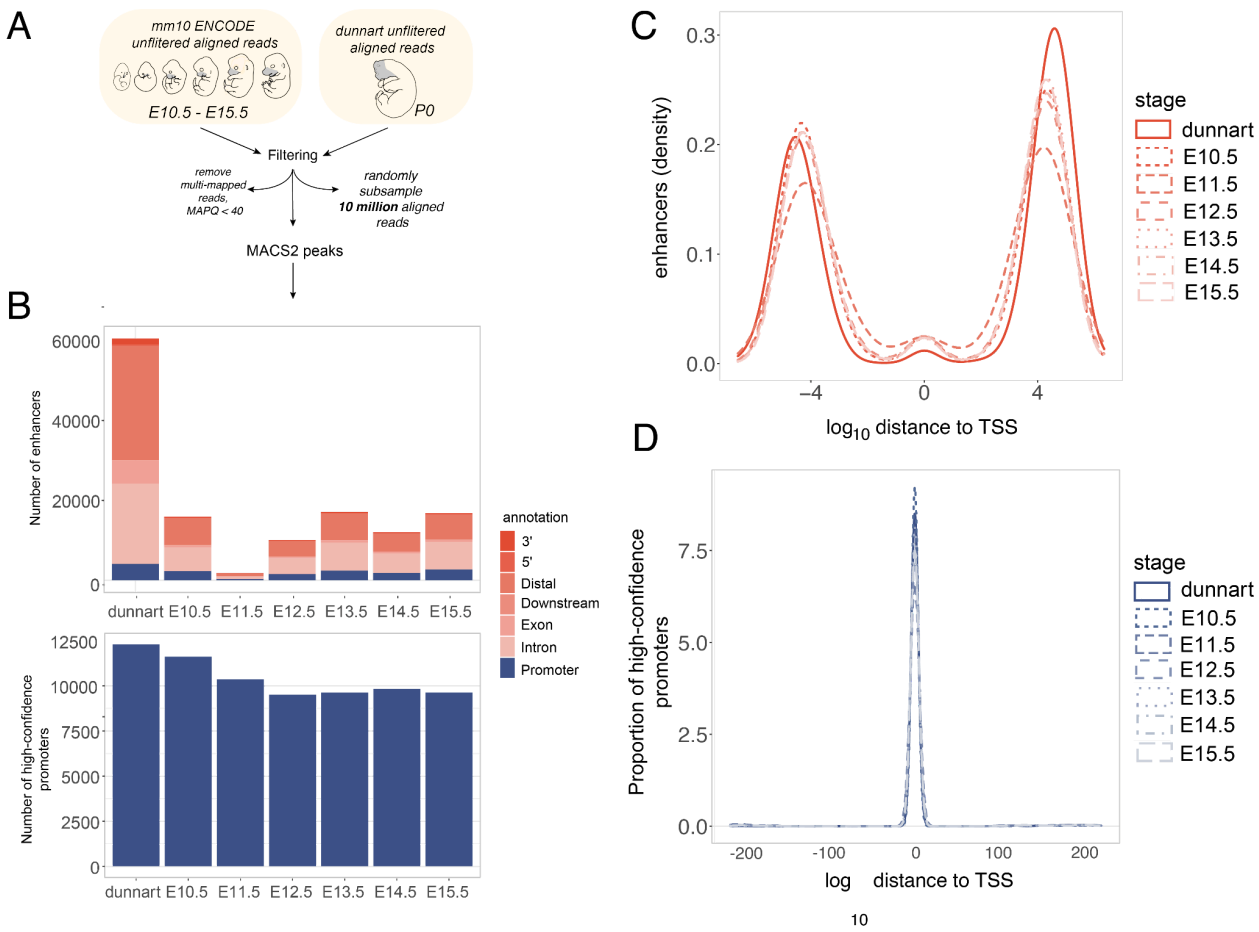


Figure 4. Analysis workflow and features of H3K4me3 and H3K27ac enhancers and promoters in dunnart and mouse. (A) Alignment filtering and peak calling workflow and (B) number of reproducible predicted regulatory elements identified in the dunnart and mouse embryonic stages. Log_{10} of distance to the nearest TSS for putative (C) enhancers and (D) promoters.

To further contextualize the regulatory patterns in the dunnart compared to mouse, we retrieved mouse gene expression data from embryonic facial prominence tissue collected from E10.5 to E15.5 in the mouse ENCODE database[39–41], and incorporated this into our comparative analyses. We considered only one-to-one orthologous genes in mouse and dunnart. Predicted dunnart enhancers with sequence conservation in mouse and matching peak activity at any of the five mouse embryonic stages (461 regions) were located near genes reproducibly expressed (>1 TPM, 89%) in both species (SupTable8). Genes in this group were enriched for core developmental and signaling processes; BMP signaling, cartilage development, ossification, skeletal development and chondrocyte differentiation (SupTable9). Taking the predicted dunnart enhancers alignable to the mouse but without a matching peak, we looked at nearby genes and compared expression between mouse and dunnart. For the 4,311 dunnart-specific enhancers we found that 2,310 (54%) were linked to genes expressed >1 TPM

in all stages and species, suggesting that these genes in mouse could be regulated by a different set of regulatory regions or could be accounted for by the reduced enrichment in the mouse ENCODE ChIP-seq face datasets for H3K27ac (SupFig2). We found a smaller subset (179 regions, 114 unique genes) where the nearest genes were highly expressed (>10 TPM) only in the dunnart with low to no activity in the mouse at any of the embryonic stages (E10.5-E15.5; SupTable10). This included genes involved in cranial neural crest proliferation and migration (*INKA2*, *TFAP2E, OVOL2, GPR161*), keratinocyte proliferation (*PLAU, HOXA1*), embryonic myogenesis (PDF4), development of the ectodermal placodes and sensory systems (*CNGA2, ELF5, EDN1, HOXA1, ATOH1, NPHP4, CFD, WNT2B*) and vomeronasal sensory neuron development (*TFAP2E*; SupTable10).

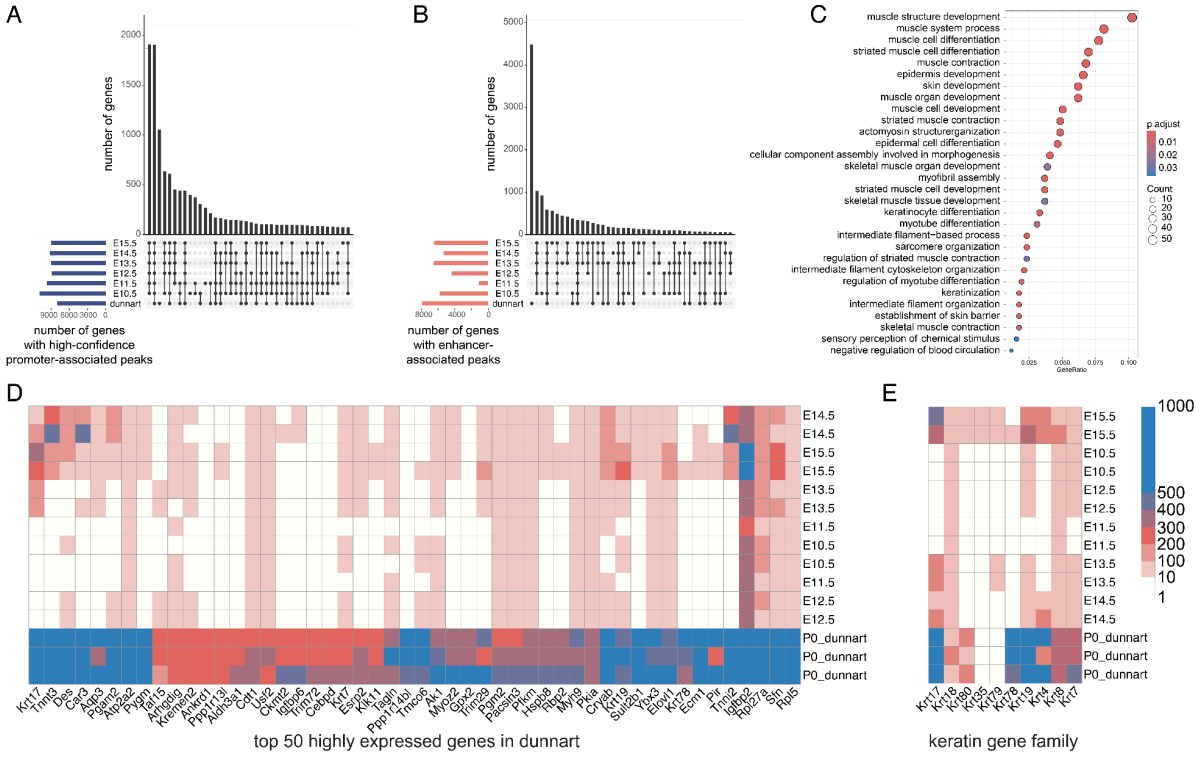


Figure 5. Genes near enhancers highly expressed only in dunnart are involved in development of the skin, muscle and mechanosensory systems. Gene set intersections across mouse (E10.5-E15.5) and dunnart (P0) for (A) genes near promoters and (B) genes near enhancers. (C) Gene ontology term enrichment for top 500 highly expressed dunnart genes. (D) Top 50 highly expressed genes (TPM) in dunnart compared with mouse embryonic stages. Scale bar in E. (E) Expression levels (TPM) for keratin genes across dunnart and mouse.

Dunnart-specific expression of genes involved in the development of the mechanosensory system

Given the large evolutionary distance between the mouse and dunnart and low recovery of alignable regions, we performed a comparison between species at the gene level, by comparing genes assigned to putative enhancers and promoters between the dunnart and mouse. The number of genes that intersect can provide an idea of the similarities in genes and pathways regulated across a larger subset of the total regulatory dataset. The largest intersection size in genes with putative promoters was between the six mouse embryonic stages (1,910 genes;

21.2%) and between the dunnart all six embryonic mouse stages (1,908 genes, 21.2%; Fig5A). Overlap between enhancers was more restricted, with 4,483 predicted target genes (56%) being unique to the dunnart at D0 (Fig5A). The top enriched terms for biological processes were largely shared across dunnart and mouse, with the exception of one GO term, "sensory system development" (SupFig3A). We further investigated this by incorporating gene expression data for mouse and dunnart for genes near putative enhancers. Genes highly expressed in dunnart but lowly or not expressed in mouse (537 total; see SupTable11) were related to three main developmental processes, "epidermis/skin development and keratinization", "sensory system development" and "muscle development and contraction" (Fig5C,D; SupTable12). The majority (70/114) of genes associated with the sequence conserved dunnart-specific enhancers (see SupTable10) overlapped with the list of genes reported here.

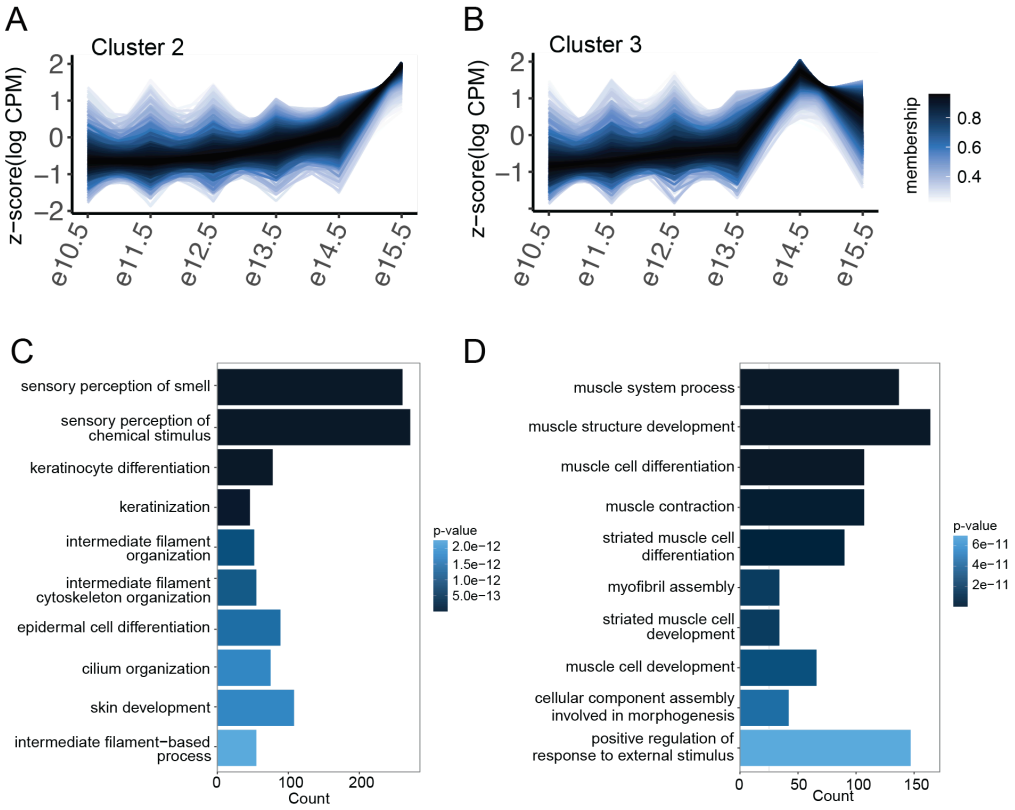


Figure 6. Dunnart expressed genes are associated with two gene clusters with distinct temporal expression patterns in the mouse. (A) Genes in cluster 2 and cluster 3 plotted with their z-scaled temporal expression (logCPM). Color-coding represents membership value (degree to which data points of a gene belong to the cluster). Gene Ontology enrichment for biological processes enriched in (B) cluster 2 and (C) cluster 3 (FDR-corrected $p<0.01$).

Genes critical for development of keratinocytes and the establishment of a skin barrier were highly expressed in dunnart facial tissue with lower expression or no transcripts expressed across the mouse embryonic stages including *IGFBP2, SFN, AQP3, HOPX, KRT17, KRT7, KRT8* and *KRT78* (Fig5D, SupTable12). Keratin genes are also critical for the development of

the mammalian mechanosensory system[42,43]. *Krt17* expressing epidermal keratinocytes are necessary to establish innervation patterns during development and *Krt8* and *Atoh1* expression is required for the specification of the Merkel cells (touch sensory cells)[43]. *KRT17, ATOH1* and *KRT8* are all very highly expressed in dunnart compared to the low expression in mouse facial tissue. Other genes in the keratin family such as *KRT35, KRT79, KRT4* and *KRT80* did not vary greatly between mouse and dunnart (Fig5E). Additionally, we observed high expression of genes involved in development of the olfactory system including *CCK*, *OTX1* and *ISLR* (expressed in olfactory epithelium)[44–47], and Ybx3[47] (expressed in the nasal epithelium; Fig5D, SupTable12). Muscle contraction genes such as *TNNI2*[48] and *TNNT3*[49,50] and genes involved in skeletal muscle development[51–54] (*CAR3, ATP2A2, IGFBP6* and *TRIM72)* were upregulated in dunnart craniofacial tissue (Fig5D, SupTable12). The majority of conserved toolkit genes involved in embryonic development had consistent expression across mouse and dunnart (SupFig4). To explore the relationship between genes expressed in the dunnart face and temporal gene expression dynamics during mouse development, we categorized mouse gene expression into five distinct temporal patterns (SupFig5A). Each of these groups appeared to reflect biological processes occurring during development (SupFig5B). Although dunnart facial development more closely resembles approximately E12.5[5] in the mouse, when compared to the temporal gene expression dynamics during mouse craniofacial dunnart expressed genes were associated with two distinct clusters: a set of genes upregulated specifically at E15.5 in mouse (cluster 2: OR = 1.30, CI = 1.15-1.46; SupFig5C; Fig6A) and a set of genes upregulated at E14.5 (cluster 3: OR = 1.99; CI = 1.78-2.24; SupFig5C; Fig6B). Cluster 2 genes were enriched for functions related to sensory perception, skin development and keratinization (Fig6C) while cluster 3 genes were enriched for muscle development (Fig6D). These results align with our observations in dunnart enhancer and gene expression data, suggesting that shifts in the developmental timing of the skin, muscle and sensory perception may play a role in marsupial early life history.

Discussion

Marsupials display advanced development of the orofacial region relative to development of the central nervous system when compared to placental mammals[3,6]. Specifically the facial skeleton and muscular tissues begin development early relative to the neurocranium differentiation[3,6,31,55–58]. Although development of the central nervous system is protracted in marsupials compared to placentals, marsupials have well-developed peripheral motor nerves and sensory nerves (eg. the trigeminal) at birth[59,60]. Despite increasing descriptions of craniofacial enhancers in model species[15,61–64], the genetic drivers of variation in craniofacial development between mammalian species is largely unexplored. We examined the similarities and differences in genomic regions marked by H3K4me3 and H3K27ac between the dunnart and mouse during early craniofacial development and incorporated comparisons of gene expression between species.

Given the large evolutionary distance (~170 million years) between the mouse and dunnart[65,66] and high turnover of non-coding DNA sequences[67,68], regulatory elements in the dunnart were largely not sequence conserved in the mouse (~8% for promoters and ~0.5% of enhancers). Despite this, regions with conserved activity between the two species were predominantly near genes consistently expressed in both, and were enriched for core craniofacial developmental and signaling processes. Additionally, 54% of dunnart enhancers aligned to the mouse genome but lacking active marks in the mouse were associated with genes expressed throughout all mouse developmental stages, suggesting these genes might be regulated by different enhancers in mice. Moreover, despite differences in experimental methods, when taking a gene level approach there was significant overlap in genes near regulatory elements in both species, with high concordance in enriched biological process terms. The majority of craniofacial developmental genes had highly consistent gene expression levels across the mouse and dunnart, highlighting the substantial conservation of gene regulatory networks driving facial development. This supports the notion that conserved genes and signaling pathways are crucial for mammalian facial development[69].

Although there was generally a high-level of concordance between the dunnart and mouse, we discovered dunnart-specific regulatory elements near highly-expressed genes in the dunnart that were lowly or not expressed in the mouse. These genes were related to three main developmental processes, ”epidermis/skin development and keratinization”, ”sensory system development” and ”muscle development and contraction”. In marsupials, pharyngeal muscles are well differentiated before birth preceding development of both the central nervous system and the skeletal system[70]. Additionally, in the first two days immediately after birth, marsupial pouch young undergo considerable differentiation of the facial muscles[70]. This pattern varies significantly from rodents where the events of skeletal and muscular development generally occur simultaneously[70]. Consistent with this, we found that dunnart expressed genes were associated with a cluster of mouse genes highly expressed at E14.5, a key timepoint for myogenesis in the mouse[71].

Our data also revealed a less evident developmental constraint experienced by marsupials, with dunnart-specific enhancer and gene expression data showing an enrichment of genes related to skin development. Newborn dunnarts exhibit high expression of KRT17 in the face and other genes essential for periderm establishment and maintenance[72–77]. Unlike placental mammals, marsupials retain the periderm post-birth[60,78,79], which may relate to neonatal transcutaneous gas exchange observed in marsupials in the first few days after birth[80,81]. In placental mammals such as mouse and human, the periderm only exists in utero and disappears late in gestation with the eruption of hair and the development of a thick keratinized stratum corneum[72]. The persistence of the periderm aligns with the observation that dunnarts rely heavily on gas exchange through the skin and only begin lung respiration around 5 days postpartum coinciding with the disappearance of the periderm[80,82].

The recurrence of genes related to development of the sensory systems in our comparative analyses spotlights a highly unique aspect of early pouch life in marsupials. Newborn marsupial young require highly developed sensory systems (mechanosensory and olfactory systems) to respond to the cues that guide them to the teat inside the mother's pouch, with the nasal-oral region making regular contact with the mother's belly during the crawl to the pouch[9,83,84]. The snout epidermis of newborn marsupials has been shown to be innervated with presence of mature Merkel cells with connected nerve terminals[7,8,83]. We observe high expression of key genes involved in the development of Merkel cells and sensory placodes uniquely in the dunnart face. In the mouse, Merkel cell development doesn’t begin until later at approximately E16.5[85]. Upregulation of Merkel cell genes in dunnart facial tissue could be a result of prioritization of sensory development required for life outside the womb in a highly underdeveloped state.

In summary, our work suggests that craniofacial developmental constraints in the dunnart may be driven by dunnart-specific gene regulation. Future studies will apply single-cell multi-modal technologies to dunnart tissues to investigate the specific regulatory differences between CNS and orofacial development.

Methods

Tissue Collection

All animal procedures, including laboratory breeding, were conducted in accordance with the current Australian Code for the Care and Use of Animals for Scientific Purposes[86] and were approved by The University of Melbourne Animal Ethics Committee (AEC: 1513686.2) and with the appropriate Wildlife Permit (number 10008652) from the Department of Environment, Land, Water and Planning. Animals were housed in a breeding colony in the School of BioSciences, The University of Melbourne. For details on animal husbandry and collection of dunnart pouch young refer to Cook et al. 2021[5]. Details of pouch young used in this study are presented in SupTable13. After removal of pouch young from the teat, young were killed by decapitation and craniofacial tissue dissected using insulin needles (Becton Dickinson). The tongue, neural tube and eye primordia were removed to limit tissue collected to the facial prominences. Specifically, the mandibular, maxillary and fronto-nasal prominences were collected and snap frozen in liquid nitrogen. All pouch young were determined to be <24 hours old but to account for variability in time since birth we scored pouch young based on head shape. Immediately after birth the dunnart has a flat neurocranium that by approximately 1 day after birth has begun to round (see Cook et al. 2021[5] for more details). We combined craniofacial tissue from 50 pouch young into two replicates ensuring that sex, head shape, and parentage were accounted for (SupTable13).

Immunofluorescence

We assessed the reactivity of the ChIP antibodies in the dunnart using immunostaining on dunnart head sections. Frontal head sections (7 µm thick) on superfrost slides (Platinum Pro, Grale) were deparaffinized and rehydrated according to standard methods[87], followed by antigen retrieval using pH 8.8 unmasking solution (Vector) at 99°C for 30 minutes. We then incubated the sections with either rabbit anti-H3K4me3 primary antibody (1:500 dilution; Abcam ab8580) or rabbit anti-H3K27ac primary antibody (1:500, Abcam ab4729). Sections were then incubated with Alexa Fluor 555 donkey anti-rabbit antibody (1:500 dilution; Abcam in 10% horse serum in PBS with 0.1% Triton X-100; Sigma). All sections were counterstained with 300 nM DAPI to visualize cell nuclei. We observed no staining in the negative controls (no primary antibody). Images were captured using fluorescence microscopy (BX51 Microscope and DP70 Camera; Olympus) and processed in ImageJ v 2.0.0[88].

Chromatin immunoprecipitation and sequencing

Chromatin immunoprecipitation (ChIP) was performed using the MAGnify Chromatin Immunoprecipitation System (Thermo Fisher, 492024) according to manufacturer's instructions. Briefly, frozen dunnart tissue samples (see SupTable13) were diced quickly with two razor blades in cold dPBS (Gibco) followed by crosslinking in 1% formaldehyde solution (Sigma) for 10 minutes. We then added 0.125 M glycine and incubated for 5 minutes at room temperature to neutralize the formaldehyde. Chromatin was fragmented to 300 bp average size by sonication on a Covaris S2 using the following parameters *[duty cycle = 5%, intensity = 2, cycles per burst = 200, cycle time = 60 seconds, cycles = 10, temperature = 4°C, power mode = frequency sweeping, degassing mode = continuous]*. In each ChIP experiment, we used sheared

chromatin from each replicate for immunoprecipitation with antibodies against H3K4me3 (Abcam ab8580) and H3K27ac (Abcam ab4729). An input control was included for each replicate. DNA was purified according to kit instructions using a DynaMag™-PCR Magnet (Thermo Fisher).

We assessed the success of the immunoprecipitation in dunnart craniofacial tissues by performing qPCR for primers designed to amplify genomic regions expected to be occupied by H3K4me3 and/or H3K27ac based on mouse and human enhancers active in facial regions of E11.5 mice in VISTA enhancer[89] and GeneHancer[90] (SupTable14). We also designed primers that amplify regions we predicted would be unoccupied by these histone modifications (enhancers active in heart tissue of E11.5 mice in VISTA enhancer browser[89] (SupTable14). qPCR using SYBR Green Supermix (Bio-Rad Laboratories) was performed in triplicate on a QuantStudio™ 5 System (Thermo Fisher) as per manufacturer's instructions (primers listed in SupTable14). The cycling conditions were *[one cycle of 95°C for 15 s, followed by 40 cycles of 95°C for 15 s, 57°C for 30 s and 72°C for 30 seconds]*. A dissociation curve was also generated for each primer pair. No-template controls were included in triplicate on each plate as a negative control. Regions expected to be enriched in the test sample were quantified by expressing the test sample as a fold change relative to a control sample (no antibody control).

Illumina sequencing libraries were prepared from ChIP-enriched DNA by GENEWIZ (Suzhou, China). Libraries were constructed following the manufacturer's protocol (NEBNext UltraTMII DNA Library Prep Kit for Illumina). For each sample, a minimum of 10 ng of ChIP product was used and libraries were multiplexed and sequenced on an Illumina HiSeq 4000 instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2x150 paired-end configuration to an average depth of 57 million read pairs per sample.

Genome assembly and annotation

Previously there was no fat-tailed dunnart genome available to map ChIP-seq reads against. In order to generate a genome for dunnart, tissue was collected and four sequencing libraries were prepared following methods [91]. Four libraries were generated to improve molecular complexity and genome representation of input DNA. These libraries were then sequenced using the following technologies: Illumina X Ten 2x150 bp, PacBio Sequel I CLR, ONT PromethION and ONT GridION (SupTable15).

Quality trimming and residual adaptor removal from dunnart Illumina libraries (Library 1) was performed using trimmomatic v0.38[92] with options *[PE SLIDINGWINDOW 5:20, MINLEN 75, AVGQUAL 30]*. Contigs were assembled using 200 Gb of PacBio CLR subreads (Library 2) using Flye v2.7[93] with options *[–iterations 4 –trestle –pacbio-raw –genome-size 3.0g]*. Removal of redundant contigs and two rounds of short-read and long-read scaffolding were performed using Redundans 0.14a[94] with options *[--nogapclosing --limit 1.0]*. Inputs for redundans scaffolding were short-insert paired-end reads (Library 1) and 6.5 gigabases of Oxford Nanotechnology reads corresponding to 2 libraries (Library 3 and Library 4). Scaffolds then underwent two rounds of polishing to improve base quality using Pilon v1.23[95] with Illumina

Library 1 as input and with options *[--vcf --diploid --chunksize 10000000 --fix snps,indexls,gaps --minqual 15]*. The resulting genome assembly had a total size of 2.84 Gb and an N50 length of 23 megabases (Mb). We used BUSCO v5.2.2 to assess genome completeness *[v3.0.2, -l mammalia_odb10 -m genome]*. BUSCO gene recovery was 89.9% for complete orthologs in the mammalia_odb10 lineage dataset which includes 9226 BUSCOs. Together, these metrics indicate that the assembly is of comparable completeness and contiguity to other recently-published marsupial genomes[96–102] and therefore represents an excellent resource for downstream functional genomic experiments. The resulting de novo assembly with a resulting scaffold N50 of 23 Mb and a total size of 2.84 Gb making it comparable to other marsupial genomes[96,98–100]. The G+C content of the de novo contigs in the dunnart (~36.25%), was similar to other marsupial species (Tasmanian devil; 36.4%[99], thylacine; ~36%[96] wallaby; 34-38.8%[98], opossum; 37.8%[100], woylie; 38.6%[101] and the brown antechinus; 36.2%[102].

Gene annotations from the high-quality genome assembly of the Tasmanian devil (*Sarcophilus harrisii*, GCF_902635505.1 - mSarHar1.11), which has a divergence time with the dunnart of approximately 40 million years, were accessed from NCBI and then lifted over to dunnart scaffolds using the program liftoff v1.0 [103] with option *[--d 4]*.

Whole genome alignment

To compute pairwise genome alignments between the mouse and dunnart, we used the mouse mm10 assembly as the reference. We first built pairwise alignments using Lastz and axtChain to generate co-linear alignment chains[104], using the previously described Lastz parameters for vertebrates, [*K = 2400 L = 3000 Y = 3400, H = 200]* with the HoxD55 scoring matrix[105]. After building chains, patchChain[105] was applied to extract all the unaligned loci and create local alignment jobs for each one. The new local alignments were combined with the original local alignments for an improved set of chains. We then applied chainCleaner[106] with the parameters *[-LRfoldThreshold = 2.5 -doPairs -LRfoldThresholdPairs = 10 -maxPairDistance = 10000 -maxSuspectScore = 100000 -minBrokenChainScore = 75000]* to improve the specificity of the alignment. After generating an improved set of chains, we applied chainPreNet, chainNet and ChainSubset to filter, produce the alignment nets and create a single chain file using only the chains that appear in the alignment nets[104]. Alignment nets are a hierarchical collection of the chains that attempt to capture orthologous alignments[104]. Chain fragments were joined using chainStitchId and dunnart to mouse chains generated using chainSwap[104]. For quality control, maf files were generated using netToAxt and axtToMaf[104]. Block counts, block lengths and pairwise divergence in the alignments were assessed using MafFilter[107].

ChIP sequencing data analysis

First, we assessed the raw sequencing read quality using FastQC v0.11.9[108]. Raw data were processed by adapter trimming and low quality read removal using Cutadapt v1.9.1[109] [*-q 20 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGA AAGAGTGT --max-n 0.10 -m 75*]. ChIP sequencing statistics for raw reads and trimmed reads are described in SupTable16. Sequencing reads were aligned to the dunnart genome with Bowtie2 v.2.3.5.1[110] *[-q -X 2000 --very-sensitive]*. Unfiltered aligned reads from ChIP-seq experiments performed using mouse embryonic facial prominence for E10.5, E11.5, E12.5,

E13.5, E14.5 and E15.5 were downloaded from ENCODE.org[39–41,111] (accession details described in SupTable17). For both dunnart and mouse aligned reads, low-quality and unpaired reads were removed using Samtools v.1.9[112] *[-q 30 -f 2]* and duplicate reads removed by the MarkDuplicates tool from picard v.2.23.1 (http://broadinstitute.github.io/picard/). Mapping statistics and library complexity for dunnart and mouse reads are described in SupTable18 and SupTable19, respectively. Effective genome size for the dunnart was calculated using the *[unique-kmers.py]* script from khmer v.2.0[113].

Peaks were called on the dunnart-aligned reads using MACS2 v.2.1.1[114] *[-f BAMPE]* and the mouse aligned reads using MACS2 v.2.1.1 with default parameters for mm10, using total DNA input as control and retaining all statistically significant enrichment regions (FDR-corrected $P < 0.01$). Reproducible consensus peaks for biological replicates within a species were determined using the ENCODE3 *[overlap_peaks.py]* script[39,41]. Enriched regions were considered reproducible when they overlapped in two biological replicates within a species by a minimum of 50% of their length using bedtools intersect v2.29.2[115]. Peak-calling statistics for dunnart and mouse are described in SupTable20 and SupTable21, respectively. Similar to Villar et al. 2015[19], we overlapped H3K4me3 and H3K27ac reproducible peaks to determine promoter-associated peaks (marked by only H3K4me3 or both H3K4me3 and H3K27ac) and enhancer-associated peaks (marked only by H3K27ac). H3K4me3 and H3K27ac reproducible peaks were overlapped to determine genomic regions enriched for H3K4me3, H3K27ac or both marks using bedtools intersect v2.29.2[115]. Double-marked H3K4me3 and H3K27ac elements were defined as regions reproducibly marked by H3K4me3 and H3K27ac and overlapping by a minimum 50% of their reciprocal length and were merged with bedtools v2.29.2[115]. Mouse and dunnart peaks called on 10 million aligned reads are deposited at 10.7554/eLife.103592.1.

Gene Ontology Enrichment Analyses

For the dunnart peaks, gene annotations lifted over from the Tasmanian devil annotation were associated with ChIP-seq peaks using the default settings for the annotatePeak function in ChIPseeker v1.26.2[116]. As there is no equivalent gene ontology database for dunnart, we converted the Tasmanian devil RefSeq IDs to Ensembl v103 IDs using biomaRt v2.46.3[119,120], and then converted these to mouse Ensembl v103. In this way we were able to use the mouse Ensembl gene ontology annotations for the dunnart gene domains. We were able to assign Devil Ensembl IDs to 74% of genes with peaks, and mouse IDs to 95% of genes with a devil Ensembl ID. For calculating enrichment in GO in the dunnart, the list of Tasmanian devil genes with an orthologous Ensembl gene in the mouse were used as the background list. All gene ontology analyses were performed using clusterProfiler v4.1.4[117], with Gene Ontology from the org.Mm.eg.db v3.12.0 database[118], setting an FDR-corrected p-value threshold of 0.01 for statistical significance.

Short sequence motifs enriched in dunnart peaks were identified with Homer v4.11.1[30] using *[findMotifsGenome.pl]*. In this case random GC- and length-match sequences for all promoters and enhancers were used as the background set to test for enrichment compared to random expectation. Enriched motifs were clustered into Homer motif families[30].

Mouse and dunnart peak comparison

Dunnart and mouse peaks called from normalized input reads were filtered to 50 bp regions centered on the peak summit. Dunnart peak coordinates were lifted over using liftOver[104] to the mouse (mm10) genome and then back to the dunnart genome. Alignable peaks were kept if after reciprocal liftOver they had the same nearest gene call (SupTable22). Alignable peaks were then intersected with enhancer and promoters at each stage in the mouse with bedtools intersect v.2.30.0[115] to assess peaks with conserved activity (SupTable23).

RNA-sequencing and analyses

The reads for three replicates were generated from facial prominences of P0 dunnart pouch young. Each replicate is a pool of two pouch young. Libraries were prepared using the Illumina stranded mRNA kit and were sequenced to a depth >=50M read pairs (i.e. 25M F + 25M R) in 2x100bp format. The raw reads were first trimmed using trimmomatic v0.39 [92] [ILLUMINACLIP:2:30:10 SLIDINGWINDOW:5:15 MINLEN:50 AVGQUAL:20]. Reads were mapped with hisat2 v2.21[121] [--fr --no-mixed --no-discordant] and alignments were filtered using samtools view[112] [-f 1 -F 2316]. Finally the count table was generated using featureCounts from the Subread package v2.0.1 [122] [-s 2 -p -t exon --minOverlap 10]. Mouse gene count tables were acquired from ENCODE (see SupTable17). Mouse and dunnart gene expression data were normalized for library size with edgeR v.4.0.16[123] and lowly expressed genes filtered (>2 cpm) for at least two out of three replicates.

For mouse gene expression data the resulting gene list was tested for differential expressed genes with the [*DBanalysis*] function in TCseq package[124] which implements edgeR to fit read counts to a negative binomial generalized linear model. Differentially expressed genes with an absolute $\log_2$ fold-change > 2 compared to the starting time-point (E10.5) were determined. The optimal division of clusters was determined using the Calinski criterion implemented with the [*cascadeKM*] function in the vegan v. package[125] with the parameters [*inf.gr = 1, sup.g = 10 iter = 1000, criterion = "calinski"*]. Time-clustering for 5 clusters was performed using the [*timeclust*] function with the parameters: [*algo = 'cm', k = 5, standardize = TRUE*] which performs unsupervised soft clustering of gene expression patterns into five clusters with similar z-scaled temporal patterns. Orthologous genes reproducibly expressed >1 TPM in the dunnart were compared to the list of genes for each cluster using Fisher's Exact Test followed by p-value corrections for multiple testing with the Holm method. Gene Ontology enrichment was performed using [*enrichGO*] as part of the clusterProfiler v4.10.1[117] with the 9933 genes present across the clusters as background and a false discovery rate of 1%.

Data availability

The data generated in this study have been deposited in NCBI's Gene Expression Omnibus[126] and are available through GEO Series accession number GSE188990. Accession IDs for previously published data sets from the ENCODE consortium[39–41] are given in SupTable17. Processed data and an IGV session for browsing is available in a figshare repository (10.7554/eLife.103592.1. All analyses described were carried out using custom bash, Python3 and R v4.1.0 scripts, and are available at github.com/lecook/chipseq-cross-species.

Acknowledgements

This research was supported by an Australian Government Research Training Program (RTP) Scholarship, the David Lachlan Hay Memorial Grant and an Australian Research Council Grant (DP160103683) awarded to A.J.P. St Vincent's Institute acknowledges the infrastructure support it receives from the National Health and Medical Research Council Independent Research Institutes Infrastructure Support Program and from the Victorian Government through its Operational Infrastructure Support Program. The authors thanks Eva Suric, Tania Long and Darren Cipolla for their technical contributions and help with animal husbandry.

Author contributions

L.E.C, I.G.R and A.J.P designed and conceived the study. L.E.C collected the tissue samples, performed the ChIP experiments, analyzed the sequencing data and wrote the manuscript. D.M.V assisted with data analysis. J.D.H and C.Y.F collected tissue samples, performed RNA-sequencing experiments and initial analyses. C.Y.F performed the genome assembly and annotation. I.G.R and A.J.P supervised the study and gave editorial assistance. All authors reviewed and commented on the final manuscript.

Supplementary Notes

Supplementary Note 1. Validation of ChIP-seq in dunnart craniofacial tissue

As the H3K4me3 and H3K27ac antibodies used in this work had not been previously used in marsupials, we first tested their reactivity in the dunnart using immunofluorescence and observed strong positive staining for both antibodies (SupFig6A,B). Therefore, before sequencing, we validated the levels of enrichment in the ChIP libraries using qPCR and primers for sequences in the dunnart that were orthologous to the craniofacial enhancers mm428, mm387, mm423, hs466 and GH07J005561 and an enhancer active in the heart as a negative control, hs222 (SupTable14). All primer sets were enriched as a percentage of the input control, and as expected no enrichment was observed in the rabbit and mouse mock IgG negative control samples (SupFig6C). Although the dunnart ortholog of the human heart enhancer (hs222) was also enriched in both ChIP samples, this enhancer has only been tested at E11.5 in mouse embryos[89] and may have alternative roles at other stages in development or in different species.

Having validated the ability of our antibodies to enrich for dunnart chromatin marks, ChIP-seq libraries were sequenced to average depth of 57 million reads. Reads were mapped to a de novo assembly of the dunnart genome generated for this study (SupTable15), which we annotated using the publicly available high-quality genome assembly of the Tasmanian devil (mSarHar1.11; see Methods for additional details). After mapping and filtering, we retained an average of 43 million mapped reads per library (SupTable16). All samples had a NRF of between 0.76-0.85 (SupTable18)[41]. Read coverage was highly correlated for replicates within each mark (Pearson's $r > 0.99$) and also between marks (Pearson's $r > 0.88$; SupFig7A). Read coverage for replicate input controls was also correlated, but as expected due to enrichment for specific genomic regions in IP samples, there was no correlation in read coverage observed between the input controls and IP samples (Pearson's r range = 0.03 to 0.08, SupFig7A). Similarly, alignment BAM files for H3K4me3 and H3K27ac IP replicates show strong and specific enrichment indicated by a prominent and steep rise of the cumulative sum towards the highest rank, whereas input control alignment BAM files are roughly diagonal, highlighting even coverage across the genomic windows (SupFig7B). Fraction of reads in peaks was approximately 60% for consensus H3K27ac and H3K4me3 peaks (SupTable20).

Supplementary Note 2. Investigating promoters located >3 kb from an annotated TSS.

To investigate the large number of putative promoters located >3kb from an annotated TSS, we first looked at the relationship between the number of peaks per gene and the distance to the next closest gene. After annotating peaks with the nearest gene call using ChIPseeker[116], 72% of genes had at least 1 predicted promoter or enhancer, with 81 genes having more than 50 promoters (SupFig8A). The number of promoters per gene was weakly positively correlated with distance to the next closest gene (Pearson's $r = 0.36$, $p < 2.2 \times 10^{-16}$), suggesting that this observation was partially due to the presence of unannotated transcripts between genes (SupFig8B). Enhancers can also be associated with H3K4me3[124]. However, enrichment levels (peak intensity) tend to be lower than that of H3K27ac[127]. Hence, peaks characterized as promoters that were located more than 3 kb from a known TSS may instead represent active

enhancer regions[127]. To assess this we compared mean peak intensity in H3K4me3 peaks located greater than 3 kb from the nearest TSS, to H3K4me3 peaks located within 3 kb of the TSS. We found that H3K4me3 peaks located closer to the TSS had a stronger peak signal (mean = 46.10) than distal H3K4me3 peaks (mean = 6.95; Wilcoxon FDR-adjusted $p < 2.2 \times 10^{-16}$). This suggests that although some distal promoter peaks may be due to missingness in the annotation, the majority likely represent peaks associated with enhancer regions.

Supplementary Note 3. Defining high-confidence promoter regions

This analysis identified 3 clusters: upstream distal peaks (cluster 3, 31,285 peaks), downstream distal peaks (cluster 2, 22,863 peaks) and peaks centered on the TSS (cluster 1, 12,295 peaks; SupFig9B). Cluster 1 peaks had a higher intensity (SupFig9C) and length (SupFig9D) when compared to clusters 2 and 3 (Wilcoxon FDR-adjusted $p < 2.2 \times 10^{-16}$) and compared to all predicted enhancers and promoters (Wilcoxon FDR-adjusted $p < 2.2 \times 10^{-16}$). Furthermore, peaks in cluster 1 (those closest to annotated TSS) had a higher GC and CpG island content than clusters 2 and 3 (Wilcoxon FDR-adjusted $p < 2.2 \times 10^{-16}$) and higher than enhancers (Wilcoxon FDR-adjusted $p < 2.2 \times 10^{-16}$), consistent with known features of mammalian promoters (SupFig9E,F). H3K4me3 ChIP-seq peak intensity has previously been correlated with transcriptional activity[128] and a higher peak intensity has also been observed in promoters in mammalian liver tissue[19]. We thus used cluster 1 (12,295 peaks) to define a set of high-confidence promoters for all of the following analyses.

Supplementary Note 4. Investigating differences in the number of mouse and dunnart peaks called.

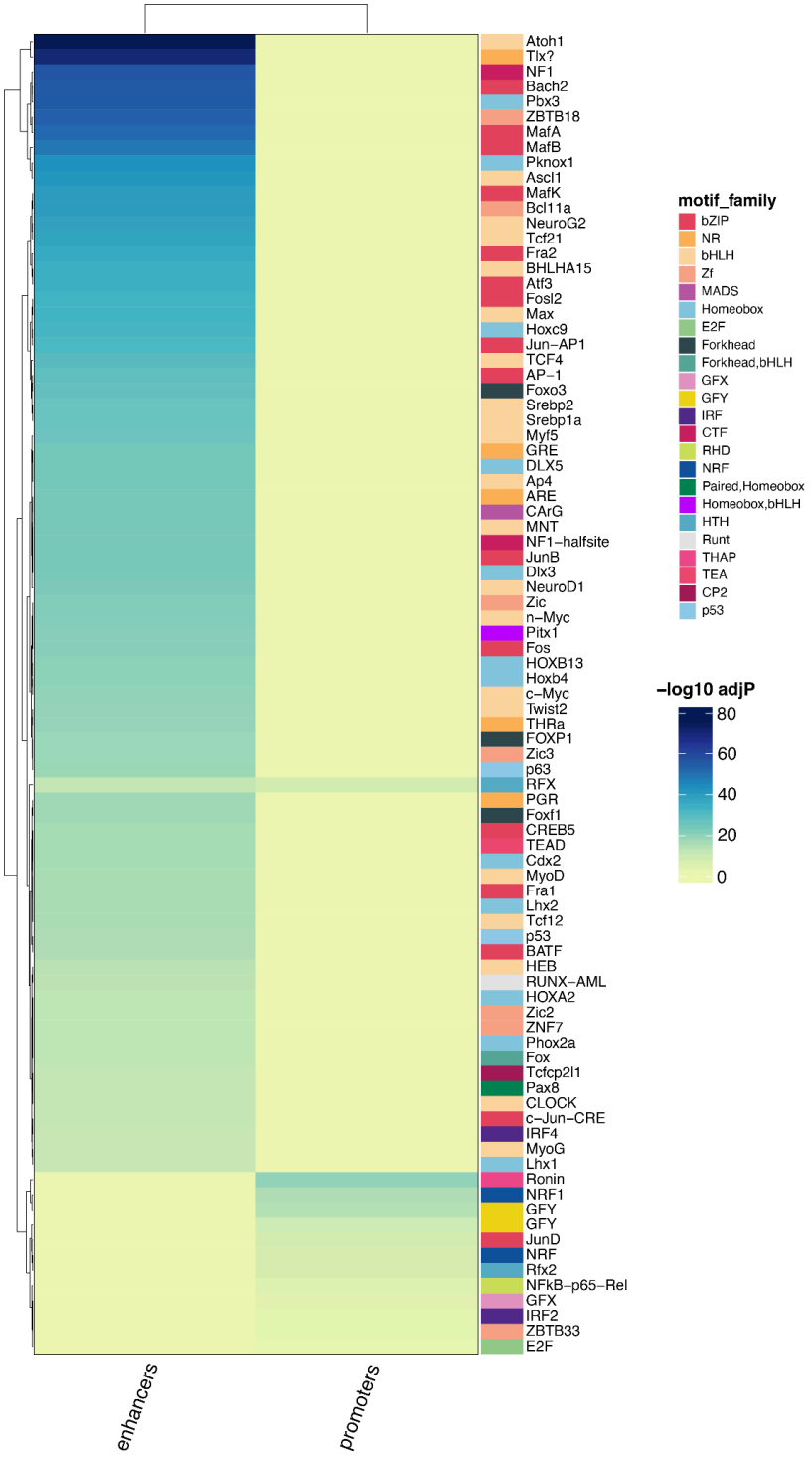
After applying consistent peak calling and filtering parameters to the dunnart ChIP-seq data we found that the number of peaks were fairly consistent across mouse embryonic stages (~20-30,000 total peaks; Fig4A). This number was significantly lower than the total number of peaks observed in the dunnart (~150,000 peaks; Fig4A). We investigated this further and found that the strength and specificity of enrichment differed between the mouse and dunnart datasets. Dunnart alignment BAM files for H3K4me3 and H3K27ac immunoprecipitation replicates show strong and specific enrichment (SupFig7B); however, mouse alignment BAM files show a weaker enrichment with the immunoprecipitation samples for H3K27ac being closer to the input (SupFig10). This was also reflected in the similarity between mouse IPs and input controls based on read coverage, with correlation coefficients higher (H3K4me3, Pearson's r range = 0.17 to 0.32, SupFig11B; H3K27ac, Pearson's r range = 0.39 to 0.64, SupFig11A) than the corresponding dunnart values (SupFig7A). Therefore, this is likely a technical confounder that may be responsible for the lower numbers of enriched regions called in the mouse, as peaks with lower enrichment signal might not be identified by the peak caller. This was consistent with the distribution of peak enrichment values in the dunnart and mouse for H3K4me3, with a lower mean peak enrichment value in the dunnart compared to all mouse stages (Kruskal-Wallis FDR-corrected $p = 6.7 \times 10^{-12}$; SupFig2A). However, this was not the case for peak enrichment values for H3K27ac peaks (SupFig2A), which generally have lower peak enrichment values than observed in H3K4me3[19,127].

Supplementary Note 5. Whole genome alignment between mouse and dunnart

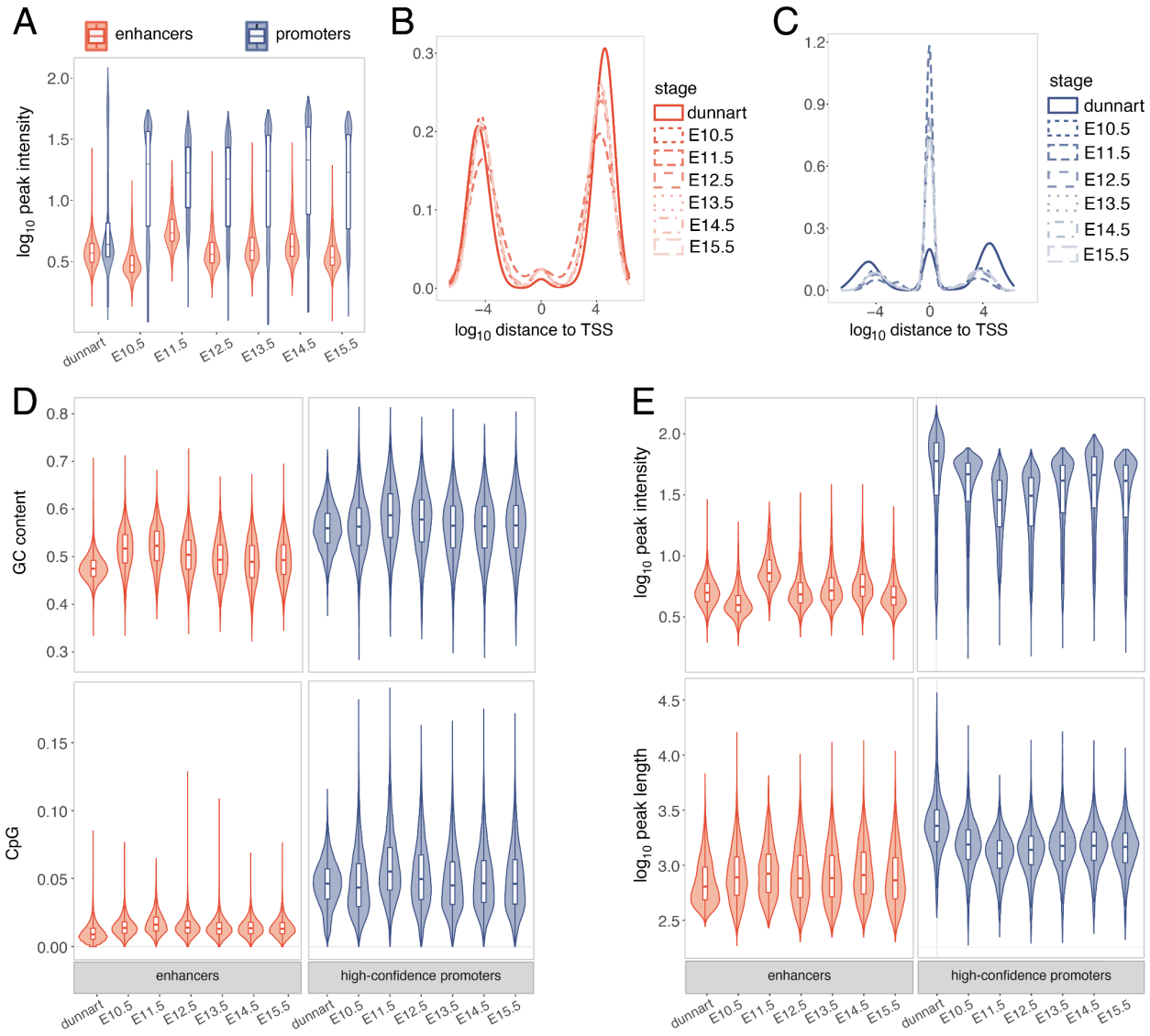
In order to compare activity in orthologous mouse and dunnart peaks, liftOver chains are needed to find the corresponding coordinates in the alternative genome. After building a dunnart-mm10 liftOver chain, we examined the quality of the WGA between mouse and dunnart. Approximately a quarter of both genomes were recovered in the WGA, with 26% of the dunnart genome and 28% of the mouse genome were present with an average mismatch between the dunnart and mouse of 37.2% (SupFig12). Previous WGAs between placentals and marsupials (opossum versus human, tammar wallaby versus human) have recovered approximately 6-8% of the genomes with LastZ[129]. The higher recovery in our WGA may be due to the use of new genome alignment tools like chainCleaner and patchChain. Given the higher conservation of coding regions than non-coding regions, exon coverage is a useful metric for assessing the quality of the alignment. 92% of mouse exon sequence and 69% of dunnart exon sequence was present in the WGA (SupFig12). Coverage of the exons recovered after liftOver were 82.7% and 57.8% in the mouse and dunnart, respectively (SupFig12). The lower exon coverage in the dunnart could be due to the missingness in the dunnart annotation. The majority of the alignment blocks fell within the 100 bp to 1000 bp size range with a total of 578.3 Mb of DNA sequence alignment out of a total of 801 Mb in the alignment within this range (SupFig12). This is very similar to the existing opossum-human WGA where 83% of alignment blocks fall within the size range of 100 to 1000 bp[129].

We then used the liftOver chains to find corresponding coordinates for dunnart peaks in the mouse genome. Given the majority of the alignment blocks for the mouse and dunnart WGA fell between 100-1000 bp in length ((SupFig12C), we first assessed the impact of peak length (centered on the peak summit, which represents the point with the highest signal in the peak region) on inter-species recovery of peaks. We defined dunnart peak regions as alignable if they could be reciprocally lifted over to the mouse genome and then back to the dunnart genome. Unsurprisingly, for both enhancers and promoters the number of alignable peaks recovered decreased with increasing peak length (SupTable22 and SupTable23). Promoters had the highest number of regions alignable (17%) using 50 bp peak summit widths which decreased to 1.8% with a 500 bp peak summit length (SupTable23). Similarly, 19% of enhancers were alignable using a 50 bp peak summit length and this decreased to 2% with 500 bp peak summit length (SupTable22). To assess the accuracy of the reciprocal liftOver, we next investigated whether alignable peaks in the mouse and dunnart mapped to the same nearest Ensembl gene before and after liftOver. Accuracy was high for both the promoters and enhancers (from 94% to 100%; (SupTable22 and SupTable23).

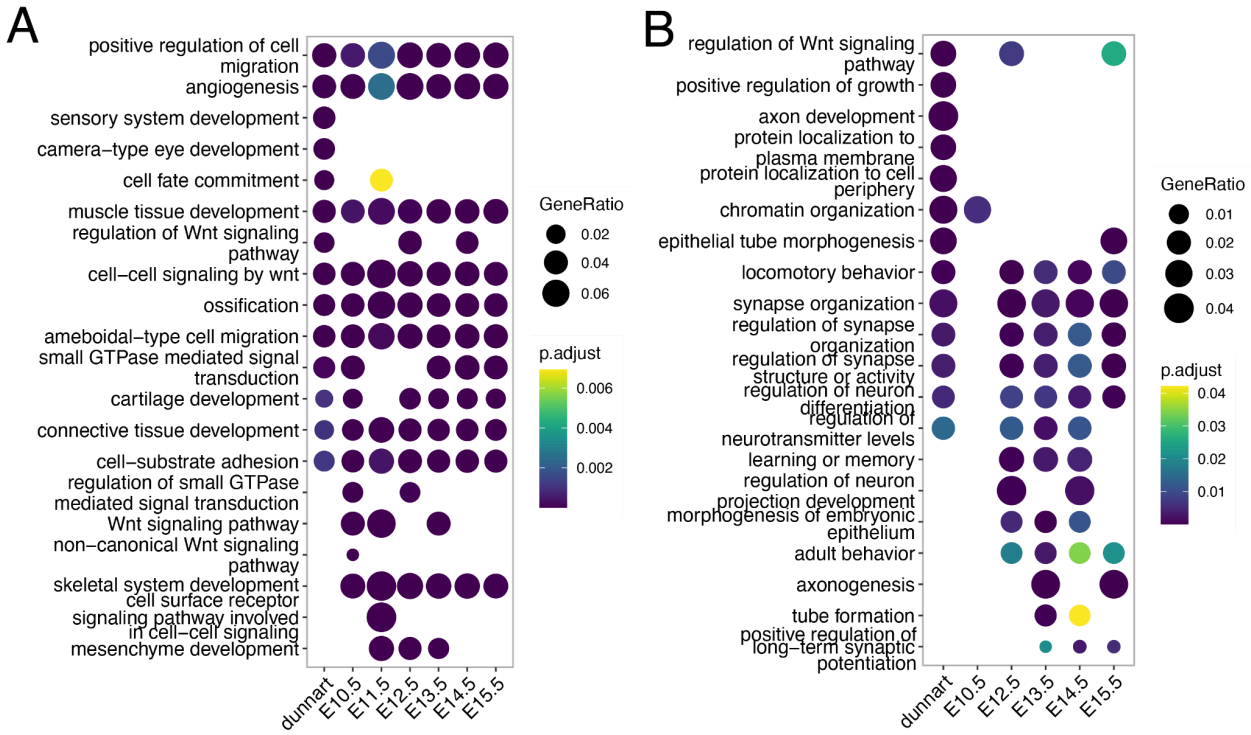
Supplementary Figures



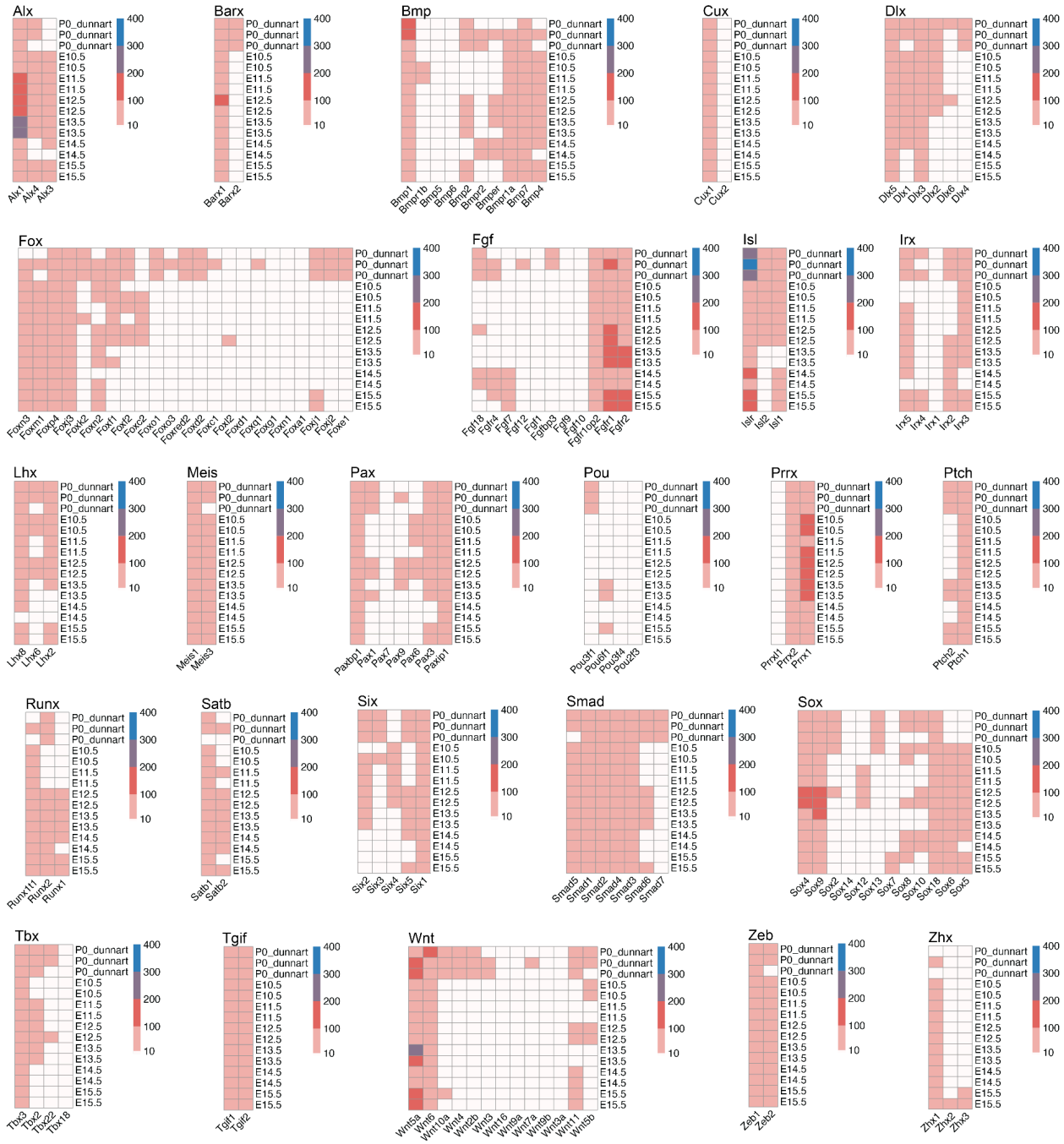
Supplementary Figure 1. Homer motif enrichment for dunnart promoters and enhancers for the top 20 enriched TF families. Log_{10} FDR adjusted p-value scores are shown as a heatmap in enhancers and promoters. Individual TFs are labeled on the right hand side of the heatmap. Distinct colors represent TF families.



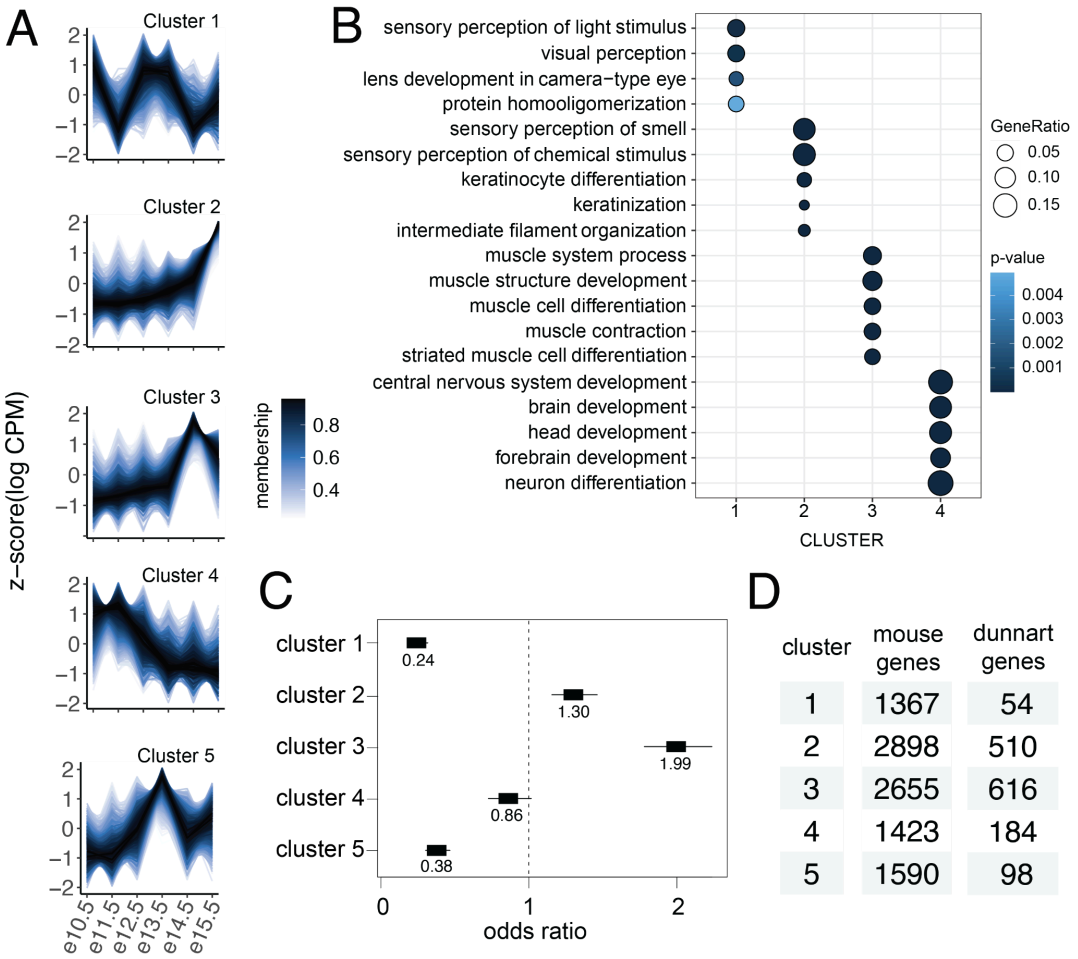
Supplementary Figure 2. Features of predicted promoters and enhancers in the dunnart and mouse. (A) Log_{10} peak intensity (measure of enrichment) for enhancers (orange) and promoters (blue) prior to filtering. (B) Log_{10} distance to the nearest TSS for enhancers (orange) and promoters (blue). After clustering and filtering for high-confidence promoters we observed consistent patterns for features including, (D) CpG and GC content and, (E) Log_{10} peak intensity and Log_{10} peak length.



Supplementary Figure 3. Top GO enriched terms in the dunnart and mouse embryonic stages for (A) genes near enhancers and, (B) genes near high-confidence promoters.

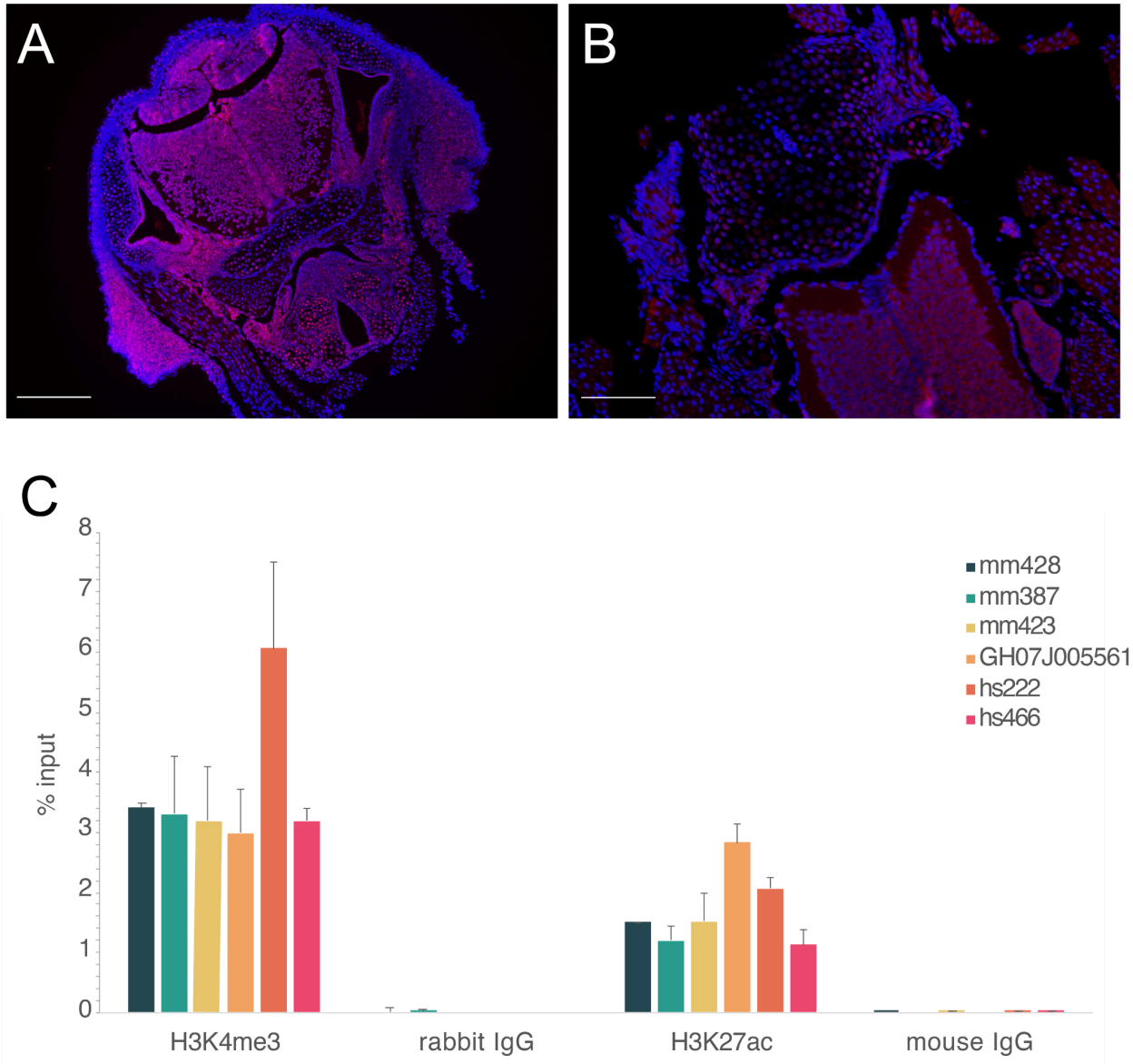


Supplementary Figure 4. Heatmaps showing gene expression values (TPM) for common developmental genes across dunnart and mouse.

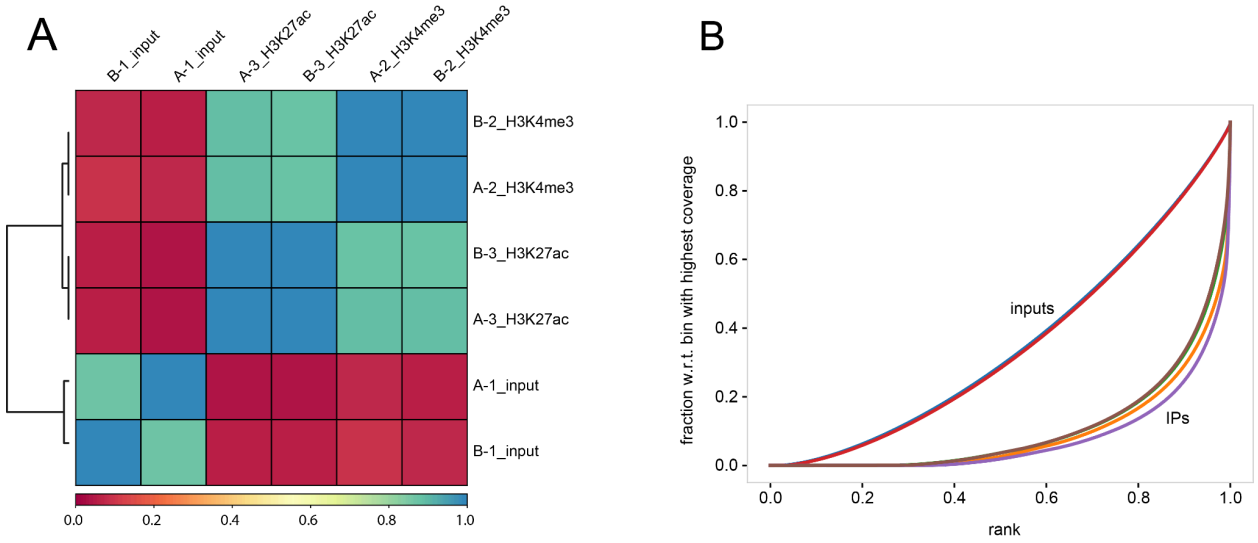


cluster	mouse genes	dunnart genes
1	1367	54
2	2898	510
3	2655	616
4	1423	184
5	1590	98

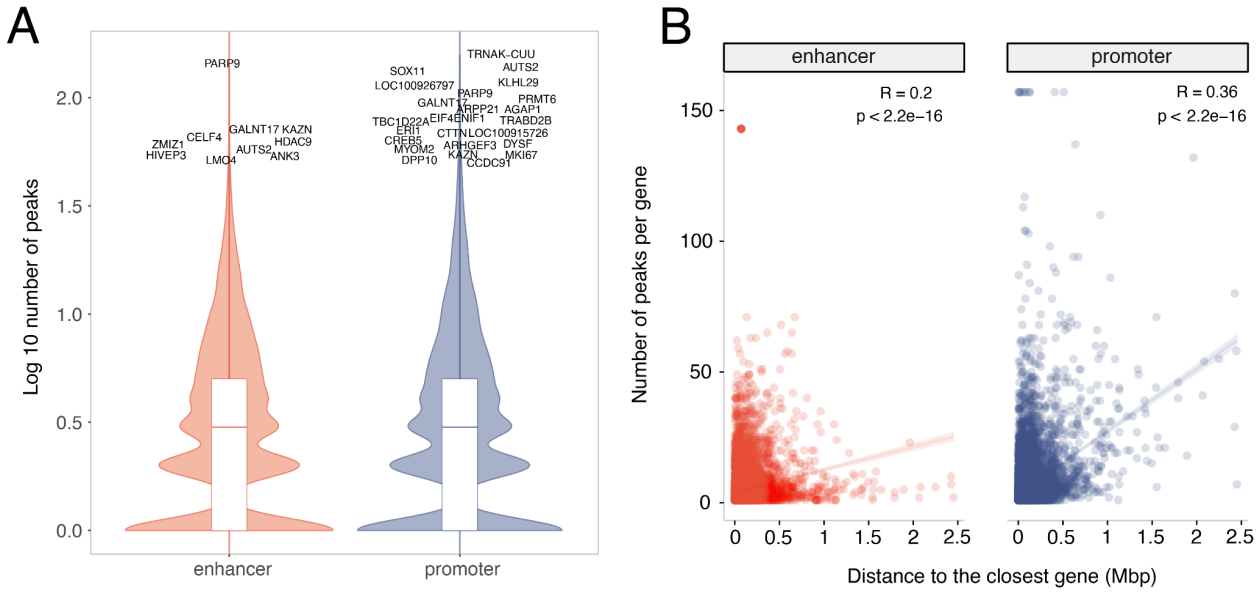
Supplementary Figure 5. Temporal gene expression dynamics. (A) Z-scaled temporal expression (logCPM) plotted across embryonic timepoints for five clusters. Membership values are used to indicate to what degree a data point belongs to each cluster. (B) Top enriched biological processes for each cluster (FDR corrected p<0.01, background gene set is all differentially expressed genes used for clustering). (C) Odds ratio estimates with 95% confidence interval for genes expressed in dunnart in each cluster. (D) Number of genes in each cluster for mouse and dunnart.



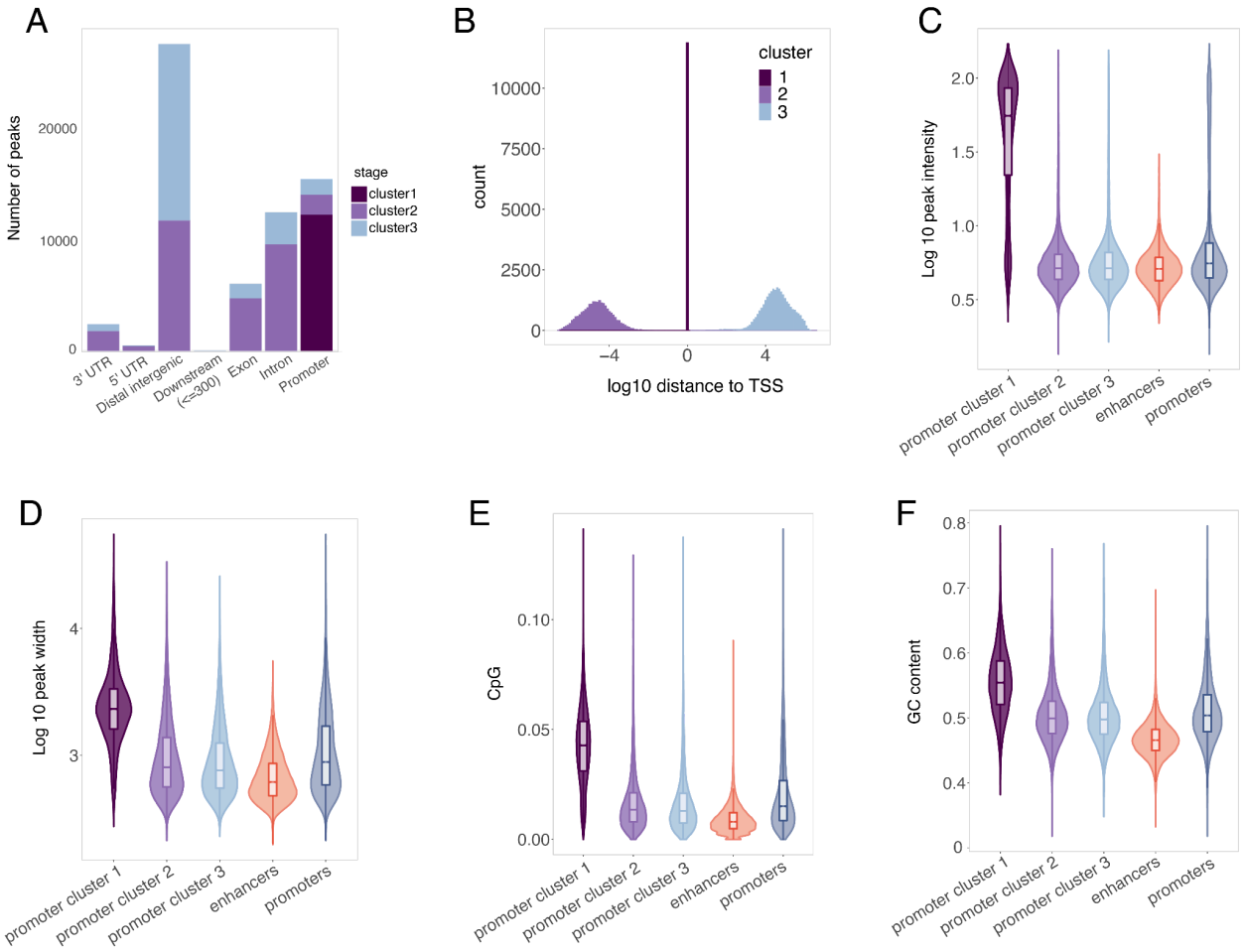
Supplementary Figure 6. Validation of antibodies in dunnart craniofacial tissue with immunofluorescence and qPCR. Localisation of (A) H3K27ac (pink, scale bar = 250um) and (B) H3K4me3 (pink, scale bar = 80um), in D0 dunnart head sections with nuclei stained with DAPI (blue) (C) enhancer regions expected to be enriched in the IP samples presented as the percentage of the input control sample as measured by qPCR.



Supplementary Figure 7. deepTools quality control plots for dunnart subsampled aligned BAM files. (A). Overall similarity between BAM files based on read coverage within genomic regions with Pearson correlation coefficients plotted for H3K27ac, H3K4me3 and input control. (B) Fingerprint plot showing a profile of cumulative read coverages for each BAM file. All reads overlapping a window (bin) of the specified length are counted, sorted and plotted for H3K27ac, H3K4me3 and input control.

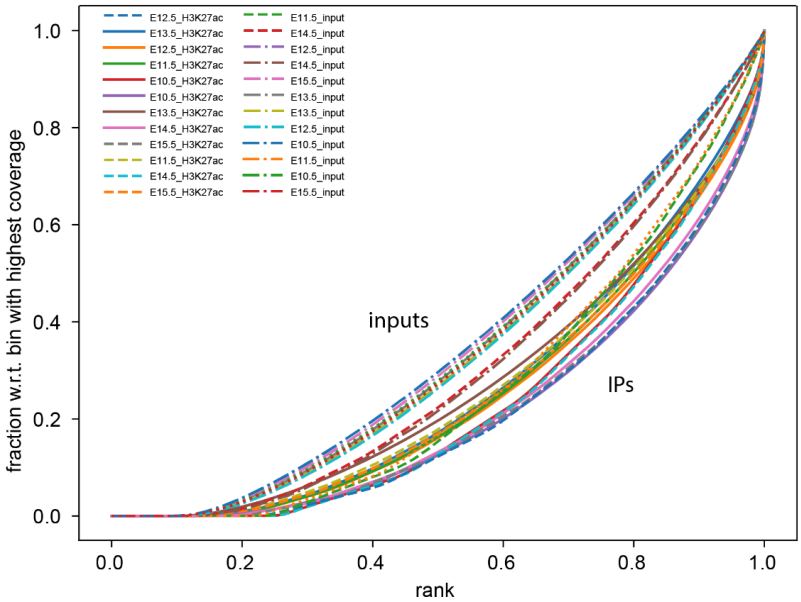


Supplementary Figure 8. Number of peaks per gene for enhancer- and promoter- associated peaks. (A) Log_{10} number of peaks per gene. Genes with greater than 50 peaks are noted. (B) Scatter-plot with distance to the closest gene on the x-axis and number of peaks per gene on the y-axis. There is a weak but significant correlation between the number of peaks per gene and the distance to the next closest gene in both enhancers and promoters.

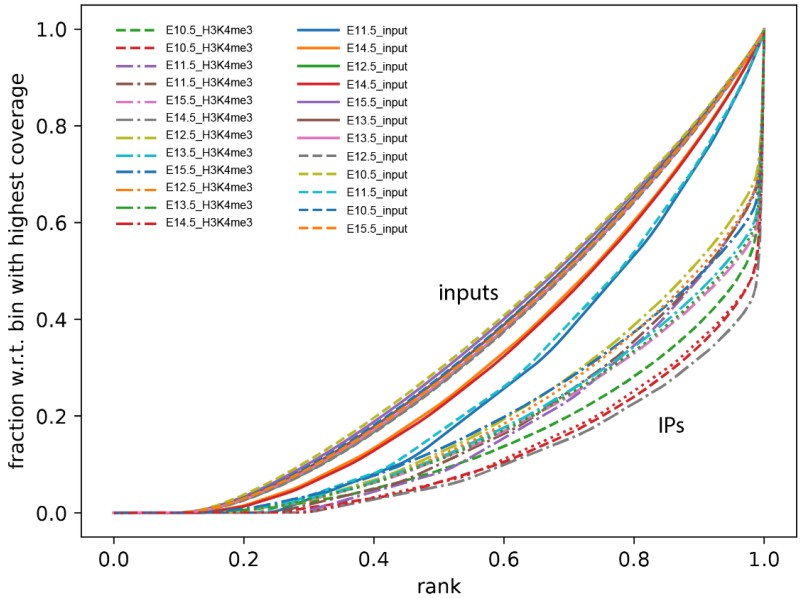


Supplementary Figure 9. Subsetting high-confidence promoter-associated peaks with k-means clustering. (A) Barplot for the number of promoter-associated peaks per cluster and histogram showing the distribution of peak distance from the nearest TSS in each cluster. (B) Genomic annotations for promoter-associated peaks in each cluster. (C) GC content, (D) Log_{10} peak length, (E) CpG content, and (F) Log_{10} of clustered promoter-associated peaks, enhancer-associated peaks, and unclustered promoter-associated peaks. Statistical significance (Wilcoxon, FDR-adjusted, $p <$ 0.00001) compared to cluster 1 promoter-associated peaks is denoted by ****.

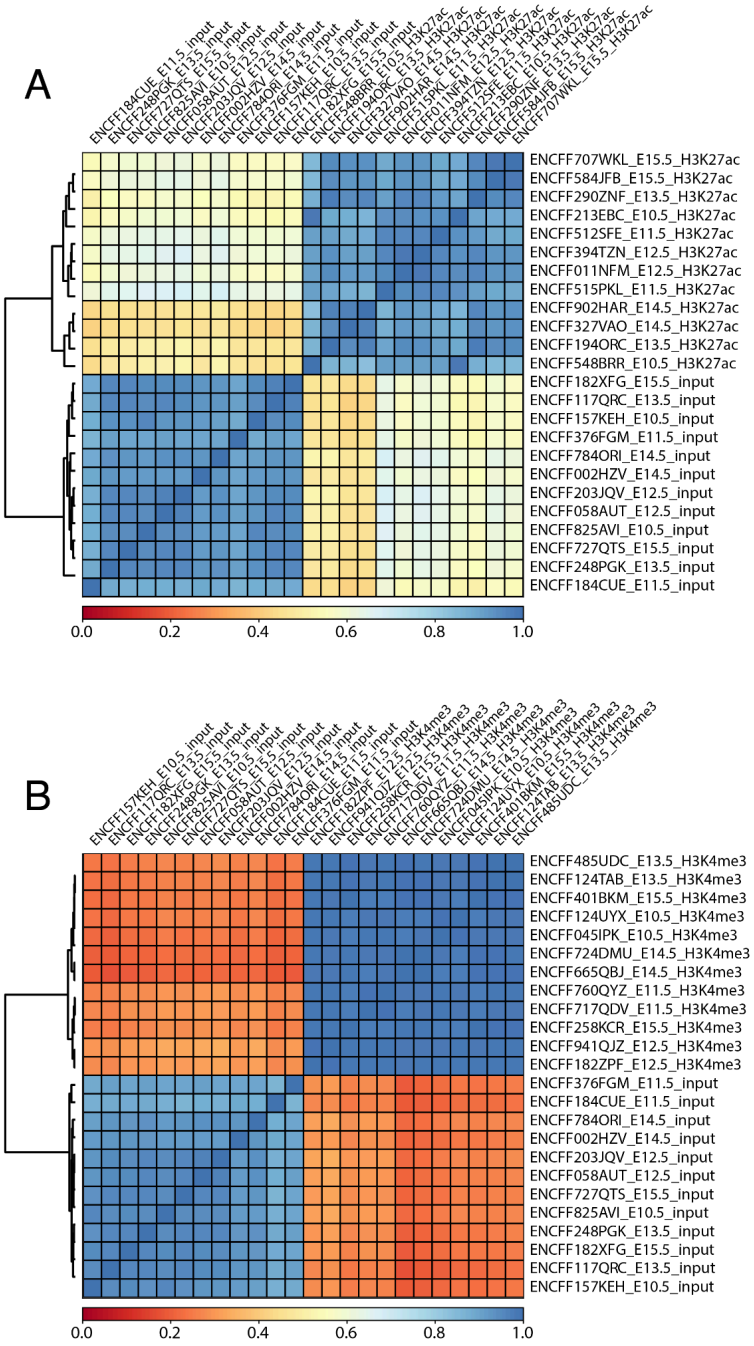
A



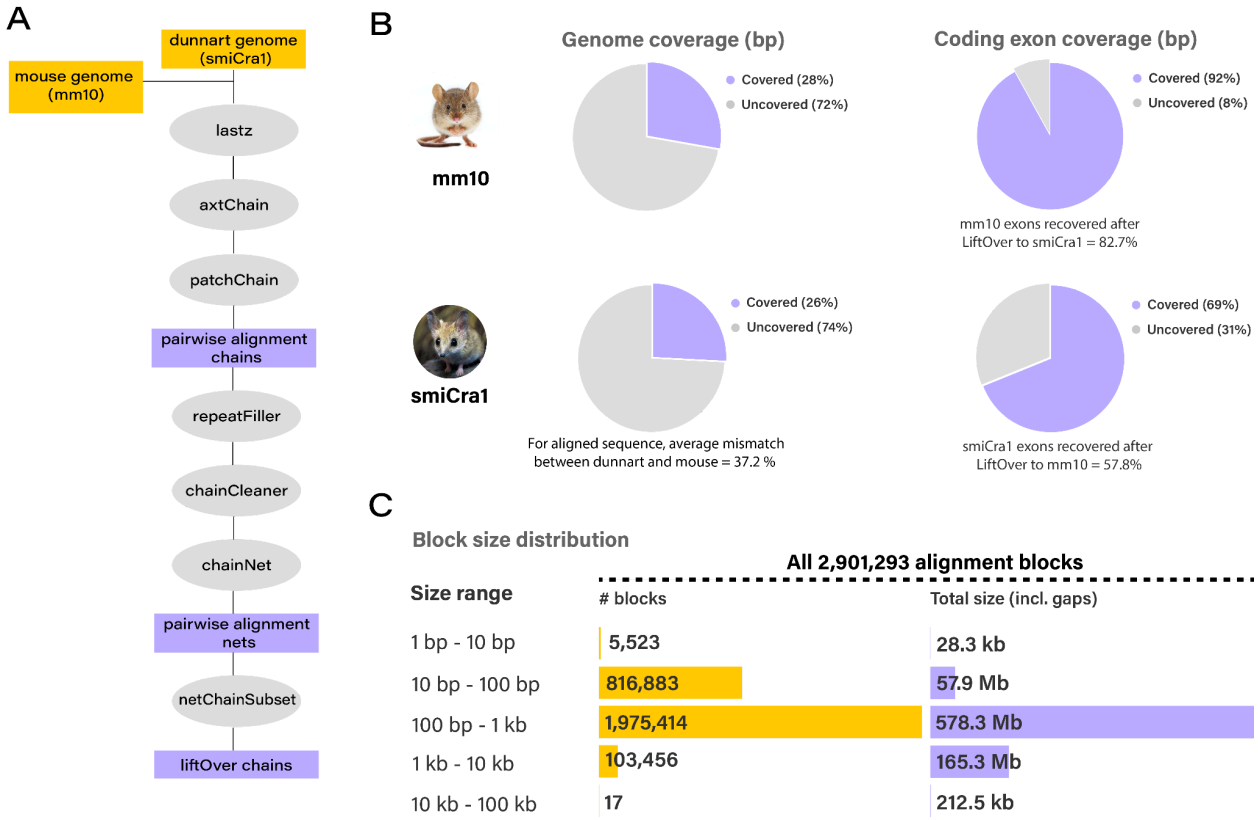
B



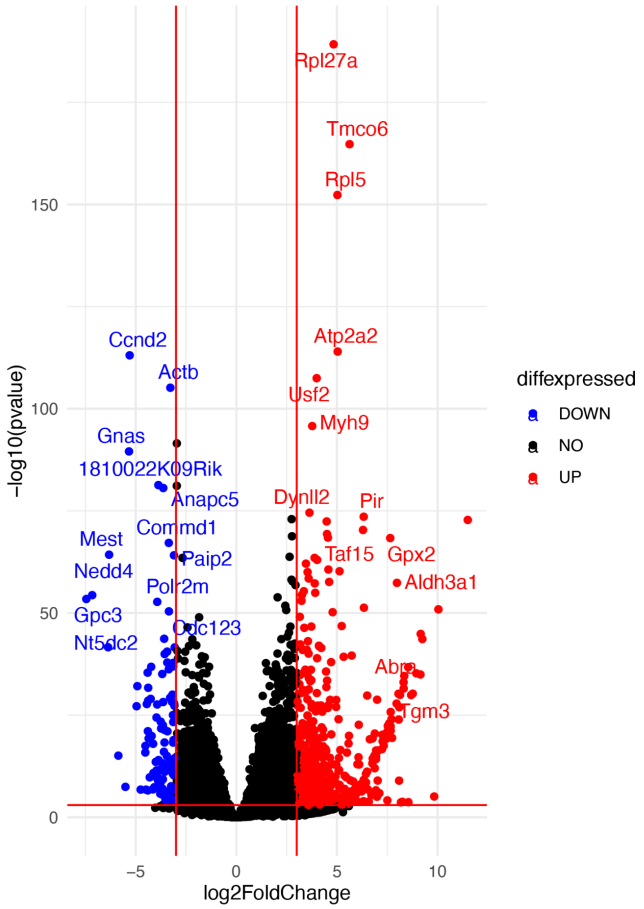
Supplementary Figure 10. deepTools fingerprint plot for mouse subsampled aligned BAM files. Cumulative read coverages for each BAM file. All reads overlapping a window (bin) of the specified length are counted, sorted and plotted for (A) H3K27ac, (B) H3K4me3.



Supplementary Figure 11. deepTools correlation plots for mouse subsampled aligned BAM files. Overall similarity between BAM files based on read coverage within genomic regions with Pearson correlation coefficients plotted for (A) H3K27ac, (B) H3K4me3



Supplementary Figure 12. Genome alignment between the fat-tailed dunnart (*Sminthopsis crassicaudata* and mouse (*Mus musculus*). (A) Genome alignment workflow. (B) Genome coverage and exon coverage (bp) between mouse and dunnart. Coverage of exons recovered after LiftOver. (C) Block size distribution for the dunnart including number of blocks and total size of the blocks.



Supplementary Figure 13. Volcano plot for both upregulated and downregulated differentially expressed genes from mouse (any stage) and dunnart. Stringent log2 fold change and pvalue cutoffs were used given the comparisons between datasets from two different species.

References

1. Francis-West, P. & Crespo-Enriquez, I. *Vertebrate Embryo: Craniofacial Development*. vol. 492 (John Wiley & Sons, Ltd, Chichester, UK, 2016).
2. Nunn, C. L. & Smith, K. K. Statistical analyses of developmental sequences: the craniofacial region in marsupial and placental mammals. *Am. Nat.* **152**, 82–101 (1998).
3. Smith, K. K. Craniofacial development in marsupial mammals: developmental origins of evolutionary change. *Dev. Dyn.* **235**, 1181–93 (2006).
4. Smith, K. K. The evolution of mammalian development. *Bull. Mus. Comp. Zool.* **156**, 119–135 (2001).
5. Cook, L. E., Newton, A. H., Hipsley, C. A. & Pask, A. J. Postnatal development in a marsupial model, the fat-tailed dunnart (Sminthopsis crassicaudata; Dasyuromorphia: Dasyuridae). *Commun Biol* **4**, 1–14 (2021).
6. Smith, K. K. Comparative patterns of craniofacial development in eutherian and metatherian mammals. *Evolution* **51**, 1663–1678 (1997).
7. Veitch, C. E., Nelson, J. & Gemmell, R. T. Birth in the brushtail possum, *Trichosurus vulpecula* (Marsupialia: Phalangeridae). *Aust. J. Zool.* **48**, 691–700 (2000).
8. Jones, T. E. & Munger, B. L. Early differentiation of the afferent nervous system in glabrous snout skin of the opossum (*Monodelphis domesticus*). *Somatosens. Res.* **3**, 169–184 (1985).
9. Waite, P. M., Marotte, L. R. & Leamey, C. A. Timecourse of development of the wallaby trigeminal pathway. I. periphery to brainstem. *J. Comp. Neurol.* **350**, 75–95 (1994).
10. Spiekman, S. N. F. & Werneburg, I. Patterns in the bony skull development of marsupials: high variation in onset of ossification and conserved regions of bone contact. *Science Reports* **7**, 11 (2017).

11. Koyabu, D. *et al.* Mammalian skull heterochrony reveals modular evolution and a link between cranial development and brain size. *Nat. Commun.* **5**, 3625 (2014).
12. Goswami, A. *et al.* Do developmental constraints and high integration limit the evolution of the marsupial oral apparatus? *Integr. Comp. Biol.* **56**, 404–415 (2016).
13. Sears, K. E. Constraints on the morphological evolution of marsupial shoulder girdles. *Evolution* **58**, 2353–70 (2004).
14. Hüppi, E., Sánchez-Villagra, M. R., Tzika, A. C. & Werneburg, I. Ontogeny and phylogeny of the mammalian chondrocranium: the cupula nasi anterior and associated structures of the anterior head region. *Zoological Letters* **4**, (2018).
15. Attanasio, C. *et al.* Fine tuning of craniofacial morphology by distant-acting developmental enhancers. *Science* **342**, 1241006 (2013).
16. Wakamatsu, Y. & Suzuki, K. Sequence alteration in the enhancer contributes to the heterochronic Sox9 expression in marsupial cranial neural crest. *Dev. Biol.* **456**, 31–39 (2019).
17. Wakamatsu, Y., Nomura, T., Osumi, N. & Suzuki, K. Comparative gene expression analyses reveal heterochrony for Sox9 expression in the cranial neural crest during marsupial development. *Evol. Dev.* **16**, 197–206 (2014).
18. Suarez, R. *et al.* Development of body, head and brain features in the australian fat-tailed dunnart (*Sminthopsis crassicaudata*; Marsupialia: Dasyuridae); A postnatal model of forebrain formation. *PLoS One* **12**, e0184450 (2017).
19. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
20. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
21. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).

22. Cain, C. E., Blekhman, R., Marioni, J. C. & Gilad, Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* **187**, 1225–1234 (2011).
23. Santos-Rosa, H. *et al.* Active genes are tri-methylated at k4 of histone h3. *Nature* **419**, 407–411 (2002).
24. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
25. Koch, F. & Andrau, J.-C. Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity. *Transcription* **2**, 263–268 (2011).
26. Pekowska, A. *et al.* H3K4 tri-methylation provides an epigenetic signature of active enhancers: Epigenetic signature of active enhancers. *EMBO J.* **30**, 4198–4210 (2011).
27. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
28. Won, H.-J., Kim, J.-W., Won, H.-S. & Shin, J.-O. Gene regulatory networks and signaling pathways in palatogenesis and cleft palate: A comprehensive review. *Cells* **12**, (2023).
29. Zarate, Y. A., Bosanko, K., Derar, N. & Fish, J. L. Abnormalities in pharyngeal arch-derived structures in SATB2-associated syndrome. *Clin. Genet.* **106**, 209–213 (2024).
30. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
31. Clark, C. T. & Smith, K. K. Cranial osteogenesis in *Monodelphis domestica* (Didelphidae) and *Macropus eugenii* (Macropodidae). *J. Morphol.* **215**, 119–49 (1993).
32. Bentzinger, C. F., Wang, Y. X. & Rudnicki, M. A. Building muscle: molecular regulation of myogenesis. *Cold Spring Harb. Perspect. Biol.* **4**, a008342–a008342 (2012).
33. Buckingham, M. Gene regulatory networks and cell lineages that underlie the formation of skeletal muscle. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5830–5837 (2017).

34. Lin, C.-Y. *et al.* Myogenic regulatory factors Myf5 and Myod function distinctly during craniofacial myogenesis of zebrafish. *Dev. Biol.* **299**, 594–608 (2006).

35. Lefebvre, V., Angelozzi, M. & Haseeb, A. SOX9 in cartilage development and disease. *Curr. Opin. Cell Biol.* **61**, 39–47 (2019).

36. Yip, R. K. H., Chan, D. & Cheah, K. S. E. Mechanistic insights into skeletal development gained from genetic disorders. *Curr. Top. Dev. Biol.* **133**, 343–385 (2019).

37. Qi, H. *et al.* Identification of genes responsible for osteoblast differentiation from human mesodermal progenitor cells. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3305–3310 (2003).

38. Huang, W., Yang, S., Shao, J. & Li, Y.-P. Signaling and transcriptional regulation in osteoblast commitment and differentiation. *Front. Biosci.* **12**, 3068–3092 (2007).

39. Davis, C. A. *et al.* The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).

40. Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

41. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

42. Xiao, Y., Williams, J. S. & Brownell, I. Merkel cells and touch domes: more than mechanosensory functions? *Exp. Dermatol.* **23**, 692–695 (2014).

43. Doucet, Y. S., Woo, S.-H., Ruiz, M. E. & Owens, D. M. The touch dome defines an epidermal niche specialized for mechanosensory signaling. *Cell Rep.* **3**, 1759–1765 (2013).

44. Liu, X. & Liu, S. Cholecystokinin selectively activates short axon cells to enhance inhibition of olfactory bulb output neurons. *J. Physiol.* **596**, 2185–2207 (2018).

45. Simeone, A. *Otx1* and *Otx2* in the development and evolution of the mammalian brain. *EMBO J.* **17**, 6790–6798 (1998).

46. Santoro, S. W. & Jakob, S. Gene expression profiling of the olfactory tissues of sex-separated and sex-combined female and male mice. *Sci Data* **5**, 180260 (2018).

47. Haering, C. *et al.* Ion transporter NKCC1, modulator of neurogenesis in murine olfactory neurons. *J. Biol. Chem.* **290**, 9767–9779 (2015).

48. Robinson, P. *et al.* Mutations in fast skeletal troponin I, troponin T, and beta-tropomyosin that cause distal arthrogryposis all increase contractile function. *FASEB J.* **21**, 896–905 (2007).

49. Wei, B. & Jin, J.-P. TNNT1, TNNT2, and TNNT3: Isoform genes, regulation, and structure-function relationships. *Gene* **582**, 1–13 (2016).

50. Sung, S. S. *et al.* Mutations in TNNT3 cause multiple congenital contractures: a second locus for distal arthrogryposis type 2B. *Am. J. Hum. Genet.* **73**, 212–214 (2003).

51. Jung, S.-Y. & Ko, Y.-G. TRIM72, a novel negative feedback regulator of myogenesis, is transcriptionally activated by the synergism of MyoD (or myogenin) and MEF2. *Biochem. Biophys. Res. Commun.* **396**, 238–245 (2010).

52. Jin, S. *et al.* Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. *Nat. Commun.* **5**, 3793 (2014).

53. Feng, H.-Z. & Jin, J.-P. Carbonic Anhydrase III Is Expressed in Mouse Skeletal Muscles Independent of Fiber Type-Specific Myofilament Protein Isoforms and Plays a Role in Fatigue Resistance. *Front. Physiol.* **7**, 597 (2016).

54. Aboalola, D. & Han, V. K. M. Insulin-like growth factor binding protein-6 alters skeletal muscle differentiation of human mesenchymal stem cells. *Stem Cells Int.* **2017**, 2348485 (2017).

55. Smith, K. K. Early cranial development in marsupial mammals: the origins of heterochrony. *Am. Zool.* **39**, 13A–13A (1999).

56. Smith, K. K. Early development of the neural plate, neural crest and facial region of marsupials. *J. Anat.* **199**, 121–131 (2001).

57. Smith, K. K. Time's arrow: heterochrony and the evolution of development. *Int. J. Dev. Biol.* **47**, 613–621 (2003).

58. Sánchez-Villagra, M. R. & Forasiepi, A. M. On the development of the chondrocranium and the histological anatomy of the head in perinatal stages of marsupial mammals. *Zoological Letters* **3**, 632 (2017).

59. Karlen, S. J. & Krubitzer, L. The functional and anatomical organization of marsupial neocortex: evidence for parallel evolution across mammals. *Prog. Neurobiol.* **82**, 122–141 (2007).

60. Smith, K. K. & Keyte, A. L. Adaptations of the Marsupial Newborn: Birth as an Extreme Environment. *Anat. Rec.* **303**, 235–249 (2020).

61. Brinkley, J. F. *et al.* The FaceBase consortium: a comprehensive resource for craniofacial researchers. *Development* **143**, 2677–2688 (2016).

62. Samuels, B. D. *et al.* FaceBase 3: analytical tools and FAIR resources for craniofacial and dental research. *Development* **147**, (2020).

63. Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83 (2015).

64. Rajderkar, S. S. *et al.* Dynamic enhancer landscapes in human craniofacial development. *Nat. Commun.* **15**, 2030 (2024).

65. Bininda-Emonds, O. R. *et al.* The delayed rise of present-day mammals. *Nature* **446**, 507–12 (2007).

66. Luo, Z.-X., Yuan, C.-X., Meng, Q.-J. & Ji, Q. A jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* **476**, 442–445 (2011).

67. Genereux, D. P. *et al.* A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).

68. Andrews, G. *et al.* Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. *Science* **380**, eabn7930 (2023).

69. Roosenboom, J., Hens, G., Mattern, B. C., Shriver, M. D. & Claes, P. Exploring the Underlying Genetics of Craniofacial Morphology through Various Sources of Knowledge.

Biomed Res. Int. **2016**, 3054578 (2016).

70. Smith, K. K. Development of craniofacial musculature in Monodelphis domestica (Marsupialia, Didelphidae). *J. Morphol.* **222**, 149–173 (1994).
71. Stockdale, F. E. Myogenic cell lineages. *Dev. Biol.* **154**, 284–298 (1992).
72. Hammond, N. L., Dixon, J. & Dixon, M. J. Periderm: Life-cycle and function during orofacial and epidermal development. *Semin. Cell Dev. Biol.* **91**, 75–83 (2019).
73. Liu, H. *et al.* Irf6 directly regulates Klf17 in zebrafish periderm and Klf4 in murine oral epithelium, and dominant-negative KLF4 variants are present in patients with cleft lip and palate. *Hum. Mol. Genet.* **25**, 766–776 (2016).
74. Noiret, M. *et al.* Ptbp1 and Exosc9 knockdowns trigger skin stability defects through different pathways. *Dev. Biol.* **409**, 489–501 (2016).
75. Kousa, Y. A. *et al.* IRF6 and SPRY4 Signaling Interact in Periderm Development. *J. Dent. Res.* **96**, 1306–1313 (2017).
76. Carroll, S. H. *et al.* Neural crest and periderm-specific requirements of Irf6 during neural tube and craniofacial development. *bioRxiv* (2024) doi:10.1101/2024.06.11.598425.
77. Vaziri Sani, F., Kaartinen, V., El Shahawy, M., Linde, A. & Gritli-Linde, A. Developmental changes in cellular and extracellular structural macromolecules in the secondary palate and in the nasal cavity of the mouse. *Eur. J. Oral Sci.* **118**, 221–236 (2010).
78. Lillegraven, J. A. Biological considerations of the marsupial-placental dichotomy. *Evolution* **29**, 707–722 (1975).
79. Pralomkarn, T., Nelson, J. & Gemmell, R. T. Postnatal development of the skin of the marsupial native cat *Dasyurus hallucatus*. *J. Morphol.* **205**, 233–242 (1990).
80. Simpson, S. J. *et al.* Structural and functional development of the respiratory system in a newborn marsupial with cutaneous gas exchange. *Physiological and Biochemical Zoology: Ecological and Evolutionary Approaches* **84**, 634–649 (2011).
81. Ferner, K. Skin structure in newborn marsupials with focus on cutaneous gas exchange. *J.*

Anat. **233**, 311–327 (2018).

82. Krause, W. J., Cutts, J. H. & Leeson, C. R. Postnatal development of the epidermis in a marsupial, Didelphis virginiana. *J. Anat.* **125**, 85–99 (1978).
83. Desmarais, M. J., Beauregard, F., Cabana, T. & Pflieger, J. F. Facial mechanosensory influence on forelimb movement in newborn opossums, *Monodelphis domestica*. *PLoS One* **11**, e0148352 (2016).
84. Gemmell, R. T. & Nelson, J. Ultrastructure of the olfactory system of three newborn marsupial species. *Anat. Rec.* **225**, 203–208 (1989).
85. Jenkins, B. A., Fontecilla, N. M., Lu, C. P., Fuchs, E. & Lumpkin, E. A. The cellular basis of mechanosensory Merkel-cell innervation during development. *Elife* **8**, (2019).
86. NHMRC, National Health And Medical Council. *Australian Code for the Care and Use of Animals for Scientific Purposes*. (2013).
87. Bancroft, J. D. & Gamble, M. *Theory and Practice of Histological Techniques*. (Elsevier Health Sciences, 2008).
88. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–5 (2012).
89. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–92 (2007).
90. Fishilevich, S. *et al.* GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, 1217 (2017).
91. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, 1989).
92. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
93. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).

94. Pryszcz, L. P. & Gabaldón, T. Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113 (2016).

95. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).

96. Feigin, C. Y. *et al.* Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. *Nature Ecology & Evolution* **2**, 182–192 (2018).

97. Feigin, C., Frankenberg, S. & Pask, A. A Chromosome-Scale Hybrid Genome Assembly of the Extinct Tasmanian Tiger (Thylacinus cynocephalus). *Genome Biol. Evol.* **14**, (2022).

98. Renfree, M. B. *et al.* Genome sequence of an australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol.* **12**, R81 (2011).

99. Murchison, E. P. *et al.* Genome sequencing and analysis of the Tasmanian devil and its transmissible cancer. *Cell* **148**, 780–791 (2012).

100. Mikkelsen, T. S. *et al.* Genome of the marsupial monodelphis domestica reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).

101. Peel, E. *et al.* A reference genome for the critically endangered woylie, Bettongia penicillata ogilbyi. *Gigabyte* **2021**, 1–15 (2021).

102. Brandies, P. A. *et al.* The first Antechinus reference genome provides a resource for investigating the genetic basis of semelparity and age-related neuropathologies. *Gigabyte* **2020**, 1–22 (2020).

103. Shumate, A. & Salzberg, S. L. Liftoff: Accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).

104. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proceedings of the National Academy of Sciences* **100**, 11484–11489 (2003).

105. Sharma, V. & Hiller, M. Increased alignment sensitivity improves the usage of genome alignments for comparative gene annotation. *Nucleic Acids Res.* **45**, 8369–8377 (2017).

106. Suarez, H. G., Langer, B. E., Ladde, P. & Hiller, M. chainCleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33**, 1596–1603 (2017).

107. Dutheil, J. Y., Gaillard, S. & Stukenbrock, E. H. MafFilter: a highly flexible and extensible multiple genome alignment files processor. *BMC Genomics* **15**, 53 (2014).

108. Andrews, S. *FastQC: A Quality Control Tool for High Throughput Sequence Data [online]*. (2010).

109. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

110. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

111. He, Y. *et al.* Spatiotemporal DNA methylome dynamics of the developing mouse fetus. *Nature* **583**, 752–759 (2020).

112. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

113. Crusoe, M. R. *et al.* The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Res.* **4**, 900 (2015).

114. Zhang, Y. *et al.* Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).

115. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

116. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

117. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov. J.* **2**, 100141 (2021).

118. Marc Carlson. *Org.mm.eg.db: Genome Wide Annotation for Mouse. R Package Version*

3.12.0. (2020).

119. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

120. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).

121. Zhang, Y., Park, C., Bennett, C., Thornton, M. & Kim, D. Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N. *Genome Res.* **31**, 1290–1295 (2021).

122. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).

123. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

124. Wu, M. & Gu, L. *TCseq: Time Course Sequencing Data Analysis*. (2020).

125. Jari Oksanen *et al. Vegan: Community Ecology Package*. (2020).

126. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).

127. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).

128. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).

129. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).