

RESEARCH ARTICLE

Divergent evolution of developmental timing in the neocortex revealed by marsupial and eutherian transcriptomes

Peter Kozulin^{1,*,\$}, Rodrigo Suárez^{1,*,\$,¶}, Qiong-Yi Zhao^{1,\$}, Annalisa Paolino¹, Linda J. Richards^{1,2,‡,¶} and Laura R. Fenlon^{1,*,\$,¶}

ABSTRACT

Only mammals evolved a neocortex, which integrates sensory-motor and cognitive functions. Significant diversifications in the cellular composition and connectivity of the neocortex occurred between the two main therian groups: marsupials and eutherians. However, the developmental mechanisms underlying these diversifications are largely unknown. Here, we compared the neocortical transcriptomes of *Sminthopsis crassicaudata*, a mouse-sized marsupial, with those of eutherian mice at two developmentally equivalent time points corresponding to deeper and upper layer neuron generation. Enrichment analyses revealed more mature gene networks in marsupials at the early stage, which reverted at the later stage, suggesting a more precocious but protracted neuronal maturation program relative to birth timing of cortical layers. We ranked genes expressed in different species and identified important differences in gene expression rankings between species. For example, genes known to be enriched in upper-layer cortical projection neuron subtypes, such as *Cux1*, *Lhx2* and *Satb2*, likely relate to corpus callosum emergence in eutherians. These results show molecular heterochronies of neocortical development in Theria, and highlight changes in gene expression and cell type composition that may underlie neocortical evolution and diversification.

This article has an associated ‘The people behind the papers’ interview.

KEY WORDS: Cortical development, Marsupial brain, Heterochrony, Developmental timing, Cortical transcriptome, Corpus callosum, *Sminthopsis crassicaudata*

INTRODUCTION

The neocortex is unique to mammals and crucially underpins the behavioural diversity and complexity exhibited by this class. The lack of brain structures with clear structural and/or functional homology to the neocortex in other vertebrate species has posed

barriers to the identification of mechanisms that contributed to its evolution (Faunes et al., 2015; Briscoe and Ragsdale, 2018). However, within mammals there exists egg-laying monotremes, marsupials and eutherians (placentals), all of which show subtle group-specific divergences in an otherwise highly conserved brain anatomy. For example, the neocortices of eutherians have more neurons than similar sized marsupials (Cheung et al., 2010; Seelke et al., 2013). Similarly, whereas the connections between the two neocortices cross via the anterior commissure in monotremes and marsupials, eutherians evolved a new commissural route: the corpus callosum (Suárez et al., 2014, 2018). The mechanisms involved in such evolutionary innovations remain largely unknown, but likely include changes in developmental processes such as molecular cell-fate specification and tissue remodelling (Gobius et al., 2016, 2017; Paolino et al., 2020).

Marsupialia are a highly diverse class of mammals endemic to Australasia and the Americas. Their evolutionary adaptations to their environment encompass vastly specialised means of locomotion, reproduction and life-cycles. Their highly distinctive development of young within the pouch have also led to specialised adaptations for early development. Despite the importance of marsupials for understanding the evolutionary history of Theria, genomic sequencing of marsupials has been limited to only four species: the South American opossum (*Monodelphis domestica*), the Australian tammar wallaby (*Macropus eugenii*), the koala (*Phascolarctos cinereus*) and the Tasmanian devil (*Sarcophilus harrisii*) (Mikkelsen et al., 2007; Miller et al., 2011; Renfree et al., 2011; Johnson et al., 2018). Transcriptomic analysis of marsupials outside these four species has, therefore, required *de novo* transcriptome assembly, of which there are few datasets collected and not all are derived from primary tissue samples, such as the rat kangaroo (*Potorous tridactylus*) kidney epithelial cell line (Udy et al., 2015). The adult marsupial transcriptome also limits insights into neocortical evolution and development, given that it may not reflect divergences in transient developmental gene expression and potential heterochronies that might underlie adult structural differences. Here, we compare the transcriptomic landscape of the developing neocortex between a marsupial and a eutherian species of similar size and ecological adaptations. Specifically, we generated a *de novo* assembled transcriptome of the marsupial fat-tailed dunnart (*Sminthopsis crassicaudata*) neocortex, at two crucial stages of its development, during the birth of infragranular (i.e. deeper layers 5–6) and supragranular (i.e. upper layers 2–3) neurons, and performed comparative sequence, ontology, enrichment and localisation analyses with mouse neocortical transcriptomes at equivalent stages of development. The molecular database in this study also provides a rich resource from which to undertake further investigations into the development and evolution of the mammalian brain.

¹Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia. ²School of Biomedical Sciences, Faculty of Medicine, The University of Queensland, Brisbane, QLD 4072, Australia.

*Present address: School of Biomedical Sciences, The University of Queensland, Brisbane, QLD 4072, Australia. †Present address: Washington University School of Medicine, Department of Neuroscience, 660 S. Euclid Ave, St Louis, MO 63110, USA.

§These authors contributed equally to this work

¶Authors for correspondence (l.fenlon@uq.edu.au; richards@uq.edu.au; r.suarez@uq.edu.au)

ORCID R.S., 0000-0001-5153-5652; L.J.R., 0000-0002-7590-7390; L.R.F., 0000-0002-5132-2419

RESULTS

Dunnart neocortical *de novo* transcriptome shows a mapping hierarchy consistent with its mammalian phylogeny

In order to compare the developmental cortical transcriptomes between species, we first determined the most developmentally relevant ages of each species for collection. We previously published a staging system of the fat-tailed dunnart that is consistent with human and mouse embryonic development (i.e. Carnegie and Theiler systems, respectively), based on highly conserved anchor points of development, such as eye and limb formation (Suárez et al., 2017). We have previously shown with birth-dating data that cortical neurogenesis is initiated and terminated at equivalent stages (S) between species (S19-S25) and, although each stage within this period spans 1 day in mouse, it spans ~3 days in dunnart (Paolino et al., 2020). Here, we chose ages corresponding to the generation of deeper and upper layer neocortical neurons, identified by injecting the thymidine analogue EdU in developing mice and dunnarts (Fig. 1A-D). Dunnarts have a six-layered cerebral cortex which develops in an inside-out fashion, with deeper and upper layer neocortical neurons generated during early and late corticogenesis, respectively, similar to mice (Suárez et al., 2017; Paolino et al., 2018). Their development is, however, protracted, and therefore comparisons between species are best performed using staging rather than days post fertilisation. During early cortical development [S20-21; embryonic day (E) 12 in mice and postnatal day (P) 12 in dunnarts, Fig. 1A,B] EdU-labelled cells were present in layers 5/6. During late cortical development (S23-24; E16 mouse or P20 dunnart, Fig. 1C,D) labelled cells were present in layers 2/3. We then used these ages as comparable benchmark periods of development for all subsequent collections and transcriptomic analyses, due to the equivalency of overall development according to our staging system, as well as the match in timing of birth for equivalent cortical layers (see Materials and Methods). We refer to these stages as early cortical development (ECD) and late cortical development (LCD) hereafter.

The dunnart RNA-seq samples had a high depth of sequencing (average 153 million reads per sample) and high level of accuracy (Table S2), resulting in a total of 2.7 million Trinity-assembled

transcripts, 183,321 of which had predicted open reading frames (Table S3). Although the dunnart genome sequence has not yet been released, we were able to validate our transcriptome dataset with those previously published in other species by aligning the dunnart transcriptome against a subset of marsupial and eutherian genomes. This showed a mapping hierarchy consistent with the phylogenetic relationships between these species. The closest alignment was with another member of the Dasyuridae family (Tasmanian devil, 94.6%), followed by other marsupials (Australidelphian Tamar wallaby 50.0% and Ameridelphian opossum 39.9%), and finally eutherian species were the least similar (Fig. 1E). We then quantified the completeness of the dunnart transcriptomic assembly and annotation using a Benchmarking Universal Single-Copy Orthologs (BUSCO) approach (Simão et al., 2015). This analysis revealed that 65.9% of the *de novo*-assembled transcripts showed sufficiently high length alignment scores to be regarded as a complete BUSCO match aligned to the 'mammalia_odb9' database of orthologous mammalian genes groups (Fig. 1F). Moreover, the majority of dunnart transcripts were full-length and had orthologous sequences present in all 50 species listed in the Mammalia database. About a quarter (25.1%) of dunnart transcripts showed length alignment scores less than that expected for 'mammalia_odb9' orthologues (fragmented BUSCO match), and 9.0% of dunnart transcripts showed no significant matches to this mammalian reference (missing BUSCO match). We further validated the accuracy of our sequences by cloning one of the few fat-tailed dunnart genes that has previously been sequenced, *Crh* (GenBank sequence KT380843.1), which showed complete alignment apart from one synonymous nucleotide variant (Fig. S1).

Given the high (91%) conservation of orthologous transcripts (complete or fragmented) between dunnart and the BUSCO mammalian reference, we focused our analysis primarily on comparing protein-coding orthologous genes between mice and dunnarts across developmental stages. We identified 12,241 orthologous gene pairs between species, consisting of 12,632 orthologous transcripts (Table S4; Fig. 1G). Transcript abundance (reported as fragments per kilobase of transcript per million mapped reads, FPKM) was calculated for all orthologous gene pairs in each sample in both species. Based on the FPKM value, rankings of the expression level (from 1 to 12,632, with 1 being the lowest and

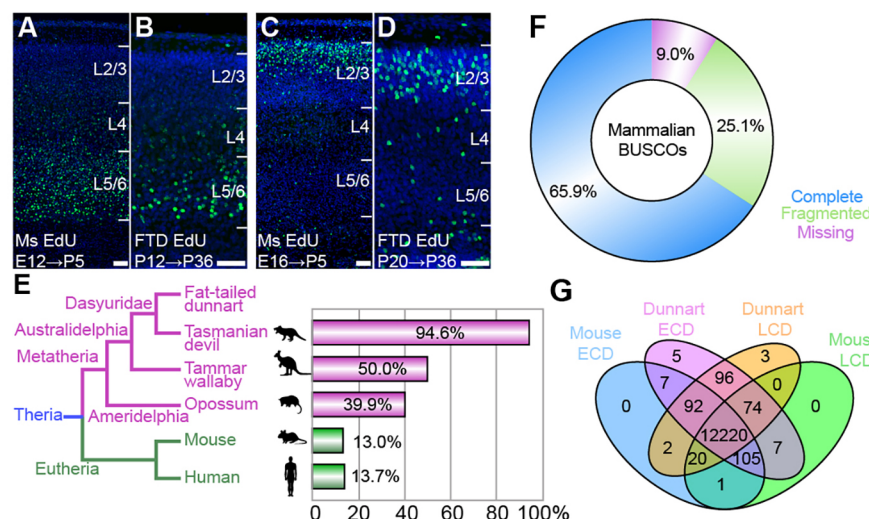


Fig. 1. Generation of an age-matched, phylogenetically relevant and high-quality dunnart developmental transcriptome. (A-D) EdU labelling at developmental stages and collection after the maturation of neocortical layers revealed a similar pattern of labelling between mouse and dunnart at E12 (A) and P12 (B), respectively, labelling deeper layer neurons (ECD), and at E16 (C) and P20 (D), respectively, labelling upper layer neurons (LCD). (E) Mapping the *de novo* dunnart transcriptome to five reference mammalian genomes (see left dendrogram for known phylogenetic relationship) revealed the highest percentage similarity to another dasyurid, the Tasmanian devil (94.6%), followed by more distantly related marsupials, and finally eutherian species. (F) BUSCO analysis revealed that the majority of dunnart transcripts showed 'complete' alignment to the 'mammalia_odb9' lineage database of orthologous mammalian gene groups. (G) The majority (12,220 out of 12,632) of orthologous gene pairs identified between dunnart and mouse had a non-zero expression in both species and at both developmental stages. Scale bars: 50 µm.

12,632 being the highest expressed) were assigned to each transcript, and for each species and developmental stage, thus allowing for direct comparison between species. We also mapped non-coding RNAs (ncRNA) by searching the top hits (lowest E-value) for each Blastn alignment between the dunnart transcriptome and the mouse GRCm38 ncRNA database (ftp://ftp.ensembl.org/pub/release-94/fasta/mus_musculus/ncrna). This produced 2550 unique Trinity transcripts that corresponded to 707 mouse ncRNAs (Table S5), which represents 33% of the whole mouse database. The smaller ncRNA dataset in dunnart is likely a result of the fragmented nature of the *de novo* assembled dunnart transcriptome and the stringent mapping criteria.

Divergent rates of change across early and late stages of neocortical development between dunnart and mouse transcriptomes

We first compared differentially expressed protein-coding genes between ECD and LCD within each species. The number of transcripts that showed a significant differential expression within species, across developmental periods (FDR-adjusted

P -value<0.01, exact test pairwise analysis) and a fold change of at least four (LCD relative to ECD) was 2443 in mouse (Fig. 2A, red and green circles) and 2178 in dunnart (Fig. 2B, red and green circles). We analysed which of these transcripts were mouse-dunnart orthologues (Table S4; Fig. 2A and B, green circles) and whether they had increased (833 orthologous transcripts) or decreased (581 orthologous transcripts) expression at LCD relative to ECD in mouse (Tables S6A and B). In contrast, only 38 orthologous transcripts had increased expression at LCD relative to ECD in dunnart (Table S6C) but 617 had decreased expression at LCD relative to ECD (Table S6D). Overall, there was a trend in mouse towards a profile of a larger number of differentially expressed orthologous transcripts with absolute fold-change values greater than four (Fig. 2C). The disparity in this fold-change profile across species, in particular the comparatively small number of transcripts upregulated in dunnarts during LCD, suggests there may be different transcriptomic dynamics between species that have resulted in an altered rate of change of the neocortical profile of mouse between these two stages compared with dunnart.

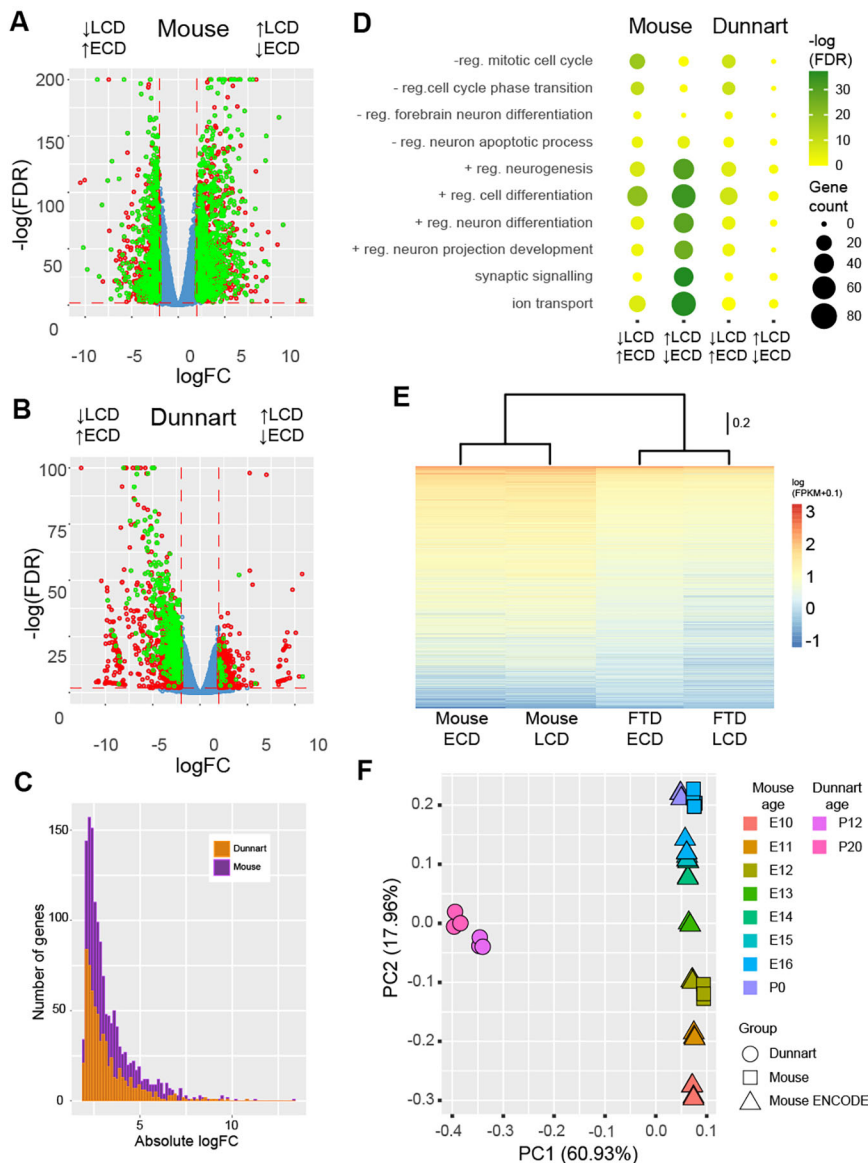


Fig. 2. Intrasppecies inter-age comparisons reveal distinct dynamics of development. (A,B) Volcano plots for mouse (A) and dunnart (B) show expression of transcripts above significance thresholds of a log2 fold change (logFC) of 2 and a P -value [adjusted for false discovery rate (FDR) to correct for multiple testing of differential gene expression] of 0.01 (red circles), and a subset of these transcripts that were also found to be mouse-dunnart orthologues (green circles). Fold changes are relative to ECD, and $-\log_{10}$ -transformed FDR P -values are graphically limited to 1e-200 in mouse and 1e-100 in dunnart. (C) Histogram of number of genes associated with absolute log2 fold change (0.15 bin size) between ECD and LCD in dunnart (orange) and mouse (magenta). (D) GO enrichment analysis shows greater enrichment of ontologies related to neuronal differentiation and function for genes upregulated at LCD in mouse relative to ECD, compared with the same developmental analysis in dunnart. ID numbers of the ontology terms are shown in Table S7A. P -values were adjusted for FDR and $-\log_{10}$ transformed. (E) A dendrogram and heatmap of log-transformed mean FPKM values of the common 7883 mouse-dunnart orthologue transcripts present in all experimental groups. The mean FPKM values of each group were transformed via addition of a 0.1 positive constant and log (base 10) applied. Dendrogram scale bar indicates 0.2 correlation dissimilarity. (F) Principal component analysis of dunnart and mouse samples, in addition to an ENCODE mouse forebrain dataset, the developmental series of which is visualised along the second principal axis (PC2). Each sample consists of count profiles normalised by library size for 8674 transcripts common to all groups. The results show a greater separation of mouse cortical profiles along PC2 between ECD (E12, squares) and LCD (E16, squares) than dunnart between these stages (P12 to P20, circles). FTD, fat-tailed dunnart; $-\log$ (FDR), negative log of FDR-adjusted P -values; +reg., positive regulation; -reg., negative regulation.

To better understand the functional significance of these differentially expressed genes, we performed gene ontology (GO) analyses of orthologue transcripts that were upregulated during LCD, relative to ECD (i.e. green circles in Fig. 2A,B) for each species. This showed a larger enrichment, from ECD to LCD, of gene ontologies related to neuronal differentiation and function in mouse compared with dunnart (Fig. 2D; Table S7A). Genes related

to neurogenesis, neuron differentiation, synaptic signalling and ion transport were more highly upregulated in mouse between these two stages. This indicates that during development from early to late stages, the mouse cortex experiences a greater change in mitotic and neuronal differentiation processes, and the onset of mature neuronal functions, than a comparison of the two stages in dunnart. In contrast, genes that were upregulated during ECD (relatively downregulated during LCD) did not show a notable change in GO enrichment between developmental periods in either species with respect to the aforementioned developmental processes. This trend may reflect a relative stability of early progenitor populations (and thereby their gene expression) across these two stages of cortical development, with the majority of interstage differences arising in mouse driven by the transcriptomes of later-born populations (such as mature projection neurons) in LCD. To further assess the broad relationships between stages and species, we created a dendrogram and heatmap of the mean transformed expression values of the orthologue transcripts identified in all samples across experimental groups (7883 transcripts; Fig. 2E). These results show a greater similarity between developmental periods within each species than any interspecies relationship. Furthermore, principal component analysis was performed for the dunnart and mouse samples, together with a developmental series of ENCODE project samples of mouse forebrain, the developmental stage variance of which was visualised along the second principal component (Fig. 2F). The results show a clear separation of species along the first principal component, and a greater separation of mouse cortical profiles between ECD and LCD along the second principal component compared with the dunnart. Taken together, Fig. 2E and F indicate that the greatest variance in these samples is determined by species, and that a smaller transcriptional change occurs across ECD and LCD in dunnart compared with mouse. These results further validate our developmental datasets and therefore allow us to directly compare similar stages between species.

Dunnart neocortical transcriptome has enriched markers of neural maturation compared with mouse during ECD

We used a reciprocal match approach to highlight the most salient interspecies differences at each ECD and LCD stage. Transcripts common to both the top 30% (i.e. expression rank 9475 to 12,632) of one species and the bottom 30% (i.e. expression rank 1 to 3158) of the other species were first identified for each stage (Fig. 3A; Table S8). We then performed GO enrichment analyses on these gene lists to get a broad indication of their likely functions (Fig. 3B)

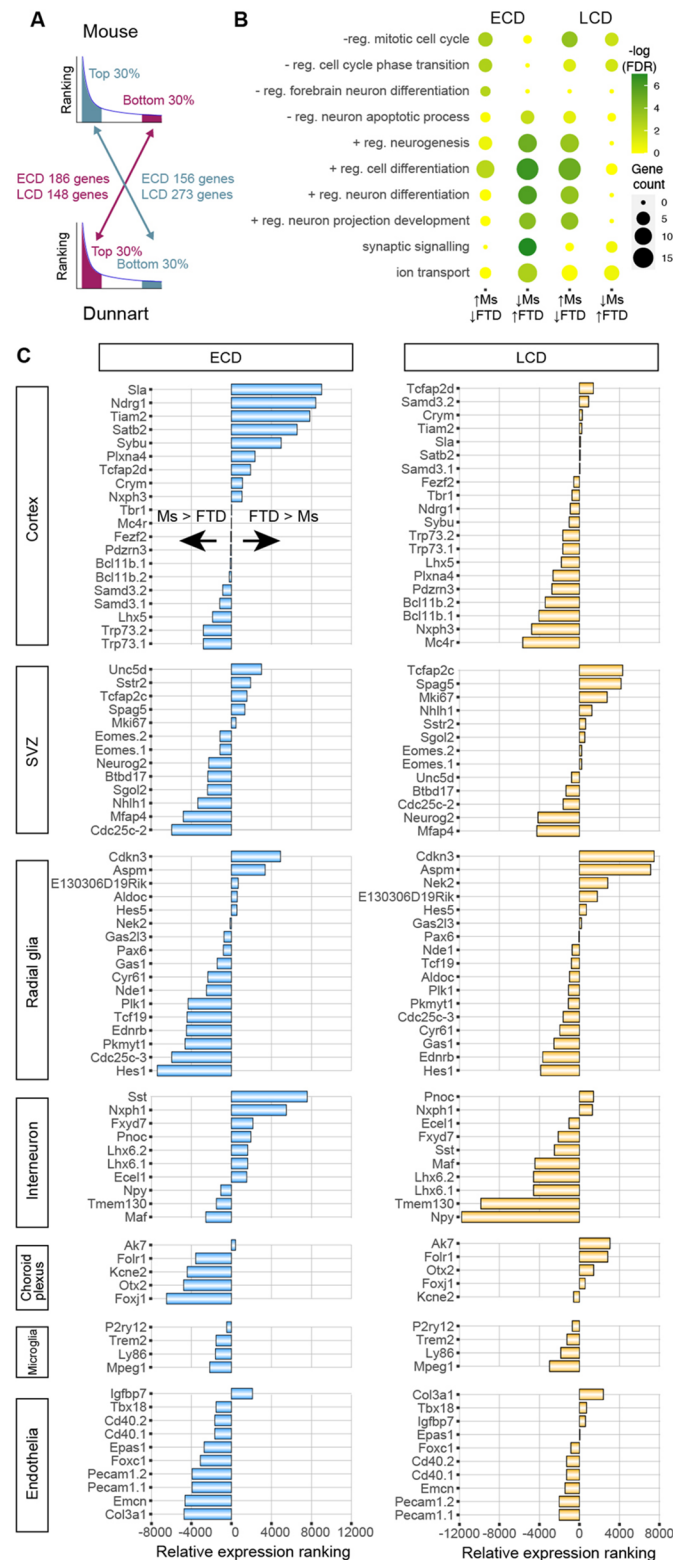


Fig. 3. Interspecies comparisons show different distributions of genes enriched in developmental processes and neocortical cell types. (A) A reciprocal match approach between the top 30% expressed transcripts in one species and bottom 30% of the other species was used to identify transcripts most differentially expressed between dunnart (FTD) and mouse (Ms) during ECD and LCD. (B) Ontology enrichment analysis of these differentially ranked transcript lists showed that during ECD there was a greater enrichment of ontologies related to neuronal differentiation and function for transcripts with a higher ranked expression in dunnart relative to mouse; this trend reversed during LCD. ID numbers of the ontology terms are shown in Table S7B. *P*-values were adjusted for FDR to correct for multiple testing and $-\log_{10}$ transformed. (C) Difference in orthologue gene expression rankings between dunnart and mouse for genes found to be enriched in seven different clusters identified by Loo et al. (2019). In E14 mouse cortex single cell analyses are presented as dunnart relative to mouse, and reveal trends towards different degrees of enrichment for genes associated with distinct neocortical cell types. +reg., positive regulation; -reg., negative regulation; SVZ, subventricular zone.

(Klopfenstein et al., 2018). Interestingly, the transcripts with a higher expression ranking in dunnarts relative to mice during ECD (Table S8A) were enriched for GO terms related to advanced stages of neural maturation, such as neuron differentiation, neuronal projection development, ion transport and synaptic signalling. In contrast, transcripts with a higher expression ranking in mice relative to dunnarts during ECD (Table S8B) were enriched for GO terms that negatively regulate processes related to mitosis and cell cycle (Fig. 3B; Table S7B). It has previously been reported that differing lengths of various phases of the cell cycle are related to differing developmental outcomes in the cerebral cortex. For example, a shorter S-phase is linked to differentiation (Arai et al., 2011). It is therefore possible that the relative enrichment of GO terms for negative regulation of mitotic cell cycle and cell cycle phase transition in mouse is functionally linked with the downregulation of differentiation via a mechanism such as lengthening S-phase and thereby inhibiting differentiation. In contrast to ECD, during LCD few ontology enrichments were observed for transcripts with a higher expression ranking in dunnarts relative to mice (Table S8C), but included negative regulation of cell cycle. Notably, a larger number of ontologies were enriched for transcripts with a higher expression ranking in mice relative to dunnarts at the later stage (Table S8D), and included similar post-mitotic neuronal ontologies to those observed for transcripts with a higher expression ranking in dunnart relative to mice during the early stage (Fig. 3B; Table S7B). The direction of regulation of these processes, together with the greater degree of interstage differences in mouse than in dunnart (Fig. 2A-D), suggests that the dunnart cortex undergoes precocious neural maturation compared with mice during ECD, when the same cortical layers are being born. This result supports the principal component analysis of these samples (Fig. 2F) in that the PC2 value of the dunnart ECD sample relative to mouse ECD suggests a precocious maturity of the dunnart neocortex at this stage, and a larger change in PC2 value between ECD and LCD in mouse than in dunnart. Given the principal component and ontology analyses were performed on genes common to all samples, this trend indicates that the change in expression profile between ECD, which shows higher expression of progenitor transcripts, and LCD, which shows higher expression of mature neuronal transcripts, is greater in mouse.

Next, we sought to further investigate whether changes in cell-type composition may underlie the aforementioned interspecies differences in GO enrichments. To achieve this, we examined the rankings of our orthologue transcript list relative to published datasets of genes enriched in subpopulations of developing cortex. We specifically used datasets from single cell RNA-seq because of its ability to identify cell types based on expression profile clustering and associated gene enrichment. We first compared against a dataset that used single cell RNA-seq to categorise genes enriched in defined cell subpopulations in the E14 mouse cortex (Loo et al., 2019), a mid-point between our ECD and LCD stages. Consistently, genes previously classified in mature cell clusters such as 'cortex' and 'interneuron' showed a higher expression ranking in dunnarts than mice during ECD (positive rank change value, Fig. 3C), whereas during LCD, the majority of these genes showed higher expression in mice (negative rank change value). In contrast, genes classified in clusters related to earlier developmental stages, such as 'subventricular zone', 'radial glia', 'choroid plexus' and 'endothelia' showed higher expression in mice compared with dunnarts during ECD, and a more balanced expression pattern between species during LCD. These findings further demonstrate that the precocious maturity of fat-tailed dunnart cortex relative

to mouse during ECD, when the equivalent cortical layers are being born in both species, indicated by the GO analysis is also reflected in the earlier expression of genes enriched for mature (cortex and interneuron) cell types in the neocortex of dunnarts versus mice.

The developmental transcriptome of the neocortex is relatively protracted in dunnart

A recent study by Telley et al. (2016) identified large numbers of genes important for cortical specification across development. To do this, apical progenitor cells in the mouse developing neocortex were labelled with FlashTag intraventricular injections, and collected at different time points to identify progenitor-, neurogenic- and neuronal-specific markers via single cell RNA-seq. We used the gene lists from Telley et al. to make ranked lists of gene expression in our mouse and dunnart datasets. This showed that the majority of transcripts enriched in progenitor cells had a higher expression ranking in mouse than dunnart cortex during ECD, but had then switched to a majority of expression of progenitor cell genes in dunnart by LCD (Fig. 4A, top panels), indicating a protracted development in dunnart compared with mouse. A similar dynamic was observed for the expression rankings of neurogenic-enriched genes (Fig. 4A, middle panels). However, for the neuronal-enriched genes, this trend was reversed, with the majority of the genes showing a higher expression ranking in dunnart during ECD, which reverted to mouse during LCD (Fig. 4A, bottom panels). This finding strengthens the conclusion that, during ECD when equivalent cortical layers are being born in each species, the mouse cortex has a transcriptomic profile more strongly associated with early cortical events, whereas the dunnart is precociously mature, and that this trend is ablated or reversed by LCD, indicating protracted development in dunnarts. We then further examined whether this change in relative expression profile between these two stages was due to a change in the proportion of cell types occupying the cortex. We immunostained and quantified developing brains of both species using known markers of progenitor cells (*Pax6* and *Sox2*) as well as mature neuron markers (*NeuroD1* and *Thrl*), some of which were enriched in the Telley et al. cell-type specific clusters (Fig. 4A). The results showed that, during ECD, the proportion of cortical thickness immunopositive for progenitor cell markers was significantly larger in mice than in dunnarts ($*P<0.05$, Fig. 4B,C,J; Fig. S2), whereas during LCD the proportion was similar between species (Fig. 4F,G,K). In contrast, the proportion of cortical thickness immunopositive for neuronal mature neuron markers during ECD was significantly larger in dunnarts than mice ($*P<0.05$, Fig. 4D,E,J; Fig. S2), whereas during LCD the relative staining pattern between species for *NeuroD1* had switched, with the proportion of immunopositive cortical thickness being significantly larger in mice than dunnarts ($*P<0.05$, Fig. 4H,K; Fig. S2). For *Thrl*, the proportion of immunopositive cortical thickness during LCD was statistically similar across species ($P>0.05$, Fig. 4I,K; Fig. S2). We have previously shown, using a marker for post-mitotic cortical neurons, that the expansion of the cortical plate in both mouse and dunnart is a broadly additive process across multiple stages of development (Paolino et al., 2020). We therefore interpret this data as two time points along a relatively linear continuum of cortical plate growth in both species.

An additional heatmap analysis of mouse-dunnart expression rank differences was performed for genes previously reported to be enriched in six successive transcriptional waves in mouse single cell analyses (Telley et al., 2019). These waves were determined by

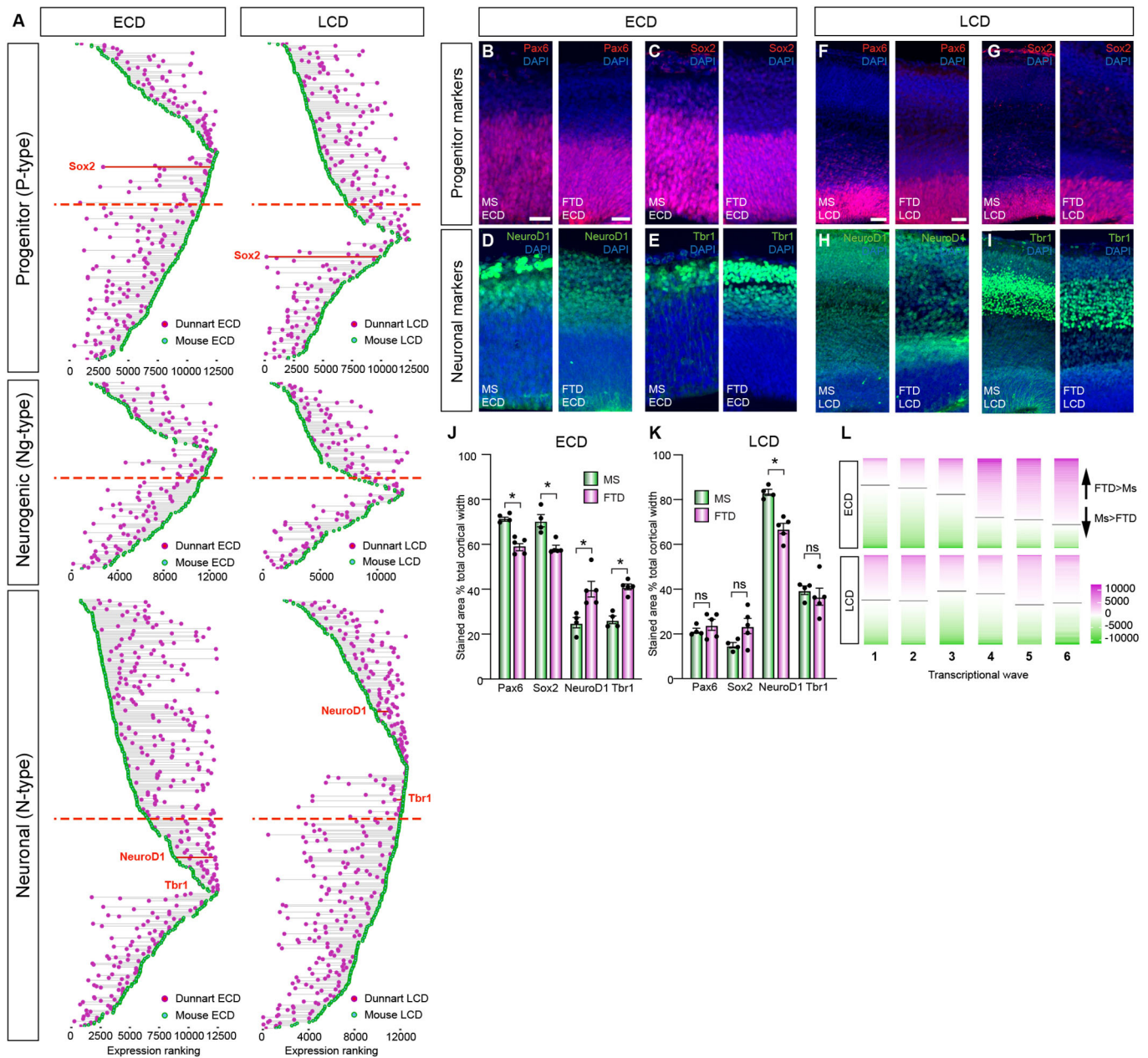


Fig. 4. Visualisation and analysis of ranked expression for genes enriched in reported single cell analyses of cortical cell fate determination. (A) Genes that were enriched in the progenitor, neurogenic or neuronal cell states identified by Telley et al. (2016), and the mouse-dunnart orthologue list, displayed as dumbbell plots of the mouse-dunnart ranked expression values. The results show a trend towards advanced maturation during ECD in dunnart (FTD) relative to mouse (Ms), which reverses during LCD. Dashed red lines indicate the midpoint of each gene list. Each dumbbell (row) in the dumbbell plots shows the expression rank values for a single gene for dunnart (magenta point) and mouse (green point) connected by a horizontal grey bar. (B-G) Immunohistochemical labelling of known markers of progenitor (B,C,F,G) and neuronal (D,E,H,I) cells show that this trend is reflected in the proportion of cortical width dedicated to each population (quantified in J,K). (L) Heatmaps of mouse-dunnart difference of expression ranking for genes found to be enriched during six successive transcriptional waves from least to most mature, previously identified in mouse E12 and E15 single cell analyses (Telley et al., 2019). The minimum absolute value indicating the inflection point between positive and negative values is marked by the horizontal grey bar, and shows a progressive interspecies shift during ECD and no distinct pattern of change during LCD. Data are mean \pm s.e.m. * $P < 0.05$, Mann-Whitney U -test. ns, not significant. Scale bars: 25 μ m, Ms and FTD (B-E), FTD (F-I); 50 μ m, Ms (F-I).

analysing patterns of single cell gene expression across cellular clusters identified along an axis of pseudo-differentiation, with clusters ranging from having a least-differentiated (wave 1) to most differentiated (wave 6) transcriptional profile. When we compared our expression ranking of mouse and dunnart at equivalent ages with the gene lists demarcating these transcriptional wave profiles, we found little interspecies differences during late cortical

development (Fig. 4L, bottom row), with the inflection point of expression rank differences (horizontal grey bar) showing no distinct pattern of change across waves. However during ECD, the majority of the genes in waves 1-3 (less differentiated) show greater ranked expression in mice (Fig. 4L, top row, green shading), whereas the majority of the genes in waves 4-6 (more differentiated) show greater ranked expression in dunnarts than in mice

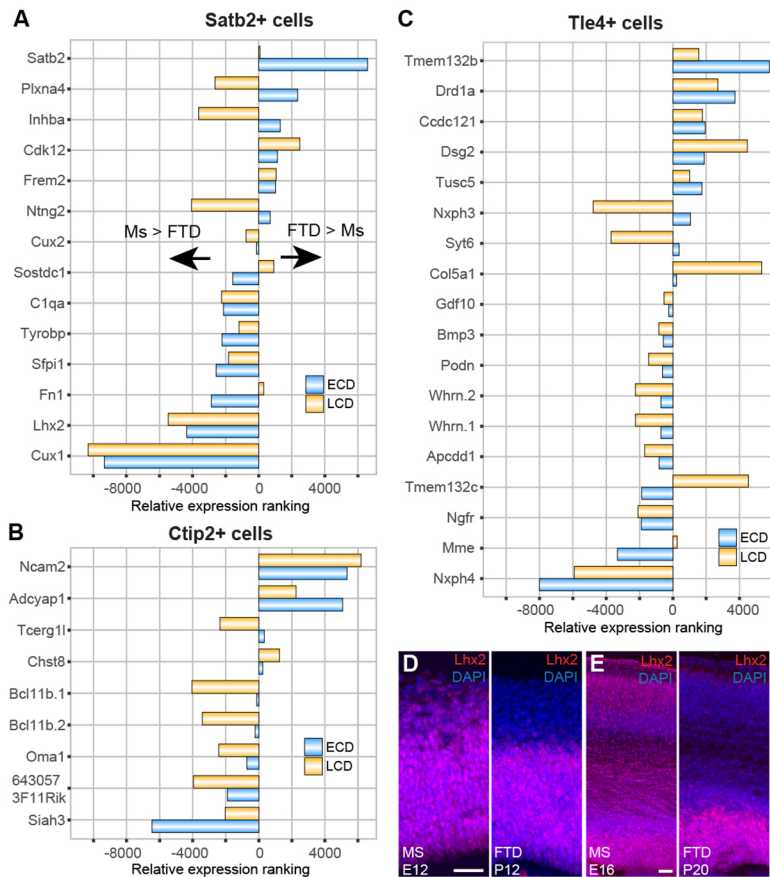


Fig. 5. Patterns of differential interspecies expression ranking for genes enriched in neocortical projection neuron subtypes. (A–C) Expression rank value differences for ECD and LCD transcriptomes between dunnart (FTD) and mouse (Ms) were plotted against previously published datasets revealing unique identifying genesets demarcating three neocortical projection neuron subtypes via immuno-isolation from mouse cortex between E15.5 and P1 (Molyneaux et al., 2014). The three subtypes correspond to corticocortical (*Satb2*+, A), subcerebral (*Ctip2*+, B) and corticothalamic (*Tle4*+, C). This comparison did not reveal any striking species-specific global trends in ranked expression difference for any of the projection neuron subtypes, but did highlight some differentially expressed candidate genes that may contribute to known differences in interspecies connectivity patterns. (D) *Lhx2*, a highly differentially expressed candidate, also showed strikingly different expression patterns upon immunohistochemical labelling in both species and at both ages. Scale bars: 50 μ m, Ms and FTD (D), Ms (E); 25 μ m FTD (E).

(Fig. 4L, top row, magenta shading). Taken together, these results further reveal a heterochronic state of cortical development between mice and dunnarts during ECD, with a relative emphasis on mitotic and progenitor-cell processes in the mouse cortex, and a relative emphasis on post-mitotic and neuronal processes in the dunnart cortex. The potential cortical heterochrony is less clear by LCD, likely because both species have shifted toward post-mitotic processes, although there is evidence, such as the *NeuroD1* staining (Fig. 4H,K), that suggests a greater degree of neuronal differentiation in mice during LCD. Given the major differences in interhemispheric connectivity between mouse and dunnart, we might expect to find major differences in defined populations of neurons in LCD, and we sought to examine this using a deeper level of analysis.

We compared our expression ranking of mouse and dunnart orthologous transcripts against a published dataset that provided transcriptomes for the three main projection neuron subtypes in the cortex, identified via immunohistochemical cell sorting: corticocortical neurons (*Satb2*+), subcerebrally projecting neurons (*Ctip2*+, also known as *Bcl11b*) and corticothalamic neurons (*Tle4*+) (Molyneaux et al., 2014). Orthologue genes in the Molyneaux et al. dataset that were not annotated in the dunnart transcriptome were identified using additional mapping methodology described in the Materials and Methods. One of the most remarkable differences in cortical circuits between therians include corticocortical connections between hemispheres exclusively via the anterior commissure in marsupials, whereas eutherians exclusively evolve the corpus callosum as the main interhemispheric connection. Surprisingly, unlike the analyses of maturation patterns, and despite the remarkably different projection

routes between marsupials and eutherians, this analysis did not reveal general interspecies trends towards differential enrichment in any of these gene lists demarcating cortical neuron subtypes (Fig. 5A–C). This indicated overall conservation of genes involved in cortical specification. However, our analysis did reveal individual genes that showed marked changes in relative ranking, such as *Satb2*, which showed a higher expression ranking in the dunnart during ECD, and similar expression rankings between species during LCD (Fig. 5A). Interestingly, this is consistent with our recent findings of differential *Satb2* expression between the two species via immunostaining (Paolino et al., 2020), as well as using species-specific knockdown and overexpression constructs to determine that this differential expression is crucially involved in differential routes of axonal projection. Other intriguing candidates that are markedly reduced in expression ranking in dunnart relative to mouse at both ages include the transcription factors *Cux1* and *Lhx2*. *Cux1* showed a substantially larger level of relative mRNA expression in the mouse cortex at both stages of cortical development. Interestingly, *Cux1* has recently been identified as a key candidate involved in translational priming of upper layer neurons, where mRNA encoding proteins that will be expressed in mature progeny accumulate but are not translated by parent cells (Zahr et al., 2018) that are present at earlier developmental stages (such as our ECD). Future studies are needed to better understand whether diversification of translational priming is an important evolutionary mechanism in the neocortex. *Lhx2* is a transcription factor known to regulate cell differentiation during cortical development (Monuki et al., 2001; Mangale et al., 2008), and our immunostaining in both stages and species revealed different localisation of the protein, with almost all of the cortex showing

immunostaining at both ages in mice, whereas the dunnarts showed a more restricted pattern of protein expression in the proliferative zones (Fig. 5D,E). Taken together, these findings confirm the validity of our analytical methods in discerning key interspecies differences in gene sets and single genes relevant to evolutionary changes in cortical development between therians.

DISCUSSION

Collectively, these results revealed a clear trend towards a greater disparity in relative gene expression between species during early cortical development (ECD; during deeper layer neurogenesis) than late cortical development (LCD; during upper layer neurogenesis). Dunnarts have a relatively more mature presumptive cortex at the time of deeper layer neurogenesis. This trend is in accord with previously published ‘event scales’ of whole brain development between species, suggesting a relatively fast maturation of marsupials at early stages compared with eutherians, followed by a more protracted rate of development at later stages (Darlington et al., 1999; Workman et al., 2013). The advanced maturation of the early cortex during our earlier stage may be due to the temporal scaling of the more protracted development of marsupials increasing the overlap of processes, and/or due to the behavioural requirements of the extraordinarily altricial birth of marsupials relative to eutherians, as they need to locate and attach to a teat to complete development. Regardless, the implications of the advanced state of the cortex during the period of deeper layer cell birth is an intriguing mechanism that could provide insight into the subtle differences in cortical architecture that exist between groups. An elegant example of how significant such differences in timing might be can be found in a hypothesis pioneered by Rakic, that evolutionary changes in proliferation kinetics of neocortical progenitors underlies human neocortical expansion (Rakic, 1995, 2009). One example of how changes in transcriptomic timing could affect the evolution of complex traits in eutherians versus marsupials is via different cellular and molecular cues presented to the pioneering interhemispheric neocortical axons, which may affect axon guidance strategies between species across the corpus callosum versus the anterior commissure. Indeed, we recently showed that the manipulation of one post-mitotic transcription factor in mouse to match the precocious marsupial timescale was sufficient to phenocopy aspects of marsupial connectivity, demonstrating that the differential timing of transcription factors between species may underlie the emergence of many complex traits (Paolino et al., 2020). Although our conclusions are predominantly based on orthologous protein-coding transcripts identified between dunnart and mouse, future studies using RNA-seq data mapped against reference genomes will be required to further discover absent/novel genes, as well as additional regulatory elements. Taken together, this work reveals a heterochronic neocortical maturation program across two stages between representative species of Theria, presents rich resources of developmental brain transcriptomes and provides a foundation for future manipulation studies to probe the functional relevance of candidate genes to the evolution of neocortical circuits.

MATERIALS AND METHODS

Animals

Fat-tailed dunnarts (*Sminthopsis crassicaudata*) were bred at the Native Wildlife Teaching and Research Facility at The University of Queensland, Australia, as previously described (Suárez et al., 2017). The pouches of breeding females were checked daily, and the day joeys were detected was regarded as P0. Two postnatal ages were used in this investigation, P12 and P20, with three joeys used at each age and each specimen from a different

litter. All animal procedures were approved by The University of Queensland Animal Ethics Committee and the Queensland Government Department of Environment and Heritage Protection.

5-ethynyl-2'-deoxyuridine injections

Intraperitoneal injections of 5-ethynyl-2'-deoxyuridine (EdU; 1 µg per 1 g of body weight, diluted in water) were performed in dunnart joeys at P12 or P20, being careful to not disturb their attachment to the dam's teat. In mice, time-mated CD1 pregnant dams were injected with EdU (5 µg per 1 g of body weight, diluted in water) at E12 or E16, and then isolated in separate cages until giving birth. The dunnart joeys and mouse pups were euthanised at an equivalent stage after cortical lamination (dunnart P36 and mouse P5) by transcardial perfusion with 4% paraformaldehyde. The EdU solution was injected via a pulled borosilicate glass capillary and using a Picospritzer II microinjection system.

RNA extraction

Joeys were removed from the pouch, anaesthetised on ice for several minutes and sacrificed by decapitation. Both hemispheres of the neocortex were micro-dissected out immediately in ice-cold sterile PBS, transferred into TRIzol reagent (Thermo Fisher Scientific) and homogenised using 23- and 27-gauge needles. Total RNA was extracted using the standard TRIzol protocol. The RNA pellet was then resuspended in RNase-free water, treated with DNaseI (79254, Qiagen) and column-purified using an RNeasy MinElute kit (74204, Qiagen). All dunnart RNA samples showed an RNA integrity number (RIN) of 8.6 or higher on a Bioanalyzer 2100 RNA Nano assay (Agilent).

The P12 dunnart data was compared with three existing age-matched (ECD) wild-type C57BL/6 mouse RNA-seq datasets of dorsal cortex at E12 from the publicly available NCBI Sequence Read Archive (SRR1509162, SRR1509163, SRR1509164). Similarly, the P20 dunnart data was compared with four existing age-matched (LCD) wild-type C57BL/6 mouse neocortical datasets at E16, also from the NCBI Sequence Read Archive (SRR5755669, SRR5755670, SRR5755671, SRR5755672) (Bunt et al., 2017).

Dunnart cDNA library preparation and sequencing

Approximately 400 ng of total RNA from each dunnart sample was used for cDNA library preparation. The RNA samples were first depleted of cytoplasmic and mitochondrial ribosomal RNA using Ribo-Zero (15065382, Illumina) and then fragmented. A Truseq Stranded Total RNA kit v4 (15031048, Illumina) was then used to perform first- and second-strand cDNA synthesis, followed by adapter ligation and PCR amplification of the cDNA products, as per the standard low sample (LS) protocol. The cDNA libraries were sequenced on an Illumina HiSeq 2000 with 125 bp paired-end reads using the standard manufacturer's instructions. Image processing and sequence data extraction were performed using the standard Illumina Genome Analyzer software and CASAVA (version 1.8.2), which generated raw sequencing reads in fastq format for each individual sample. In total, 916.8 million RNA-seq reads were produced for the six samples (an average of 152.4 million reads per sample). The dataset is available from Gene Expression Omnibus (GEO) under accession number GSE161274.

Data preprocessing

In dunnart, cutadapt (version 1.18) (cutadapt.readthedocs.io) was used to trim low-quality nucleotides at both ends of the reads, with the parameter setting of ‘-q 20,20 –minimum-length=36’. After trimming, the clean P12 dunnart datasets 436, 451 and 458, and clean P20 datasets 666, 746 and 758, were 24.8, 23.9, 22.4, 10.1, 10.3, and 17.3 GB, respectively.

In mouse, the ‘fastq-dump’ function of the SRA Toolkit (ncbi.github.io/sra-tools) was used to extract the fastq read files with the parameters ‘-gzip –read-filter pass –skip-technical –dumpbase –clip’ for the E12 dataset, and an additional parameter of ‘-split-3’ for the E16 dataset. Cutadapt was then used to trim low-quality nucleotides in the E12 data with ‘-q 20,20 –minimum-length=12’. To avoid bias of read mapping and length between mouse stages, only the forward read of the paired-end E16 data was

used, as well as trimming the 3' half of the 100 bp E16 reads (additional cutadapt parameter of '-u -50'), in order to match the 50 bp single-end reads of the E12 data.

Dunnart *de novo* transcriptome assembly and functional annotation

An annotated reference genome for dunnarts has not yet been published, therefore, the transcriptome was assembled *de novo* across both stages and used as a reference for the mapping of dunnart reads. The Trinity software package (v2.3.2) (Grabherr et al., 2011; Haas et al., 2013) was used to perform the *de novo* assembly with the parameter setting of '-seqType fq -SS_lib_type RF -max_memory 400G -left read1.fastq -right read2.fastq -output output_directory -CPU 8'. Following assembly, transcript abundance was estimated using RSEM (RNA-Seq by Expectation Maximization; v1.3.1) (Li and Dewey, 2011) and protein-coding regions were extracted using TransDecoder (v3.0.1; github.com/TransDecoder/TransDecoder/wiki). Annotation of the Trinity-generated transcripts and TransDecoder-generated peptides was performed using Trinotate (v3.0.2) (Bryant et al., 2017) with three methods: a Blastx (v2.6.0) (Camacho et al., 2009) search against SwissProt and UniProt databases, protein domain identification using HMMER (v3.1b2; hmmer.org) against the PFAM database, and transmembrane regions prediction using tmHMM (v2.0c; www.cbs.dtu.dk/services/TMHMM). The Trinotate annotations were primarily used for the identification of orthologous sequences to assist in the cloning of transcripts into overexpression vectors or for the generation of *in situ* hybridisation probes.

Mammalian genome alignment

Genome sequences were downloaded from NCBI, including human (GRCh38), mouse (GRCm38), Tasmanian devil (DEVIL7), wallaby (Meug1.0) and opossum (MonDom5). Gene annotation files were downloaded from the Ensembl database (release 81). The program GMAP (version: 2015-07-23) was used to align dunnart transcripts against these five reference genomes, separately, with the parameter setting of '-cross-species -batch=4 -prunelevel=3 -tolerant -npaths=0 -format=gff3_gene -ordered -nofails -t 16'. The program BEDOPS (v2.4.25) (Neph et al., 2012) was then used to extract the intersection of the GMAP alignment results with Ensembl gene features, requiring at least 50% coverage of the Ensembl feature, and the output was filtered in the bash shell to produce a non-redundant list of Ensembl gene identifiers.

BUSCO analysis

In order to assess the completeness of content of the dunnart transcriptome assembly, an alignment to the mammalian lineage file of OrthoDB orthologous genes groups (mammalia_odb9, www.orthodb.org) was performed using the BUSCO (version 3.0.2) python script 'run_BUSCO.py' (Seppey et al., 2019).

Cloning dunnart coding mRNA

Fasta sequences annotated with the protein symbol of interest were isolated from the dunnart Trinity database and then submitted to the University of California, Santa Cruz (UCSC) BLAT online tool for alignment against the sarHar1 Tasmanian devil genome, the only Dasyurid reference genome available. The alignment results were then used to generate a consensus sequence that was aligned against mouse Refseq sequences to determine the extent of its completion. RT-PCR products were generated from these consensus sequences using P45 dunnart cortical total RNA template, and cloned into pGEM-T vectors (Promega). The first genes, *Satb2* and *Ctip2*, were previously successfully cloned with this method and used to generate uracil-bound digoxigenin *in situ* hybridisation probes, then subcloned into a CAG promoter-based vector for cellular overexpression (Paolino et al., 2020). The mRNA of *Crh* (corticotropin releasing hormone) was cloned with this method, given that it is one of the few dunnart transcripts that has been previously sequenced and uploaded to a public repository (GenBank number KT380843.1). PCR primers used were: forward AGTTCCTA-GACCATGAAGCTCC, reverse AAACGGGATGTCTCACTTTC.

Identification of orthologous gene pairs between dunnart and mouse

To determine the orthologous gene pairs between dunnart and mouse, reciprocal Blastp searches were performed between dunnart and mouse. First, the dunnart TransDecoder-generated peptides were used as query sequences for a Blastp search against all mouse protein sequences, and the top hit for each dunnart peptide was assigned. Secondly, mouse protein sequences were used as queries for a Blastp search against all dunnart peptides, and the top hit for each mouse protein sequence was assigned. Only Blastp results with a one-to-one top hit (multi-mapping transcripts were not considered) and an E-value less than a threshold of 10E-5 were used. Transcripts for which the Blastp hits were reciprocally matched between species were deemed to be present in both species and parsed as orthologous gene pairs between dunnart and mouse.

Intra-species differential expression analysis

RSEM was used to perform alignment and counting of the *de novo*-assembled dunnart transcripts. The functions 'extract-transcript-to-gene-map-from-trinity' and 'rsem-prepare-reference' were used to generate a reference against which the reads were mapped and counted using Bowtie2 (version 2.3.4.3) (Langmead and Salzberg, 2012) and the function 'rsem-calculate-expression'. In mouse, the reads were mapped against the mm10 genome (igenomes.illumina.com.s3-website-us-east-1.amazonaws.com/Mus_musculus/UCSC/mm10/Mus_musculus_UCSC_mm10.tar.gz) using Hisat2 (version 2.1.0) (Kim et al., 2019). The resulting SAM files were then sorted by alignment position using samtools (version 1.9) (Li et al., 2009) and the reads counted using the 'htseq-count' function of HTSeq (version 0.11.2) (Anders et al., 2015).

Differential gene expression across developmental stages within each species was calculated with an exact test pairwise comparison of the counts of mapped reads, normalised against the library size factor of each sample, using the *edgeR* Bioconductor package (Robinson et al., 2010; McCarthy et al., 2012). The lowest expressed transcripts were filtered out by only keeping transcripts that showed a raw counts-per-million (CPM) value >1 in at least two samples. *P*-values were adjusted for FDR to correct for multiple testing of differential gene expression. Volcano plots of log2 fold change values against -log10-transformed FDR-adjusted *P*-values were generated using the ggplot2 R package.

For the calculation of FPKM in dunnart, the 'effective length' of the RSEM mapping results was used, which corresponds to the mean number of positions within a transcript that align with the mapped fragments. Furthermore, the effective lengths of genes were calculated as the weighted value of transcript isoform effective lengths based on the percentage of individual isoform abundance relative to the total transcript abundance for that gene. A similar method of gene length determination relative to isoform abundance was used for mouse, with transcript lengths extracted from the EBI Refseq database (www.ebi.ac.uk/Tools/dbfetch) using a custom script, which used the urllib URL handling python module for Refseq sequence reading.

Heatmaps were created to visualise the expression profile of mouse-dunnart orthologue transcripts across all experimental groups (both species and developmental stages). First, orthologue transcripts present in all samples were identified. Then the mean FPKM values for each group were transformed by addition of a 0.1 positive constant and log (base 10) applied. Orthologue genes were sorted by increasing mean transformed FPKM value, and expression heatmaps and dendrograms generated using the pheatmap R package. Dendrograms were determined based on correlation dissimilarity of transformed FPKM values across groups.

Gene ontology analysis

For the intra- and interspecies expression analyses, enrichment of the 'biological process' branch of gene ontologies was determined using the GOATOOLS package (version 0.9.9) (Klopfenstein et al., 2018), which consists of several python scripts. This was performed in three steps. First, enrichment of ontologies in a gene list was determined with the script 'find_enrichment.py', which uses Fisher's exact test with *P*-values adjusted for FDR to correct for multiple testing, against the background of the

‘org.Mm.eg.db’ mouse genome-wide gene annotation list and NCBI’s ‘gene2go’ association list (ftp.ncbi.nlm.nih.gov). Second, ‘header’ ontology terms that summarise and group the list of ontologies enriched in the previous step are determined with the script ‘wr_sections.py’. Finally, header ontologies were filtered to remove headers at level 2 or higher in the ontology hierarchy, given these described functions that were too broad for meaningful interpretation. For headers at level 2 in the intra-species comparison, all enriched ontology children at levels 3–5 were retained and became enriched upon running the ‘wr_sections.py’ script again. In the interspecies comparison, a smaller number of ontologies were enriched given the smaller gene lists analysed, therefore, the level 2 enriched ontologies were mined to extract more information. For headers at level 2, the highest level children were retained and used to make new headers that became enriched upon running the ‘wr_sections.py’ script again. The process of identifying and enriching the highest level children was repeated until all children in the original level 2 group were represented by the newly enriched child ontologies that are level 3 or lower.

ENCODE dataset processing

A developmental series of B6NCRl mouse forebrain polyA-enriched RNA-seq data from the ENCODE project (www.encodeproject.org) was used for principal components analysis with the dunnart and NCBI-derived mouse datasets. The developmental series included data at E10 (ENCFF920CNZ, ENCFF320FJX, ENCFF528EVC, ENCFF663SNC), E11 (ENCFF329ACL, ENCFF896COV, ENCFF251LNG), E12 (ENCFF920QAY, ENCFF294JRP, ENCFF203BWA, ENCFF700OLU), E13 (ENCFF235DNM, ENCFF959PSX), E14 (ENCFF270GKY, ENCFF460TCF, ENCFF126IRS, ENCFF748SRJ), E15 (ENCFF179JEC, ENCFF891HIX), E16 (ENCFF931IVO, ENCFF114DRT) and P0 (ENCFF358MFI, ENCFF037JQC, ENCFF447EXU, ENCFF458NWF). The ENCODE dataset was processed similarly to the NCBI mouse data with respect to read trimming, alignment to the mm10 genome, sorting and read counting. To allow principal component analysis between the dunnart, NCBI mouse and ENCODE mouse datasets, each sample was normalised by library size and the common transcript Refseq IDs present in all samples were used.

Interspecies expression ranking analysis

To analyse the most highly upregulated transcripts in dunnart relative to mouse at each developmental stage, common transcripts between the top 30% expressed transcripts in dunnart and the bottom 30% expressed transcripts in the mouse were identified from the ranked data shown in Table S4. Similarly, to analyse the most highly upregulated transcripts in mouse relative to dunnart at each developmental stage, common transcripts were identified between the top 30% expressed transcripts in mouse and the bottom 30% expressed transcripts in dunnart.

Immunohistochemistry

Preparation of cortical tissue for immunohistochemistry is described in Paolino et al. (2020). Briefly, adult female dunnarts with joeys at P12 or P20 were anaesthetised in a gas induction chamber using 5% isoflurane in oxygen. The dunnart was transferred to a silicone mask and the anaesthesia maintained at 2–5% isoflurane. Joeys ($N=5$ each stage) were then collected by gently removing from the teat using forceps. They were exposed to 5–10 min of ice anaesthesia and transcardially perfused with 0.9% saline (NaCl), followed by 4% paraformaldehyde (PFA) in $1\times$ PBS. Mouse embryos at E12 or E16 ($N=5$ each stage) were collected from time-mated pregnant C57BL/7 dams that were anaesthetised via intraperitoneal injection of sodium pentobarbitone (190 mg/kg). The embryos were decapitated and drop fixed in 4% PFA. The brains were dissected out of the skull, embedded in 3.4% agarose, and 50 μ m coronal sections made using a vibratome. Sections containing the primary somatosensory cortex were mounted onto glass slides, dried enough to ensure tissue adherence to the slide, and post-fixed with 4% PFA for 10 min. Antigen retrieval was performed using a solution of 0.01M sodium citrate and 0.05% Tween 20 (pH 6.0), and a thermal program of 110°C for 4 min. The slides were blocked in a solution of 10% v/v normal donkey serum (NDS) and 0.2% Triton X-100 in PBS (pH 7.4) for 2 h, followed by overnight incubation in

blocking solution with primary antibodies (Table S1). The slides were incubated in the appropriate fluorescent secondary antibody for 3 h (Table S1), followed by a 10 min incubation in 0.1% 4',6-diamidino-2'-phenylindole dihydrochloride (DAPI), then coverslipped with an antifade mounting media. The slides were washed three times for 20 min in PBS after each antibody incubation, and all procedures were performed at room temperature.

Microscopy and image analysis

Microscopy equipment is described in Paolino et al. (2020). Briefly, high resolution fluorescence imaging was performed using a Spectral Applied Research Discovery spinning disk system on a Nikon TiE microscope using a 20 \times air objective. Cortical thickness measurements were performed manually using the native length measurement tool in Fiji (ImageJ). Only primary antibodies that showed dense laminar labelling, and therefore allowed for thickness quantification relative to the whole cortical plate, were used in this study. For each antibody, the mean percentage of immunopositive cortical labelling for each animal ($N\geq 4$ per age and species) was determined and visualised \pm s.e.m. using bar plots. Mann–Whitney U -tests were performed between these mean values across species for each antibody (Prism 7, GraphPad Software).

Mapping of selected non-annotated dunnart transcripts

Genes previously identified as highly enriched and potentially important cofactors in the development of cortical projection neuron subpopulations in mouse (Molyneaux et al., 2014) were analysed in the current dunnart and mouse datasets. Some of these genes were not annotated in the dunnart dataset, particularly the long non-coding RNAs (lincRNAs). The genomic locations of these genes were downloaded from the DeCON database (Molyneaux et al., 2014) and GMAP (version 2015-12-31) (Wu and Watanabe, 2005) was used to map these locations to dunnart transcripts. The GMAP parameter settings used were: ‘–nofails –Z –split-output=output_file –g query_fasta_file subject_fasta_file’. Trinity transcripts that map with at least 30% similarity were retained. To identify these DeCON sequences in mouse, the mm9 genomic locations were first converted to mm10 locations using the UCSC ‘liftOver’ webtool (genome.ucsc.edu/cgi-bin/hgLiftOver). The ‘GenomicRanges’ and ‘EnsDb.Mmusculus.v79’ Bioconductor packages were then used to find overlaps between the genomic locations and the mouse transcriptome. For lincRNA loci, only overlapping non-coding transcripts were retained.

Acknowledgements

We thank The University of Queensland Biological Resources and the Native Wildlife Teaching and Research Facility for establishing and maintaining the dunnart husbandry. We thank the Queensland Brain Institute Histology Facility and Advanced Microscopy Facility for assistance with histology preparation and imaging, and the Queensland Brain Institute DNA Sequencing Facility for sequencing the dunnart samples.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: P.K., R.S., L.J.R., L.R.F.; Formal analysis: P.K., Q.-Y.Z., L.R.F.; Investigation: P.K., A.P., L.R.F.; Data curation: P.K., Q.-Y.Z., A.P.; Writing - original draft: P.K., R.S., L.J.R., L.R.F.; Writing - review & editing: P.K., R.S., Q.-Y.Z., A.P., L.J.R., L.R.F.; Funding acquisition: R.S., L.J.R., L.R.F.

Funding

This work was supported by the National Health and Medical Research Council (NHMRC) (CJ Martin Fellowship 1035093 to P.K. and Principal Research Fellowship 1120615 to L.J.R.), Australian Research Council (Discovery Early Career Researcher Award DE160101394 to R.S., Discovery Project Grant DP160103958 to L.J.R. and R.S., and Discovery Project Grant DP200103093 to R.S. and L.R.F.), a University of Queensland (UQ)-Queensland Brain Institute (QBI) Doctoral Scholarship (A.P.), a UQ Development Fellowship and NHMRC Investigator Grant 1175825 (L.R.F.), and a UQ Amplify Fellowship (R.S.).

Data availability

The dunnart transcriptomic reads have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE161274.

Peer review history

The peer review history is available online at <https://journals.biologists.com/dev/article-lookup/doi/10.1242/dev.200212>

References

- Anders, S., Pyl, P. T. and Huber, W. (2015). HTSeq - a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169. doi:10.1093/bioinformatics/btu638
- Arai, Y., Pulvers, J. N., Haffner, C., Schilling, B., Nüsslein, I., Calegari, F. and Huttner, W. B. (2011). Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nat. Commun.* **2**, 154. doi:10.1038/ncomms1155
- Briscoe, S. D. and Ragsdale, C. W. (2018). Homology, neocortex, and the evolution of developmental mechanisms. *Science* **362**, 190-193. doi:10.1126/science.aau3711
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G. et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell. Rep.* **18**, 762-776. doi:10.1016/j.celrep.2016.12.063
- Bunt, J., Osinski, J. M., Lim, J. W., Vidovic, D., Ye, Y., Zalucki, O., O'Connor, T. R., Harris, L., Gronostajski, R. M., Richards, L. J., et al. (2017). Combined allelic dosage of Nfia and Nfib regulates cortical development. *Brain Neurosci. Adv.* **1**, 21. doi:10.1177/2398212817739433
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421. doi:10.1186/1471-2105-10-421
- Cheung, A. F. P., Kondo, S., Abdel-Mannan, O., Chodroff, R. A., Sirey, T. M., Bluy, L. E., Webber, N., DeProto, J., Karlen, S. J., Krubitzer, L. et al. (2010). The subventricular zone is the developmental milestone of a 6-layered neocortex: comparisons in metatherian and eutherian mammals. *Cereb. Cortex* **20**, 1071-1081. doi:10.1093/cercor/bhp168
- Darlington, R. B., Dunlop, S. A. and Finlay, B. L. (1999). Neural development in metatherian and eutherian mammals: variation and constraint. *J. Comp. Neurol.* **411**, 359-368. doi:10.1002/(SICI)1096-9861(19990830)411:3<359::AID-CNE1>3.0.CO;2-J
- Faunes, M., Francisco Botelho, J., Ahumada Galleguillos, P. and Mpodozis, J. (2015). On the hodological criterion for homology. *Front. Neurosci.* **9**, 223. doi:10.3389/fnins.2015.00223
- Gobius, I., Morcom, L., Suárez, R., Bunt, J., Bukshpun, P., Reardon, W., Dobyns, W. B., Rubenstein, J. L. R., Barkovich, A. J., Sherr, E. H. et al. (2016). Astroglial-mediated remodeling of the interhemispheric midline is required for the formation of the Corpus callosum. *Cell. Rep.* **17**, 735-747. doi:10.1016/j.celrep.2016.09.033
- Gobius, I., Suárez, R., Morcom, L., Paolino, A., Edwards, T. J., Kozulin, P. and Richards, L. J. (2017). Astroglial-mediated remodeling of the interhemispheric midline during telencephalic development is exclusive to eutherian mammals. *Neural Dev.* **12**, 9. doi:10.1186/s13064-017-0086-1
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652. doi:10.1038/nbt.1883
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M. et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494-1512. doi:10.1038/nprot.2013.084
- Johnson, R. N., O'Meally, D., Chen, Z., Etherington, G. J., Ho, S. Y. W., Nash, W. J., Grueber, C. E., Cheng, Y., Whittington, C. M., Dennison, S. et al. (2018). Adaptation and conservation insights from the koala genome. *Nat. Genet.* **50**, 1102-1111. doi:10.1038/s41588-018-0153-5
- Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907-915. doi:10.1038/s41587-019-0201-4
- Klopfenstein, D. V., Zhang, L., Pedersen, B. S., Ramirez, F., Vesztrocy, A. W., Naldi, A., Mungall, C. J., Yunes, J. M., Botvinnik, O., Weigel, M. et al. (2018). GOATOOLS: a python library for gene ontology analyses. *Sci. Rep.* **8**, 10872. doi:10.1038/s41598-018-28948-z
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923
- Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323. doi:10.1186/1471-2105-12-323
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079. doi:10.1093/bioinformatics/btp352
- Loo, L., Simon, J. M., Xing, L., McCoy, E. S., Niehaus, J. K., Guo, J., Anton, E. S. and Zylka, M. J. (2019). Single-cell transcriptomic analysis of mouse neocortical development. *Nat. Commun.* **10**, 134. doi:10.1038/s41467-018-08079-9
- Mangale, V. S., Hirokawa, K. E., Satyaki, P. R. V., Gokulchandran, N., Chikbire, S., Subramanian, L., Shetty, A. S., Martynoga, B., Paul, J., Mai, M. V. et al. (2008). Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. *Science* **319**, 304-309. doi:10.1126/science.1151695
- McCarthy, D. J., Chen, Y. and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288-4297. doi:10.1093/nar/gks042
- Mikkelsen, T. S., Wakefield, M. J., Aken, B., Amemiya, C. T., Chang, J. L., Duke, S., Garber, M., Gentles, A. J., Goodstadt, L., Heger, A. et al. (2007). Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167-177. doi:10.1038/nature05805
- Miller, W., Hayes, V. M., Ratan, A., Petersen, D. C., Wittekindt, N. E., Miller, J., Walenz, B., Knight, J., Qi, J., Zhao, F. et al. (2011). Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proc. Natl. Acad. Sci. USA* **108**, 12348-12353. doi:10.1073/pnas.1102838108
- Molyneaux, B. J., Goff, L. A., Brettler, A. C., Chen, H.-H., Brown, J. R., Hrvatin, S., Rinn, J. L. and Arlotta, P. (2014). DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85**, 275-288. doi:10.1016/j.neuron.2014.12.024
- Monuki, E. S., Porter, F. D. and Walsh, C. A. (2001). Patterning of the dorsal telencephalon and cerebral cortex by a roof Plate-Lhx2 pathway. *Neuron* **32**, 591-604. doi:10.1016/S0896-6273(01)00504-9
- Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919-1920. doi:10.1093/bioinformatics/bts277
- Paolino, A., Fenlon, L. R., Kozulin, P., Richards, L. J. and Suarez, R. (2018). Multiple events of gene manipulation via in pouch electroporation in a marsupial model of mammalian forebrain development. *J. Neurosci. Methods* **293**, 45-52. doi:10.1016/j.jneumeth.2017.09.004
- Paolino, A., Fenlon, L. R., Kozulin, P., Haines, E., Lim, J. W. C., Richards, L. J. and Suárez, R. (2020). Differential timing of a conserved transcriptional network underlies divergent cortical projection routes across mammalian brain evolution. *Proc. Natl. Acad. Sci. USA* **117**, 10554-10564. doi:10.1073/pnas.1922422117
- Rakic, P. (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* **18**, 383-388. doi:10.1016/0166-2236(95)93934-P
- Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724-735. doi:10.1038/nrn2719
- Renfree, M. B., Papenfuss, A. T., Deakin, J. E., Lindsay, J., Heider, T., Belov, K., Rens, W., Waters, P. D., Pharo, E. A., Shaw, G. et al. (2011). Genome sequence of an Australian kangaroo, *Macropus eugenii*, provides insight into the evolution of mammalian reproduction and development. *Genome Biol.* **12**, R81. doi:10.1186/gb-2011-12-8-r81
- Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140. doi:10.1093/bioinformatics/btp616
- Seelke, A. M. H., Dooley, J. C. and Krubitzer, L. A. (2013). Differential changes in the cellular composition of the developing marsupial brain. *J. Comp. Neurol.* **521**, 2602-2620. doi:10.1002/cne.23301
- Seppely, M., Manni, M. and Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227-245. doi:10.1007/978-1-4939-9173-0_14
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212. doi:10.1093/bioinformatics/btv351
- Suárez, R., Gobius, I. and Richards, L. J. (2014). Evolution and development of interhemispheric connections in the vertebrate forebrain. *Front. Hum. Neurosci.* **8**, 497. doi:10.3389/fnhum.2014.00497
- Suárez, R., Paolino, A., Kozulin, P., Fenlon, L. R., Morcom, L. R., Englebright, R., O'Hara, P. J., Murray, P. J. and Richards, L. J. (2017). 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. doi:10.1371/journal.pone.0184450
- Suárez, R., Paolino, A., Fenlon, L. R., Morcom, L. R., Kozulin, P., Kurniawan, N. D. and Richards, L. J. (2018). A pan-mammalian map of interhemispheric brain connections predates the evolution of the corpus callosum. *Proc. Natl. Acad. Sci. USA* **115**, 9622-9627. doi:10.1073/pnas.1808262115
- Telley, L., Govindan, S., Prados, J., Stevant, I., Nef, S., Dermitzakis, E., Dayer, A. and Jabaudon, D. (2016). Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351**, 1443-1446. doi:10.1126/science.aad8361
- Telley, L., Agirman, G., Prados, J., Amberg, N., Fièvre, S., Oberst, P., Bartolini, G., Vitali, I., Cadilhac, C., Hippenmeyer, S. et al. (2019). Temporal patterning of apical progenitors and their daughter neurons in the developing neocortex. *Science* **364**, eaav2522. doi:10.1126/science.aav2522

- Udy, D. B., Voorhies, M., Chan, P. P., Lowe, T. M. and Dumont, S.** (2015). Draft de novo transcriptome of the rat kangaroo *Potorous tridactylus* as a tool for cell biology. *PLoS ONE* **10**, e0134738. doi:10.1371/journal.pone.0134738
- Workman, A. D., Charvet, C. J., Clancy, B., Darlington, R. B. and Finlay, B. L.** (2013). Modeling transformations of neurodevelopmental sequences across mammalian species. *J. Neurosci.* **33**, 7368-7383. doi:10.1523/JNEUROSCI.5746-12.2013
- Wu, T. D. and Watanabe, C. K.** (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-1875. doi:10.1093/bioinformatics/bti310
- Zahr, S. K., Yang, G., Kazan, H., Borrett, M. J., Yuzwa, S. A., Voronova, A., Kaplan, D. R. and Miller, F. D.** (2018). A translational repression complex in developing mammalian neural stem cells that regulates neuronal specification. *Neuron* **97**, 520-537.e6. doi:10.1016/j.neuron.2017.12.045