Spring 2021 – Epigenetics and Systems Biology
Discussion Session (Evolutionary Biology)
Michael K. Skinner – Biol 476/576
Week 14 (April 22)

Epigenetics and Evolutionary Biology

Primary Papers

1. Bogutz, et al. (2019) Nature Communication 10(1):5674. (PMID: 31831741)
2. Green, et al. (2019) Mol Ecol 28(16):3642. (PMID: 31338928)
3. Skinner, et al. (2014) Genome Biology and Evolution 6:1972-1989. (PMID: 25062919)

Discussion

Student 38 – Ref #1 above

- What was the role of the endogenous retroviruses?
- What epigenetic differences were observed between the species?
- What is the integration of genetics, epigenetics and evolution suggested?

Student 39 – Ref #2 above

- What is the model system and experimental design?
- What molecular and hormonal changes were observed in this environmental response?
- How could epigenetics be involved in the potential adaptive response?

Student 40 – Ref #3 above

- What was the model system and experimental design?
- What epigenetic observations were provided and how might environmental epigenetics impact evolution?
- Is this a Lamarckian contribution to evolution?

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-13662-9 OPEN

Evolution of imprinting via lineage-specific insertion of retroviral promoters

Aaron B. Bogutz[1,5], Julie Brind'Amour[1,5], Hisato Kobayashi[2,5], Kristoffer N. Jensen[1], Kazuhiko Nakabayashi[3], Hiroo Imai[4], Matthew C. Lorincz[1*] & Louis Lefebvre[1*]

Imprinted genes are expressed from a single parental allele, with the other allele often silenced by DNA methylation (DNAme) established in the germline. While species-specific imprinted orthologues have been documented, the molecular mechanisms underlying the evolutionary switch from biallelic to imprinted expression are unknown. During mouse oogenesis, gametic differentially methylated regions (gDMRs) acquire DNAme in a transcription-guided manner. Here we show that oocyte transcription initiating in lineage-specific endogenous retroviruses (ERVs) is likely responsible for DNAme establishment at 4/6 mouse-specific and 17/110 human-specific imprinted gDMRs. The latter are divided into Catarrhini- or Hominoidea-specific gDMRs embedded within transcripts initiating in ERVs specific to these primate lineages. Strikingly, imprinting of the maternally methylated genes *Impact* and *Slc38a4* was lost in the offspring of female mice harboring deletions of the relevant murine-specific ERVs upstream of these genes. Our work reveals an evolutionary mechanism whereby maternally silenced genes arise from biallelically expressed progenitors.

[1] Department of Medical Genetics, Molecular Epigenetics Group, Life Sciences Institute, University of British Columbia, Vancouver, BC V6T 1Z3, Canada. [2] Department of Embryology, Nara Medical University, Kashihara, Nara 634-8521, Japan. [3] Division of Developmental Genomics, Research Institute, National Center for Child Health and Development, Setagaya, Tokyo 157-8535, Japan. [4] Molecular Biology Section, Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan. [5] These authors contributed equally: Aaron B. Bogutz, Julie Brind'Amour, Hisato Kobayashi. *email: matthew.lorincz@ubc.ca; louis.lefebvre@ubc.ca

During postnatal oocyte growth in mammals, transcribed genomic regions acquire two characteristic epigenetic marks: transcription-coupled trimethylation at lysine 36 of histone H3 (H3K36me3)[1] and DNAme[2–5]. In both human and mouse female germline, the overall levels of CpG methylation increase from less than 5% in post-migratory primordial germ cells (PGCs) to ~40–50% in fully-grown, germinal vesicle oocytes (GVOs)[4,6–9]. In growing murine oocytes, 85–90% of this DNAme is deposited over H3K36me3-marked transcribed regions of the genome[3]. We recently demonstrated that SETD2-dependent deposition of H3K36me3 is required for de novo DNAme of transcribed gene bodies in mouse oocytes[10]. Such transcription-coupled de novo DNAme is also responsible for the establishment of regions of differential DNAme between the mature gametes (gametic differentially methylated regions, gDMRs), a subset of which are maintained throughout preimplantation development. This unique class of gDMRs, the imprinted gDMRs (igDMRs), can direct imprinted paternal allele-specific expression of the gene (s) under their regulation, the imprinted genes[3,11,12]. Importantly, transcription initiating within upstream oocyte-specific promoters has been reported to play a critical role in de novo DNAme at the maternal igDMRs of a number of mouse imprinted genes[3,13–17], but the evolutionary origin of such promoters remains unexplored.

Specific families of ERVs, also known as long terminal repeat (LTR) retrotransposons, are highly transcribed in mouse oocytes[3,18]. Such LTRs function as oocyte-specific promoters for both novel protein-coding genes and non-coding transcripts, the latter of which are widely distributed in intergenic regions[3,18,19]. We recently identified numerous LTR-initiated transcription units (LITs) in mouse oocytes associated with H3K36me3-coupled de novo DNAme[20]. Notably, CpG islands (CGIs) embedded within such LITs also become hypermethylated by this mechanism in oocytes[20] and the majority of maternal igDMRs associated with paternally expressed genes in both mice and humans overlap with CGIs[15,21]. Since ERVs are highly variable among mammalian species and de novo DNAme at maternal igDMRs is transcription-coupled, we hypothesized that active LTRs and their associated LITs may have played an important role in the genesis of lineage-specific maternal imprinting in mammals.

Here, we identify primate and rodent-specific igDMRs that appear to be de novo DNA methylated in oocytes as a consequence of transcription initiating within nearby LTR promoters. Our analysis of data from macaque and chimpanzee identifies Catarrhini- or Hominoidea-specific igDMRs embedded in oocyte transcripts emanating from species-specific LTRs. We further validate this phenomenon in two mouse mutants carrying deletions of upstream LTR promoters at *Impact* and *Slc38a4*, both of which lead to loss of oocyte DNAme acquisition at the igDMR in mutant females and loss of imprinted expression in the offspring. Together, our data suggest a model in which species-specific imprinted genes emerge from biallelically expressed progenitors via the acquisition of novel LTR promoters active during oocyte growth.

Results

LITs and the establishment of species-specific imprints. To investigate the contribution of LITs to maternal gametic imprinting in humans and mice, we first curated a list that includes well characterized and putative maternally methylated igDMRs[22–28] by interrogating published whole-genome bisulfite sequencing (WGBS) data from gametes, placenta and somatic tissues (detailed in Supplementary Data 1). In total, we identified 21 mouse and 125 human igDMRs that were previously validated through Sanger bisulfite sequencing, pyrosequencing, or whole-genome bisulfite sequencing (WGBS, see Methods and Supplementary Data 2). Among these, 6 and 110 maternal igDMRs are restricted to the mouse or the human lineage, respectively, while only 15 are conserved between the two species (Fig. 1a). We next applied de novo transcriptome assembly[29] to published human[30] and mouse[20] oocyte RNA-sequencing (RNA-seq) datasets and identified the transcript(s) and transcription start site (TSS) likely responsible for transcription-coupled deposition of DNAme over 20/21 mouse and 90/125 human igDMRs (Supplementary Data 2 and Supplementary Fig. 1a; identified using LIONS or de novo transcript assembly, see Methods)[29,31,32]. Among these, four mouse and 17 human maternal igDMRs, all of which are specific to the mouse or human genome, respectively, are embedded within or immediately downstream of LITs (Fig. 1a, Supplementary Fig. 1a, b and Supplementary Data 2). Compared with all genomic CGIs, this represents a significant enrichment of LITs at igDMRs (mouse: 4/21 igDMRs vs 152/16023 CGIs, Chi-square $p = 1.17 \times 10^{-17}$; human: 17/125 igDMRs vs 70/31144 CGIs, Chi-square $p = 7.42 \times 10^{-219}$). Thus, the presence of a lineage-specific proximal LTR that initiates an oocyte transcript overlapping with a genic gDMR is associated exclusively with genes showing evidence of a species-specific igDMR.

Of the 17 LITs associated with human-specific igDMRs, 12 initiate within primate (Hominoidea or Catarrhini)-specific ERV families (Fig. 1a and Supplementary Fig. 1c). Moreover, the four LITs apparently responsible for transcription-coupled de novo DNAme of the mouse-specific igDMRs, namely at *retro-Coro1c* (also known as *2010001K21Rik* and *AK008011*), *Cdh15*[27], *Slc38a4* (also known as *Ata3*)[33,34] and *Impact*[35], all initiate in rodent-specific ERVs (Fig. 1a and Supplementary Fig. 1c). While several such lineage-specific LTR families (i.e., LTR12C/MER51E or MTC/RMER19B) are actively transcribed in oocytes[19,20], others are generally expressed at low levels (Supplementary Fig. 1d, e), indicating that igDMR-coupled LITs do not necessarily initiate from LTR families that are widely expressed in oocytes. As hypothesized, the igDMRs associated with lineage-specific LITs also show species-specific hypermethylation in oocytes, which is retained (>35% DNAme) in the blastocyst[9], the placenta[26] or at least one somatic tissue surveyed, indicating that these genomic regions are indeed likely to carry imprinted DNAme marks (Fig. 1b and Supplementary Data 2). In accordance with previous reports of preferential maintenance of maternal germline-derived DNAme in human placenta[22,23,25,26,36], most of the LIT-associated human igDMRs retain DNAme in purified cytotrophoblast (CT; cells isolated from first-trimester human placenta)[26] but are hypomethylated in adult tissues (Fig. 1b). The remaining 5/17 human-specific maternal igDMRs (at *HTR5A, AGBL3, CLDN23, ZC3H12C*, and *SVOPL*) are associated with LITs driven by LTR families that colonized the common ancestor of the Euarchontoglires (rodents and primates, Fig. 1b). These specific LTR insertions, however, are not detected at the syntenic loci of the mouse genome and only 1/5 of the regions is methylated in mouse oocytes (the region syntenic to the human igDMR at *HTR5A*), which may be due to a distinct non-LTR-initiated transcript (Fig. 1b, Supplementary Fig. 1b and Supplementary Data 2).

This species-specific pattern of transcription initiation from an upstream LTR element leading to transcription-coupled establishment of a DNAme domain that includes the downstream igDMR/CGI can be visualized in a genome-browser view of orthologous regions (Fig. 1c, d). Note that we observe both sense and antisense configurations of the relevant LITs, driven by LTRs located either upstream or downstream (within an intron or 3′) of the regulated gene, respectively (Supplementary Fig. 2a). Of the 17 LITs putatively implicated in the induction of human igDMRs, 12 are in the 5′ sense configuration and five in an antisense

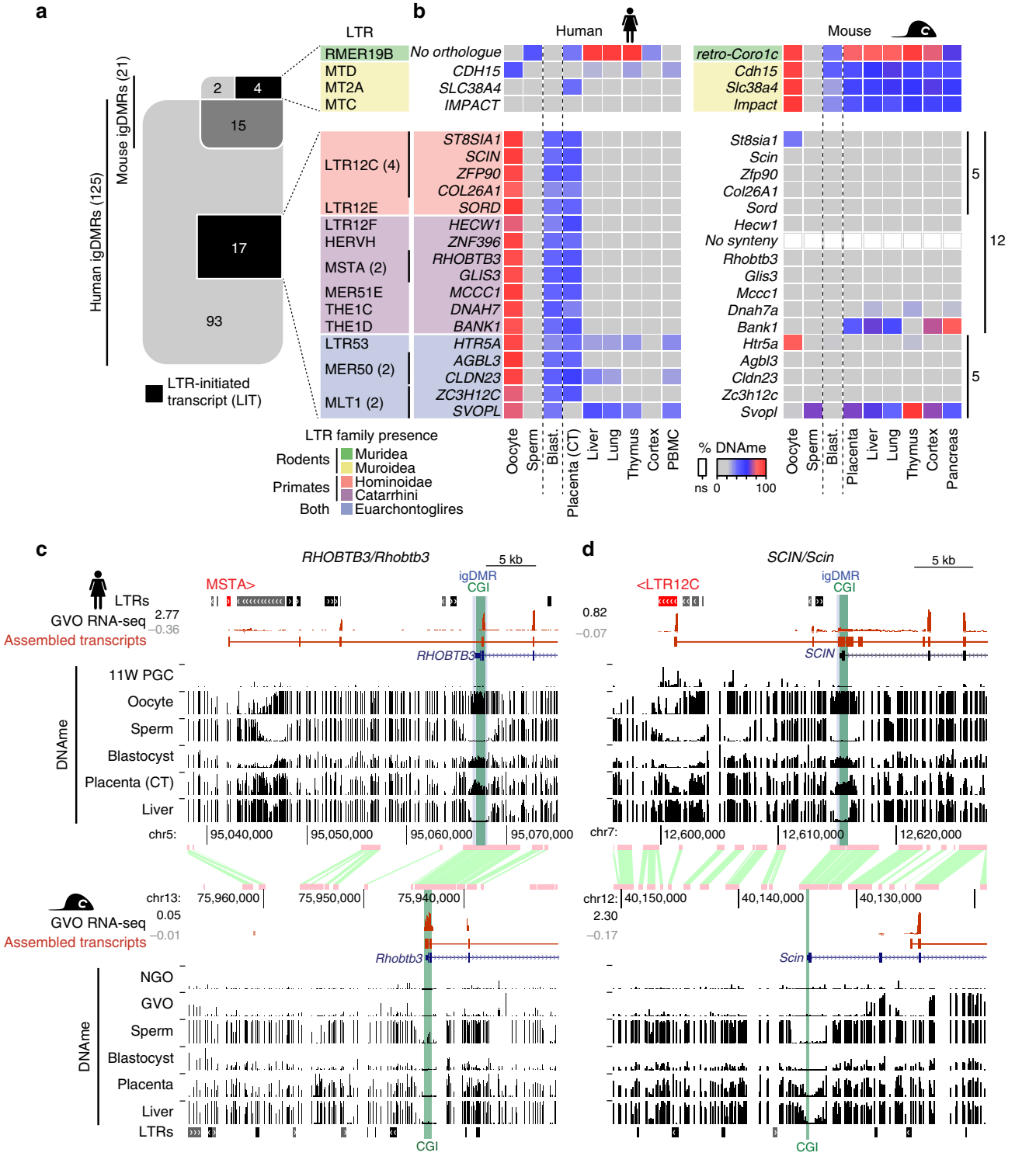


Fig. 1 Identification of human and mouse maternal igDMRs embedded within lineage-specific LITs. a Venn diagram showing the intersection of known maternal igDMRs in mouse and human, along with the subset of igDMRs in each species embedded within a LIT. For each LIT-associated igDMR, the family of the LTR in which transcription initiates in oocytes is shown on the right. The presence of each LTR family in relevant mammalian lineages is color-coded as in Supplementary Fig. 1c. **b** List of imprinted genes/igDMRs associated with LITs. Maternal igDMRs unique to mouse (4) or human (17) are shown, along with DNAme levels (heat map) for each igDMR in syntenic regions in the gametes, blastocyst, placenta and adult tissues in human and mouse. The retrogene *retro-Coro1c* is absent in the syntenic human region on chromosome 6p22.3 (*No orthologue*). *ZNF396* does not have a syntenic CGI in mice (*No synteny*). CT: cytotrophoblast; PBMC: peripheral blood mononuclear cell. **c, d** Screenshots of the human and mouse *RHOBTB3/Rhobtb3* and *SCIN/Scin* loci, including locations of annotated genes, LTR retrotransposons, and regions of syntenic homology. The relevant CGI, igDMR, and upstream LTR in human are highlighted in green, blue, and red respectively. For each species, RNA-seq data from GVOs are shown, along with assembled transcripts, including LITs and their 5′ LTR exons (red) for the human genes. DNAme levels in gametes, blastocyst, placenta, and liver are shown across each locus in both species. For the human DNAme data, profiles from female 11-week primordial germs cells are also shown (11W PGC) and oocyte DNAme is from a mixture of GVO and MII oocytes. Details of all the datasets used in this study are presented in Supplementary Data 1.

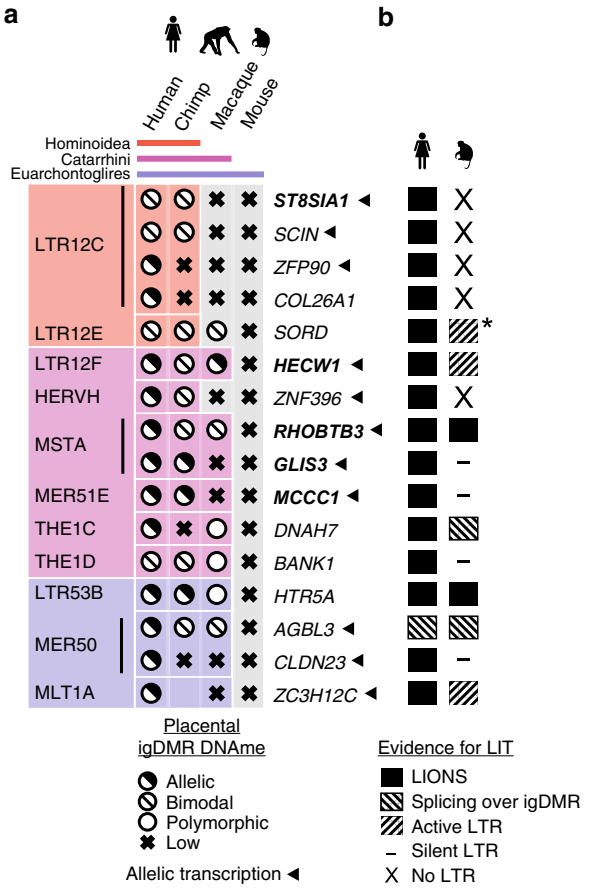


Fig. 2 Conservation of oocyte LTR-initiated transcription and gametic imprinting in primates. a Table of 16 human genes with igDMRs embedded within LITs active in oocytes and showing maternal/allelic DNAme in blastocyst and cytotrophoblast. The family of the initiating LTR is shown on the left, color-coded according to the phylogenetic distribution of the ERV family (top), as in Fig. 1a. For each species, the presence of the LTR insertion at each locus is indicated by a matching colored box and the igDMR DNAme status in human or the syntenic region in chimp, macaque and mouse placenta is shown. An empty box indicates no data. Arrows indicate genes for which evidence of allelic transcription has been published (Supplementary Data 2). Gene names in bold were analyzed in greater detail. **b** Conservation of LITs in human and macaque oocytes for the 16 igDMRs from panel **a**. Solid boxes indicate LITs discovered by LIONS, boxed hatches indicate LITs with evidence of splicing from the LTR over the igDMR, and unboxed hatches indicate evidence of transcription from the LTR. Dashes indicate LTRs from which no transcription is seen, and loci for which the relevant LTR is absent from the macaque genome are also shown (X: No LTR). Asterisk denotes an LTR12F that may initiate a LIT in macaque oocytes (Supplementary Fig. 4b).

configuration (Supplementary Table 1). At the *RHOBTB3/Rhobtb3* locus for example, transcription in human oocytes initiates within an unmethylated primate-specific MSTA element located ~25 kb upstream of the promoter CGI/igDMR, forming a chimeric transcript that splices to the downstream genic exons of *RHOBTB3* (Fig. 1c). Coincident with this LIT, a large block of DNAme is deposited in oocytes over the promoter CGI and overlapping igDMR. Importantly, these regions are hypomethylated in human female 11-week gonadal PGCs[8] and in sperm. As previously documented for many human igDMRs[22,23,36], this imprint is maintained in the blastocyst and cytotrophoblast[26], but is hypomethylated (<2% DNAme) in adult tissues (Fig. 1b-c and Supplementary Data 2). Notably, *RHOBTB3* is expressed predominantly from the paternal allele in human placenta[24,26,37,38]. In contrast, in mouse oocytes *Rhobtb3* transcription initiates at the promoter CGI, which is unmethylated in oocytes, placenta and adult tissues (Fig. 1b-c). Similarly, at the *SCIN* locus, a LIT initiates in an unmethylated LTR12C element ~14 kb upstream of the igDMR in human oocytes and extends into the gene, concomitant with de novo DNAme of this region between the PGC and mature oocyte stages (Fig. 1d). While the *SCIN* igDMR shows ~50% DNAme in human blastocyst and placenta (CT), the syntenic region in mice, including the *Scin* CGI promoter, is hypomethylated in each of these cell types and no upstream initiating transcript is observed in mouse oocytes (Fig. 1b, d). Consistent with the DNAme status of the locus in each species, *SCIN/Scin* is expressed exclusively from the paternal allele in human but not in mouse placenta[24,39]. Importantly, unlike in oocytes, eight of the nine genes that are associated with igDMRs and expressed in purified human cytotrophoblast (CT) show no evidence of transcription initiating from the proximal LTR in this cell type. Rather, for all six genes (*GLIS3*, *MCCC1*, *RHOBTB3*, *ZFP90*, *CLDN23*, *ZC3H12C*) that show clear paternally-biased expression (>70%) in CT (Supplementary Table 1), transcription initiates predominantly within the unmethylated igDMR promoter (Supplementary Fig. 2b, 3). Thus, imprinted expression of these genes in the placenta is correlated with LIT-associated deposition of DNAme over the igDMR in the oocyte and concomitant silencing of the maternal allele in the extraembryonic trophoblast lineage in the offspring.

Primate LITs and de novo DNAme at maternal igDMRs. Excluding the *SVOPL* igDMR, which is unmethylated in cytotrophoblast, we focused on the 16 human loci with evidence for maintenance of the maternal igDMRs in this placental lineage (Fig. 1b). Intriguingly, five of these, including *SCIN*, initiate in human ERV (HERV) families which colonized the common ancestor of the Hominoidea[40] (LTR12C or LTR12E), while seven of them initiate within families that colonized the primate lineage, including the common ancestors of the Catarrhini. The remaining four initiate in more ancient elements derived from LTR families common to both primates and rodents (Euarchontoglires) (Fig. 2a and Supplementary Fig. 1c). To characterize the relationship between LITs and the imprinting status of these gDMRs in non-human primates, we first determined whether the specific LTR insertions associated with these human placental igDMRs are annotated in the genomes of chimpanzees (*Pan troglodytes*; Hominoidea lineage) and rhesus macaques (*Macaca mulatta*; Catarrhini lineage). All 16 relevant LTR insertions are present in the chimpanzee genome. Moreover, all four insertions from LTR families that colonized the Euarchontoglires common ancestor and 6/7 LTR insertions from families that colonized the common ancestor of the Catarrhini are also present in the orthologous loci in macaque. Similarly, the Hominoidea-specific LTR12C and LTR12E families, which include LTRs driving transcripts overlapping the *ST8SIA1*, *SCIN*, *ZFP90*, *COL26A1*, and *SORD* igDMRs in human oocytes, are absent from the macaque genome (Fig. 2a and Supplementary Fig. 1c).

To assess the conservation of LITs and DNAme of associated gDMRs in macaque oocytes, we analyzed published RNA-sequencing[41] and genome-wide DNA methylome data[42]. As expected from the phylogeny of LTR12C elements, no LITs overlapping the region orthologous to the human igDMRs at the *ST8SIA1*, *SCIN*, or *COL26A1* loci were detected in macaque oocytes, and their CGIs remain hypomethylated (Figs. 2a, b, 3a and Supplementary Fig. 4a). Intriguingly, despite the absence of an LTR12E element, the region syntenic to the human *SORD* igDMR exhibits high levels of DNAme in macaque oocytes.

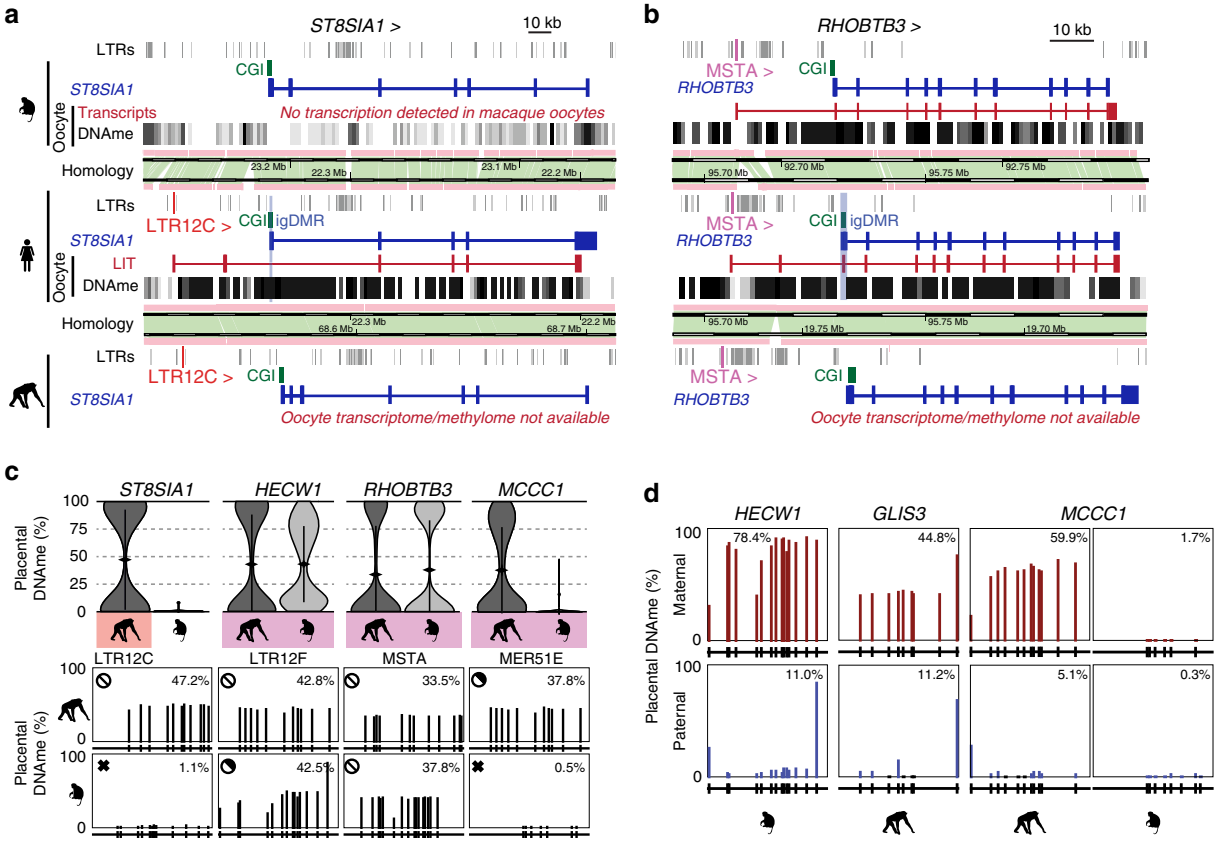


Fig. 3 Chimp and macaque LITs and placental DNAme. a Screenshot of the human *ST8SIA1* locus, showing its CGI promoter (green), annotated gene (blue), oocyte LIT (red), and DNAme in human and macaque oocytes (black). Highlighted are the location of the upstream LTR12C in human and chimp (red) and the human igDMR (blue). **b** Screenshot of the human *RHOBTB3* locus annotated as in **a** and highlighting the upstream MSTA LTR promoter (pink). **c** Violin plots of the distribution of mean DNAme levels per strand in placenta (chimp *MCCC1*: $n = 2$ all others: $n = 3$ biologically independent samples) for individual bisulphite-sequencing reads covering the igDMRs of the orthologous chimp and macaque *ST8SIA1*, *HECW1*, *RHOBTB3*, and *MCCC1* genes. Coloured boxes indicate the presence of the proximal LTR. For each gene, the mean DNAme level at each of the CpGs surveyed is shown below. Symbols for DNAme are as in Fig. 2a. Note that the MER51E in the macaque *MCCC1* locus is transcriptionally inert, likely due to macaque-specific SNPs rendering it transcriptionally inactive (Supplementary Fig. 4d–f). Source data are provided as a Source Data file. **d** Placental DNAme data at the *HECW1*, *GLIS3*, and *MCCC1* loci for informative samples heterozygous at SNPs of known parental origin, are shown for the species indicated.

Inspection of oocyte transcripts around the macaque *SORD* CGI reveals the presence of a highly transcribed LTR12F element oriented towards the putative igDMR, which may be responsible for deposition of DNAme at this locus (Supplementary Fig. 4a, b).

In contrast, LITs initiating in older LTR families that colonized the common ancestor of the Catarrhini or Euarchontoglire lineages show greater conservation between human and macaque oocytes. Indeed, 8/13 of such LTR insertions are also transcriptionally active in macaque oocytes, a subset of which display similar splicing events in both species (Fig. 2b). Remarkably, at least six of the associated CGIs are also hypermethylated in macaque oocytes (Supplementary Fig. 4a). For example, de novo methylation of the *RHOBTB3* CGI promoter/gDMR in both human and macaque oocytes appears to be the result of a conserved LIT initiating in a MSTA situated >20 kb upstream of the promoter, which clearly splices into the gene in both species (Fig. 3b). Other examples of LITs apparently conserved between the human and macaque lineages include those associated with the *HECW1* (LTR12F), *DNAH7* (THE1C), *HTR5A* (LTR53B), *AGBL3* (MER50), and *ZC3H12C* (MLT1A) putative igDMRs. In contrast, no LITs were detected at 4/11 loci despite the presence of a conserved LTR insertion in the orthologous macaque locus (Supplementary Figs. 4 and 5). For example, while the igDMR at the 5′ end of *GLIS3* is embedded within a LIT initiating in an active upstream MSTA in human oocytes, no LIT is detected in the orthologous region in the macaque locus nor is any RNA-seq coverage detected over the MSTA itself (Supplementary Fig. 5a, b).

The lineage-specific expression of orthologous LTRs may be explained by the accumulation over evolutionary time of indels and/or base substitutions that impact their promoter and/or splice donor activity. Indeed, sequence alignment of the orthologous MSTA insertions reveals a number of small deletions and single nucleotide polymorphisms (SNPs), which may impact transcription of the macaque LTR (Supplementary Fig. 5c). Similarly, despite the presence of a MER51E upstream of the *MCCC1* gene, no transcript initiating in this LTR is detected in macaque oocytes and no RNA-seq coverage is detected over the element itself (Supplementary Fig. 5d, e). Closer inspection of the MER51E insertions in the macaque locus reveals a number of SNPs and short INDELs relative to the orthologous MER51E in chimp and human, including mutations that likely disrupt an otherwise conserved PBX3 binding site that may render the LTR in the former inactive (Supplementary Fig. 5f). Taken together, these data indicate that the establishment of DNAme in oocytes at the igDMRs of a number of Hominoid- and primate-specific maternal igDMRs is likely induced by LITs originating in proximal lineage-specific LTR elements actively transcribed during oogenesis.

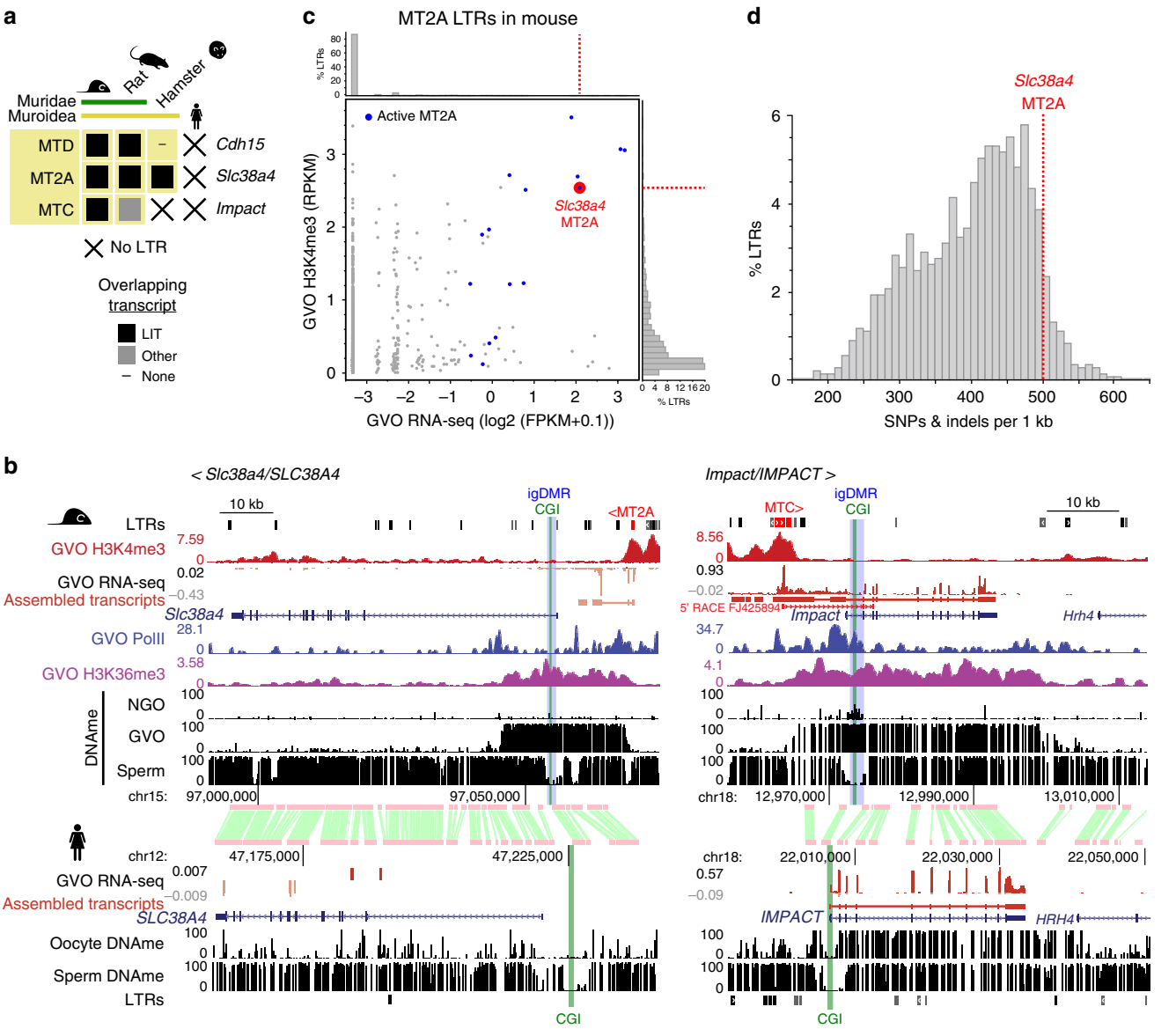


Fig. 4 Phylogenetic relationship between LTR, LIT, and maternal DNAme at mouse-specific igDMRs. a The presence or absence of a LIT overlapping the CGI at *Cdh15*, *Slc38a4*, and *Impact* are indicated with closed boxes and dashes, respectively. X: species in which the LTR is absent. **b** Screenshots of *Slc38a4/SLC38A4* and *Impact/IMPACT* loci showing GVO RNA-seq data (and the associated de novo Cufflinks transcript assembly), DNAme profiles in NGO, GVO and sperm, and H3K4me3, PolII, and H3K36me3 ChIP-seq tracks for mouse GVO. Highlighted are the upstream initiating LTR (red), CGI promoter (green), and mouse igDMR (blue). The syntenic region in human is also shown, including GVO RNA-seq and DNAme from oocytes and sperm. A 5′ RACE gene model initiating within the MTC element at the *Impact* locus is included in the right panel[14]. **c** Scatter plot of oocyte H3K4me3 and transcription levels for all mouse MT2A elements, with those acting as transcription start sites in oocytes highlighted in blue. The remaining transcribed elements reflect exonization events. The MT2A element initiating the LIT at *Slc38a4* (red) is amongst the most active elements of this family. **d** Histogram of the distribution of mouse MT2A LTRs as a function of divergence from the consensus sequence. The LTR driving expression at *Slc38a4* is amongst the most highly diverged.

Conservation of LITs and igDMRs in apes versus monkeys. To establish the imprinting status of these gDMRs in chimp and macaque placenta, we determined their methylation using a targeted high-throughput sodium bisulfite sequencing approach. We focused initially on the *ST8SIA1* gene, shown previously to be maternally methylated and paternally expressed in human placenta[24]. In chimpanzee, the genomic region syntenic to the human *ST8SIA1* igDMR is hypomethylated in adult tissues (Supplementary Fig. 4a), but shows a bimodal distribution of hypermethylated and hypomethylated sequenced reads in the placenta (Fig. 3c and Supplementary Fig. 6a), indicative of conserved placental-specific *ST8SIA1* imprinting within the Hominoidea. In contrast, the syntenic CGI in macaque, as in mouse, is hypomethylated in the placenta (Figs. 2a and 3a, c). Therefore, the presence of an active LTR12C element ~40 kb upstream of the gDMR and its associated LIT are correlated with the imprinting status of the *ST8SIA1* promoter. *SCIN*, which is imprinted and paternally expressed in human placenta[24], also shows a bimodal distribution of DNAme in the orthologous region in chimp placenta. In contrast, in the absence of an LTR12C insertion or a LIT, the orthologous region in macaque placenta is hypomethylated (Supplementary Fig. 6a). On the other hand, despite the presence of a proximal LTR12C element in the chimp (as in the human genome), the regions syntenic to the human igDMRs at the *ZFP90* and *COL26A1* genes are

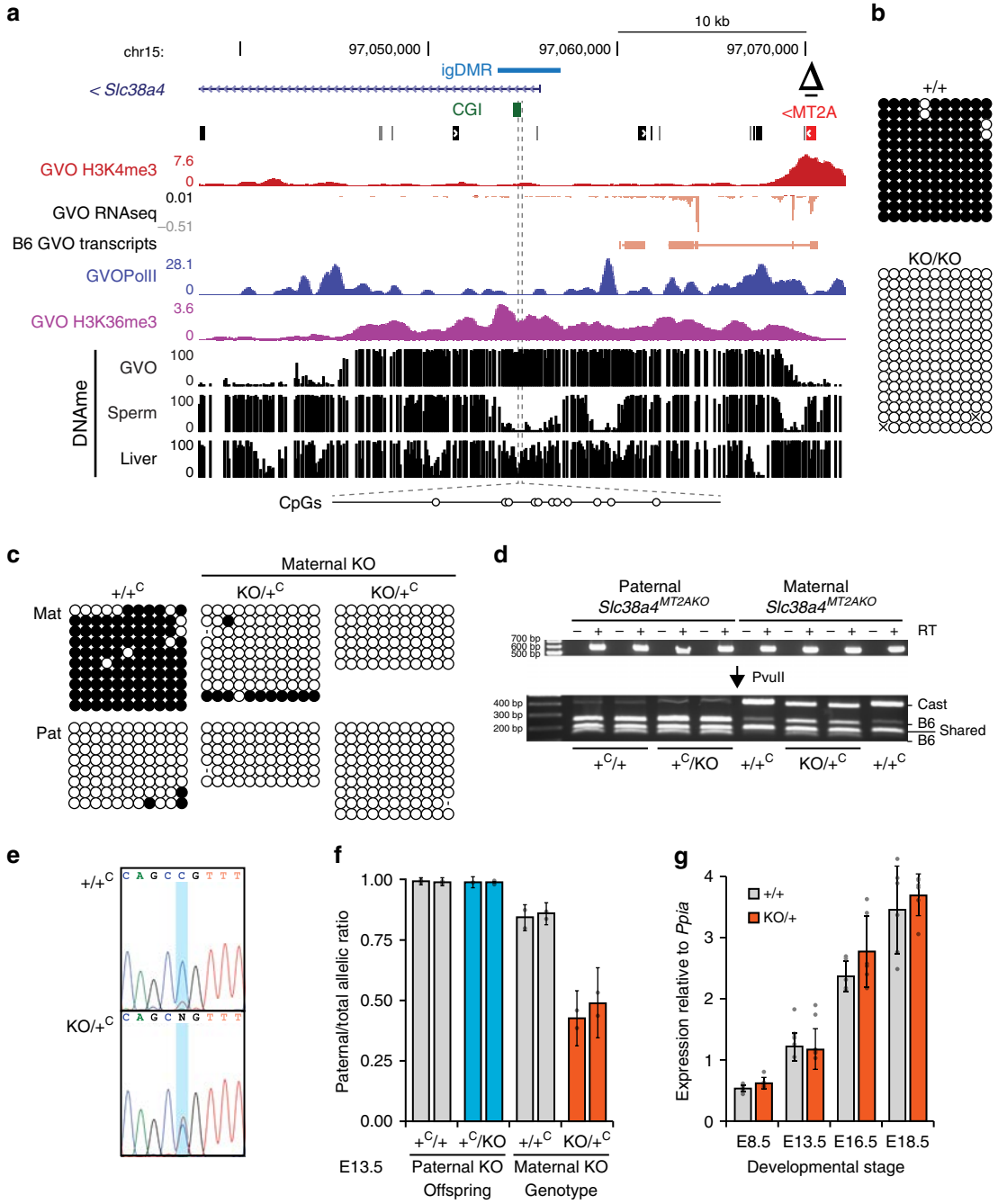


Fig. 5 Loss of imprinting at *Slc38a4* upon maternal transmission of the MT2A KO allele. a Genome-browser screenshot of the mouse *Slc38a4* promoter and upstream region, including the MT2A LTR (red), annotated *Slc38a4* exon 1, CGI (green), and igDMR (blue). GVO RNA-seq as well as RNA pol II, H3K4me3 and H3K36me3 ChIP-seq tracks are shown, along with DNAme data for GVO, sperm and adult liver. The region within the igDMR analyzed by sodium bisulfite sequencing (SBS), which includes 11 CpG sites, is shown at the bottom. Δ: extent of the MT2AKO deletion allele. **b** DNAme of the *Slc38a4* igDMR in GVO from wild-type and *Slc38a4*$^{MT2AKO/MT2AKO}$ females determined by SBS. **c** DNAme of the *Slc38a4* igDMR in E13.5 (*Slc38a4*$^{+/MT2AKO}$ × CAST)F1 embryos determined by SBS. Data for control (+/+^{C}) and heterozygous (KO/+^{C}) littermates with a maternally inherited MT2AKO are shown. +^{C}: wild-type CAST allele; KO: *Slc38a4*^{MT2AKO}. A polymorphic insertion in the amplified region allows for discrimination of maternal (Mat) and paternal (Pat) strands. Allele-specific expression analyses of F1 E13.5 placental RNA by **d** RT-PCR followed by PvuII RFLP analysis (RT reverse transcriptase), and **e, f** Sanger sequencing of a T ⟷ C transition in the 3′UTR of the *Slc38a4* cDNA (maternal B6: T allele; and paternal CAST: C allele). Each bar in **f** shows the mean of individual samples, and error bars show S.D. of two SNPs analyzed. Source data are provided as a Source Data file. **g** Total *Slc38a4* mRNA levels in E8.5, E13.5, E16.5, and E18.5 placentae, as determined by RT-qPCR ($n = 6$ biologically independent samples for each datapoint). Expression levels are relative to the housekeeping gene *Ppia*. Graph shows mean ± S.D. Source data are provided as a Source Data file.

unmethylated in chimp placenta (Fig. 2a), suggesting that, similar to the LTR insertions proximal to the *GLIS3* and *MCCC1* CGI in macaques (Supplementary Fig. 5), these orthologous LTR may be transcriptionally inert in chimp oocytes. This remains to be determined, however, as RNA-seq data from chimp oocytes is currently not available. The regions syntenic to the human *SORD* igDMR show bimodal DNAme in chimp as well as macaque, consistent with the presence of proximal active LTRs in human and

macaque oocytes. While the methylation of the macaque *SORD* locus is likely explained by the alternative LTR12F-initiated transcript described above (Supplementary Fig. 4b), it remains to be determined whether an antisense LTR12E-initiated LIT orthologous to that observed in human oocytes is responsible for depositing DNAme over the promoter CGI of this locus in chimpanzees.

As predicted by the presence of an orthologous LTR12F insertion oriented towards the *HECW1* gene in both the chimp and macaque loci, the genomic region of shared synteny to the human *HECW1* igDMR shows a clear bimodal distribution of hypermethylated and hypomethylated reads in the placentae of both species, indicative of conservation of LIT-associated *HECW1* imprinting in Catarrhines (Fig. 3c and Supplementary Fig. 4c). Moreover, exploiting a single SNP within the *HECW1* CGI, we observed clear allelic methylation (Fig. 3d and Supplementary Fig. 6b), confirming maternal allele-specific DNAme at the orthologous igDMR. Similarly, analysis of the chimp and macaque genomes in the regions syntenic to the *RHOBTB3* igDMR, which is embedded within an MSTA-initiated transcript in both human and macaque oocytes (Fig. 3b), reveals that the orthologous locus is also likely imprinted in these species (Fig. 3c). Furthermore, analysis of DNAme data from informative trios also reveals maternal-specific DNAme of the igDMRs at the *GLIS3* and *MCCC1* genes in the chimp placenta (Fig. 3c-d and Supplementary Fig. 6b). In agreement with the cognate LTR insertions being transcriptionally inactive in macaque oocytes (Supplementary Fig. 5), the genomic regions syntenic to the human and chimp igDMRs at both of these loci are hypomethylated in macaque placentae. Taken together, these observations reveal that transcription initiating within lineage-specific LTR elements in the oocytes of primates likely plays a critical role in the establishment of DNAme at proximal CGIs. The persistence of this oocyte-derived imprint in the placenta of their progeny yields maternal igDMRs that can potentially direct imprinted expression in this extraembryonic tissue, as shown for several human genes.

Muroidea-specific LTRs and maternal igDMRs. If the LTRs upstream of the mouse-specific igDMRs shown in Fig. 1 are indeed responsible for the establishment of maternal imprinting at these loci, then the orthologous genes should also be imprinted in those species harboring the same active LTR insertions. As the *retro-Coro1c* retrogene is absent from rats and more distantly related rodents (Supplementary Fig. 7a), we focused on the remaining three genes, namely *Cdh15, Slc38a4*, and *Impact* and their associated LTRs: MTD, MT2A, and MTC, respectively. All three LTR insertions are present in the orthologous regions of the rat (*Rattus norvegicus*) genome (Fig. 4a), consistent with the fact that these LTR families colonized the common ancestor of the Muroidea (Supplementary Fig. 1c, 7)[19]. In contrast, while the relevant MTD and MT2A elements are also present at the orthologous *Cdh15* and *Slc38a4* loci in the golden hamster (*Mesocricetus auratus)*, the MTC element upstream of *Impact* is absent.

To assess whether the LITs identified at these imprinted loci in mice are also present in rat and golden hamster, we mined published oocyte RNA-seq data[19,20]. LITs emanating from the relevant LTRs were clearly detected at *Cdh15* and *Slc38a4* genes in rat oocytes, while in hamster, a LIT was detected only at the *Slc38a4* locus (Fig. 4a and Supplementary Fig. 7b–d). In rat, as in mouse[20], the *Slc38a4* LIT overlaps the DNAme block that extends into the 5′ end of the gene. In hamster, the MT2A-initiated LIT upstream of *Slc38a4* splices into exon 2 of the gene, creating a chimaeric transcript covering the entire gene, including the region syntenic to the mouse igDMR (Supplementary Fig. 7c). In contrast, in human oocytes, the *SLC38A4* locus is not transcribed and the 5′ CGI is unmethylated (Fig. 4b). In the case of *Impact*, a LIT initiating in a co-oriented upstream MTC element encompasses the entire annotated gene in mouse and demarcates a hypermethylated domain, including the igDMR (Fig. 4b). Intriguingly, while the syntenic *Impact* CGI is hypermethylated in rat oocytes, the orthologous rat MTC insertion is apparently oriented on the opposite strand relative to the *Impact* gene (Supplementary Fig. 7d)[43]. Closer analysis of the locus reveals that a transcript encompassing the *Impact* CGI may originate from a distinct (non-LTR) upstream start site in rat oocytes, but a gap in the reference genomic sequence precludes detailed analysis of the upstream transcript at this locus (Supplementary Fig. 7d). Regardless, as in the mouse, *Impact* is expressed from the paternal allele in rat brain[43], consistent with epigenetic imprinting of its CGI and imprinted expression. Given the probable conservation of imprinting at *Slc38a4* and *Impact* in the Muroidea lineage, and the location of the intragenic igDMRs near their 5′ ends, we focused on these genes for further functional experiments in the mouse.

An MT2A is required for *Slc38a4* imprinting. In mouse oocytes, transcription over the *Slc38a4* igDMR originates in an H3K4me3 marked MT2A LTR located ~14 kb upstream of the gene, in a sense configuration. The 5′ end of the locus is embedded within the LIT, enriched for H3K36me3, and overlaps with a de novo DNAme block that encompasses the igDMR (Fig. 4b). In addition, our ChIP-seq analysis of RNA pol II in mouse GVOs reveals enrichment consistent with transcription extending from the MT2A through intron 1 of *Slc38a4* (Fig. 4b). MT2A elements are not generally transcribed at high levels in mouse oocytes, and analysis of all insertions >450 bp (1458 in total) indicates that only 16 initiate transcripts in GVOs at detectable levels (Fig. 4c and Supplementary Fig. 1e). As the MT2A upstream of *Slc38a4* has significantly diverged from the consensus sequence of this relatively old LTR family (Fig. 4d) and this LTR is not closely related to the few other MT2A elements active in oocytes (Supplementary Fig. 8), its transcriptional activity in the female germline may be due to the acquisition of novel transcription factor binding sites.

To test directly the role of the upstream MT2A insertion in *Slc38a4* imprinting, we generated a 638-bp deletion of the MT2A LTR in C57BL/6 N (B6) embryonic stem cells using CRISPR-Cas9 and flanking guide RNAs (Fig. 5a, Δ; Supplementary Fig. 9a). Germline transmission from male chimeras allowed us to establish the $Slc38a4^{MT2AKO}$ mutant mouse line on a pure B6 background. Heterozygous and homozygous animals were normal and fertile under standard husbandry conditions. In reciprocal crosses involving a wild-type and a heterozygous parent, heterozygotes were recovered at the expected frequency (Supplementary Table 2). We first assessed the effect of the MT2A deletion on DNAme at the *Slc38a4* igDMR in fully-grown oocytes using sodium bisulfite sequencing of 11 CpG sites within the igDMR located at the beginning of intron 1 (Fig. 5a). Whereas these sites are hypermethylated in wild-type oocytes, they remain unmethylated in oocytes from females homozygous for the MT2A deletion (Fig. 5b). To test for allele-specific transcription, we generated F1 progeny from reciprocal crosses with CAST/EiJ (CAST) mice, a hybrid strain background in which *Slc38a4* was previously shown to be imprinted[34]. Sodium bisulfite analysis confirmed that DNAme at the igDMR is absent in F1 embryos (E13.5) upon maternal transmission of the $Slc38a4^{MT2AKO}$ allele (Fig. 5c). Consistent with the absence of maternal DNAme, *Slc38a4* is expressed from both alleles in E13.5 placenta when the

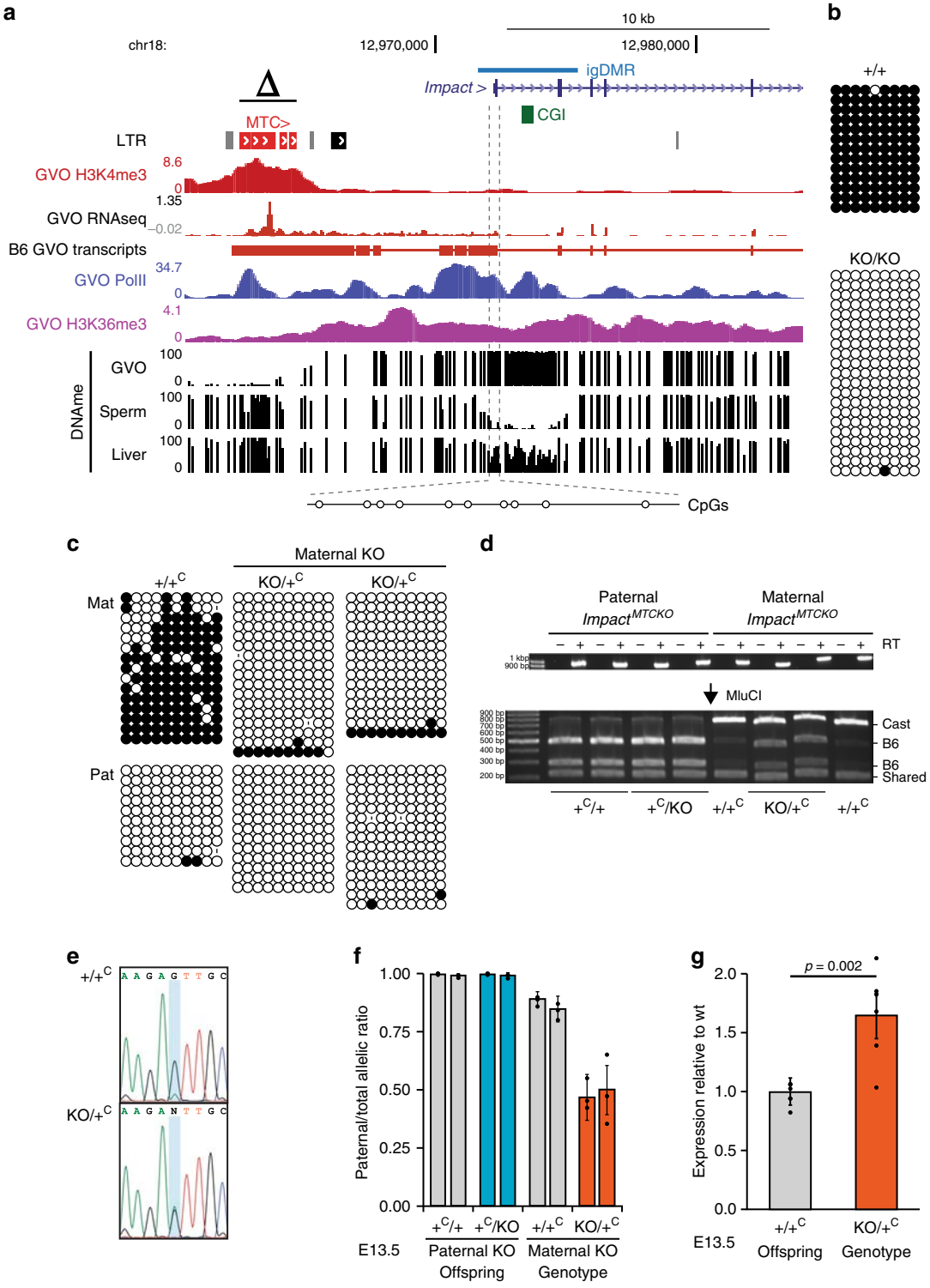


Fig. 6 Loss of imprinting at *Impact* upon maternal transmission of the MTC KO allele. a Genome-browser screenshot of the mouse *Impact* locus, including the upstream MTC LTR (red), CGI (green), and igDMR (blue). GVO RNA-seq as well as RNA pol II, H3K4me3, and H3K36me3 ChIP-seq tracks are shown, along with DNAme data for GVO, sperm, and adult liver. The region within the igDMR analyzed by sodium bisulfite sequencing (SBS), which includes 10 CpG sites, is shown at the bottom. Δ: extent of the upstream MTCKO deletion allele. **b** DNAme of the *Impact* igDMR in GVO from wild-type and $Impact^{MTCKO/MTCKO}$ females determined by SBS. **c** DNAme of the *Impact* igDMR in E13.5 ($Impact^{+/MTCKO}$ × CAST)F1 embryos determined by SBS. Data for control ($+/+^{C}$) and heterozygous ($KO/+^{C}$) littermates with a maternally inherited MTCKO allele are shown. $+^{C}$: wild-type CAST allele; KO: $Impact^{MTCKO}$. A polymorphic insertion and a SNP in the amplified region allow for discrimination of maternal (Mat) and paternal (Pat) strands. **d, e** Allele-specific expression analysis of F1 E13.5 embryonic head RNA by **d** RT-PCR followed by MluCI RFLP analysis, and **e** Sanger sequencing of a A ⟷ G transition in the 3′UTR of the *Impact* mRNA (maternal B6: A allele; and paternal CAST: G allele). RT: reverse transcriptase. Source data are provided as a Source Data file. **f** Quantification of relative levels of expression from the paternal *Impact* allele based on the analysis of embryos as in **e**. Graph shows mean ± S.D of three SNPs. Source data are provided as a Source Data file. **g** *Impact* mRNA levels analyzed by RT-qPCR on E13.5 embryonic RNA ($n = 6$ biologically independent samples). Expression levels are relative to those for the wild-type allele. Graph shows mean ± S.E.M. Source data are provided as a Source Data file.

deletion is maternally inherited, while monoallelic imprinted expression from the paternal allele is maintained in paternal heterozygotes (Fig. 5d-f). *Slc38a4* codes for an amino acid transporter (ATA3) highly expressed in the mouse placenta[34], suggesting that variations in its overall expression levels might lead to placental and/or growth abnormalities, as described for other imprinted genes. However, we did not observe either phenotype in maternal heterozygotes (Supplementary Fig. 9c–f). Notably, *Slc38a4* mRNA levels are high from E8.5 to E18.5 in wild-type placenta (Fig. 5g). However, consistent with the absence of placental or embryonic growth abnormalities, we found that loss of imprinting and biallelic expression at *Slc38a4* are not accompanied by an increase in total mRNA levels in maternal heterozygotes at those stages (Fig. 5g). These results, in accord with previously published findings[44], suggest that *Slc38a4* levels are normalized by transcriptional or post-transcriptional mechanisms in the placenta. Nevertheless, the possibility that a decrease in *Slc38a4* mRNA would bring its levels within the dosage-sensitive zone in which imprinting of the locus might provide a selective advantage is supported by the abnormal placental and embryonic growth phenotypes recently observed in *Slc38a4*-null conceptuses[45].

An upstream MTC directs imprinting at *Impact*. In mouse oocytes, the *Impact* gene is transcribed from an unmethylated upstream MTC element in the sense configuration. This LIT splices onto canonical exon 2 of *Impact*, as supported by RNA-seq data[3,20] and a 5′ RACE product from GVOs[15] (Fig. 4b). As seen at the *Slc38a4* locus, the MTC itself is marked by H3K4me3 in oocytes[46]. RNA pol II, H3K36me3, and DNAme are enriched over the transcribed region, which encompasses the entire *Impact* locus, including its igDMR and intronic CGI. Using a CRISPR-Cas9 mutagenesis approach similar to the one described for the MT2A at *Slc38a4*, we generated a mouse line in which a ~3 kb region upstream of the *Impact* gene, including the full-length MTC element, was deleted to generate the $Impact^{MTCKO}$ allele (Fig. 6a, **Δ**; Supplementary Fig. 9b). Following germline transmission through male chimeras, heterozygous males and females were bred to wild-type B6 mice and heterozygotes were recovered at the expected Mendelian ratios (Supplementary Table 2). Maternal and paternal heterozygotes, as well as homozygous mice of both sexes, appear normal and are fertile. As we observed for the MT2A deletion at *Slc38a4*, the igDMR at *Impact* fails to acquire de novo DNAme in mature oocytes from $Impact^{MTCKO/MTCKO}$ homozygous females (Fig. 6b).

To study the effect of the MTC deletion on imprinting of the downstream *Impact* gene, reciprocal crosses were performed between $Impact^{+/MTCKO}$ and CAST mice, and embryos were collected at E13.5. DNAme analysis of the *Impact* igDMR confirmed that DNAme is absent in embryos carrying a maternally inherited KO allele ($Impact^{MTCKO/+}$, Fig. 6c). Furthermore, allele-specific RT-PCR of head cDNA generated from the same E13.5 F1 progeny revealed that *Impact* is paternally expressed in all samples analyzed, with the exception of embryos in which the MTCKO allele is maternally inherited, where *Impact* is expressed from both parental alleles (Fig. 6d-f). Consistent with biallelic expression in these $Impact^{MTCKO/C+}$ embryos, quantitative analysis of total expression levels reveals that *Impact* mRNA levels are nearly doubled relative to those measured in wild-type controls (Fig. 6g). To determine whether this loss of imprinting at *Impact* influences the growth of mutant mice, we measured the weights of males and female wild-type and $Impact^{MTCKO/+}$progeny. No significant weight difference was detected through postnatal week 60 (Supplementary Fig. 9g), indicating that under standard husbandry conditions, increased *Impact* levels do not affect postnatal growth. Regardless, our analyses clearly reveal that DNAme and expression imprinting of the *Impact* gene in mice is dependent upon the presence of the upstream MTC element.

Discussion

Genomic imprinting is an epigenetic mechanism required for normal development and postnatal survival in mammals. Although more than a hundred imprinted genes have been identified and much has been learned about the epigenetic mechanisms regulating their monoallelic expression, it is still unknown how an ancestral gene, expressed from both alleles, acquires an imprinted expression pattern during evolution. Previous work has shown that through retrotransposition new retrogenes can acquire a maternal igDMR and show paternal allele-specific expression when inserted within a host gene expressed during oogenesis[47]. Although a number of such examples have been described in mammals[48], this phenomenon involves the creation of a new gene. Thus, it does not provide a model for the evolutionary switch from biallelic to imprinted expression manifest at species-specific imprinted genes, as explored in this study.

Here, we uncovered a novel mechanism whereby lineage-specific insertions of LTR retrotransposons, transcriptionally active during oocyte growth, can drive de novo DNAme and imprinting at a nearby gene. CRISPR-Cas9 mutagenesis at two mouse-specific imprinted genes, *Impact* and *Slc38a4*, confirmed the key role played by LTR-promoted transcription in guiding imprinting at these two mouse loci. Interestingly, *Slc38a4* was previously reported to be imprinted in mouse preimplantation embryos and early extraembryonic tissues in an H3K27me3-dependent manner[49]. However, the broad intragenic domain of H3K27me3 observed in oocytes and on the maternal *Slc38a4* allele in early embryos[49] is mutually exclusive of the upstream H3K36me3- and DNAme-marked domain studied here (Supplementary Fig. 10). This observation is consistent with a previous report showing that H3K36me3 inhibits PRC2 activity[50]. Thus, as with canonical imprinted genes, the *Slc38a4* igDMR in mice, which overlaps with an annotated TSS of this gene, is enriched for DNAme but devoid of H3K27me3. In addition to our results on the MT2A deletion allele, the importance of the LIT-induced igDMR for allelic usage is also supported by the observation of imprinted expression at *Slc38a4* in embryonic and adult epiblast-derived tissues in which the igDMR, but not the intragenic H3K27me3 domain, is present (Supplementary Fig. 10)[34,51]. The paradoxical critical role of two different maternal epigenetic marks laid down in two different regions of the *Slc38a4* locus may be reconciled by the fact that an alternative annotated TSS of the *Slc38a4* gene is located downstream of the igDMR and is embedded within the intragenic region enriched for H3K27me3 described by Inoue and colleagues[49]. This alternative TSS is used in adult liver, the tissue showing the highest *Slc38a4* expression levels after placenta (Supplementary Fig. 10). Although the igDMR is present in adult liver, *Slc38a4* expression is biallelic in this tissue, suggesting that tissue-specific loss of imprinted expression occurs via promoter switching[34]. Thus, depending on the tissue, imprinted expression of this gene likely requires one or the other epigenetic mark. Further studies are required to address this intriguing possibility and to determine whether establishment or maintenance of the igDMR at *Slc38a4* requires maternal H3K27me3. Notably, although preimplantation maintenance of the maternal DNAme imprint at *Impact* requires the protective action of the

KRAB-zinc finger proteins ZFP57 and ZFP445, the igDMR at *Slc38a4* does not[52]. This observation is consistent with a report showing that DNAme at the *Slc38a4* igDMR in postimplantation embryos is sensitive to loss of the H3K9 methyltransferase EHMT2/G9A[53].

Our observation that 15.5% (17/110) of human-specific and 66.7% (4/6) of mouse-specific maternal igDMRs are potentially induced by ERV promoters highlights the importance of this mechanism and its impact on the evolution of species-specific imprinted genes in mammals. Furthermore, our analysis also revealed evidence for convergent evolution of imprinting, where a distinct LTR may be responsible for imprinting in different species (LTR12E/F at the *SORD* locus in human/macaque for example). Further studies, such as those applied here in the mouse, will be required to determine whether alternative LTRs play a role in the imprinting of such loci in other species. Those species-specific maternal igDMRs not associated with a LIT are likely methylated as a consequence of transcription initiating in novel, single-copy, TSSs active in oocytes (as seen at the rat *Impact* locus).

While disruption of imprinting at some maternally silenced genes is associated with severe phenotypic outcomes, consistent with our observations at *Slc38a4* and *Impact*, loss of imprinting and biallelic expression of imprinted genes is not necessarily associated with obvious abnormal phenotypes. Previous studies of the paternally expressed genes *Plagl1*[3], *Peg3*[17], and *Zrsr1*[16] for example, revealed only subtle impacts on reproductive fitness following loss of imprinting, although detailed analyses of potential growth-related phenotypes have not always been reported.

While the role of active transcription in guiding DNAme of the oocyte genome, including at maternal igDMRs, is well established, the molecular mechanism involved still remains to be elucidated. We recently found that transcription-coupled deposition of H3K36me3 by SETD2 plays a critical role in this process, as DNAme at all maternal igDMRs, including at *Impact* and *Slc38a4*, is lost in mouse oocytes in which *Setd2* is deleted[10]. Thus, transcription-coupled H3K36me3 deposition is likely the critical common feature for the establishment of DNAme at maternal igDMRs in mammalian oocytes, including at igDMRs embedded within lineage-specific active LTR-initiated transcription units (Supplementary Fig. 1b).

A link between genomic imprinting and repetitive elements was explored previously by several groups, who proposed that the function of DNAme as a host defense mechanism might have been co-opted for allelic silencing at imprinted genes[54–57]. However, this hypothesis is inconsistent with our current understanding of de novo DNAme in the female germline, a process that is intimately associated with transcribed regions, which include imprinted gDMRs. Indeed, the mechanism we uncovered explains the establishment of DNAme imprints at single-copy CpG-rich regions that are unremarkable with respect to their repetitive element content. Rather, deposition of DNAme *in cis* is dependent upon the transcriptional activity of nearby ERVs that evade silencing in oocytes. In contrast, de novo DNA methylation at the *Rasgrf1* locus in male germ cells involves a *trans* mechanism, whereby an RMER4B LTR at the 3′-end of the gDMR is targeted for DNA methylation by a yet to be fully characterized nuclear piRNA pathway active in prospermatogonia[58]. Since the mechanism we described is not targeting repeat sequences, once the specific requirements for post-fertilization imprint maintenance are met, active ERVs, which propagate via retrotransposition, can theoretically induce imprinting of any downstream gene, with the prerequisite that the LTR be active in growing oocytes, when DNAme is established de novo by the DNMT3A/3L complex.

Methods

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Validation of mouse and human maternally methylated igDMRs. To validate previously identified maternally methylated imprinted gDMRs (igDMRs)[22–25], we interrogated DNAme profiles from published whole-genome bisulfite sequencing data from gametes, placenta and somatic tissues (Supplementary Data 1)[59]. Specifically, we identified regions that are hypermethylated (>70% DNAme) in oocytes, hypomethylated (<30% DNAme) in sperm, retain DNAme (>25%) in the blastocyst, and show 35–65% DNAme in placenta (purified first-trimester cytotrophoblast; CT) or at least one adult somatic tissue[4,5,8,9,19,20,30,41,42,46,49,51,60–72]. We further validated and refined this list by only including gDMRs harboring either reported maternal, monoallelic or bimodal DNAme in the placenta and/or at least one somatic tissue. A bimodal DNAme pattern is a specific class of sequences methylated at ~50% for which individual DNA sequencing strands from WGBS datasets, although not overlapping a SNP with known parental origin, are either fully- (hyper) methylated or un- (hypo-) methylated. If the DNA sequence analyzed also encompasses a SNP, the bimodal pattern can be described as allelic when all hypo- or all hypermethylated strands contain the same variant at the SNP: only one of the parental alleles is methylated. Only when the parental origin of the sequence variants of this SNP is known in a given sample, can the allelic DMR be referred to as imprinted (maternally or paternally methylated). In total, 125 human igDMRs (46 of which are associated with reported paternal-biased transcription of the nearby gene) and 21 mouse igDMRs (all with reported paternal-biased transcription) are presented and analyzed in Supplementary Data 2. Syntenic regions between the mouse (mm10) and human (hg19), as well as chimpanzee (panTro4), macaque (rheMac8), rat (rn6), and golden hamster (mesAur1) genomes were obtained using the Liftover tool from the UCSC Genome Browser (http://genome.ucsc.edu/)[73].

Identification of LTR transcripts overlapping igDMRs. Oocyte RNA-seq libraries (see Supplementary Data 1) were aligned to the mm10 (mouse), hg19 (human), rhemac8 (macaque), mesAur1 (golden hamster), and rn6 (rat) assemblies using STAR v.2.4.0.i42[74]. De novo transcriptome assemblies were produced using Stringtie v.1.3.5 (mouse, human, golden hamster and rat)[32] or Cufflinks v.2.1.121 (macaque)[29] with default parameters. All de novo transcripts overlapping or in proximity of putative imprinted gDMRs were identified, and their transcription initiation site was confirmed by visual inspection of splice sites. Transcripts initiating in transposable elements were identified using LIONS[20,31]. The 5′ ends of de novo assembled transcripts were classified based on overlaps with UCSC Repeat Masker annotation and our putative igDMR list, and transcripts with the Up or UpEdge classification (per LIONS raw output) were taken into consideration. For manual inspection of LITs over specific CGI promoters, LITs were either identified by manually validating transcripts with EInside transposable element contribution in the LIONS raw output or by intersecting de novo transcripts with the boundaries of hypermethylated domains.

Evolutionary tree of LTR families. Phylogenetic tree of species shown was generated from TimeTree (http://timetreebeta.igem.temple.edu/). Integration of LTR families was imputed based on the presence or absence of the family in the species investigated, and colonization time is indicated at the root of those species in which it was found in common. For examining the evolutionary history of MT2A family, including the MT2A apparently responsible for de novo methylation of the *Slc38a4* igDMR in mouse oocytes, we identified all annotated (UCSC RepeatMasker) MT2A elements >450 bp (the MT2A consensus sequence is 533 bp). Multiple sequence analysis and phylogenetic tree construction were carried out with MEGA X[75] using MUSCLE UPMGA algorithm and by Neighbour-Joining method, respectively. Visualization and annotation of the phylogenetic tree were carried out with iTOL (https://itol.embl.de/)[76].

Ethical approval for animal work. All mouse experiments were approved by the UBC Animal Care Committee under certificates A15-0291 and A15-0181, and complied with the national Canadian Council on Animal Care guidelines for the ethical care and use of experimental animals. All primate experiments were approved by the animal experiment committee of Primate Research Institute (PRI) of Kyoto University (Approval No: 2018-004). Three rhesus macaque (Macaca mulatta) placentas and their parental blood DNA were collected and stored at −80 °C before use. Three chimpanzee (Pan troglodytes verus) placentas and their parental blood DNA were provided from Kumamoto Zoo and Kumamoto Sanctuary via Great Ape Information Network (GAIN).

SNP genotyping and targeted bisulfite analysis in primates. Regions orthologous to human placental igDMR were PCR-amplified with TaKaRa EX Taq HS (TaKaRa Bio) with primers specific for chimpanzee and rhesus macaque CGIs (Supplementary Data 3). PCR were performed under the following conditions: 1 min at 94 °C, then 40 cycles of 30 s at 94 °C, 30 s at 57 °C, 30 s at 72 °C, and 5 min final elongation at 72 °C. PCR products were purified with QIAquick PCR

Purification Kit (Qiagen). Sanger sequencing was carried out for each amplicon using the BigDye Terminator v3.1 Cycle Sequencing Kit and the ABI 3130xl DNA Analyzer (Thermo Fisher Scientific). The sequencing data were assembled using ATGC sequence assembly software (bundled with GENETYX Ver. 14, Genetyx Corporation) to identify SNPs at a given loci and determine the genotypes for each individual sample. Genomic DNA (100 ng per sample) was treated with sodium bisulfite using the DNA Methylation Gold Kit (Zymo Research). The bisulfite-treated DNA was PCR-amplified with TaKaRa EpiTaq HS (TaKaRa Bio) with primers specific for chimpanzee and rhesus macaque CGIs (Supplementary Data 3). PCR were performed under the following conditions: 1 min at 94 °C, then 35–37 cycles of 30 s at 94 °C, 30 s at 57 °C, 30 s at 72 °C, and 5 min final elongation at 72 °C. PCR products were purified with QIAquick PCR Purification Kit (Qiagen). Amplicons were tagged using NEB Next Ultra II DNA Library Prep Kit for Illumina (NEB) with five thermal cycles for amplification and subjected to paired-end sequencing on a MiSeq platform using MiSeq Reagent Nano Kit v2, 300 Cycles (Illumina). We used the QUMA website (http://quma.cdb.riken.jp/top/index.html) to quantify CpG methylation levels of each CGI. From each sequence data (fastq files), the top 400 forward reads (R1) were extracted and mapped to each CGI with QUMA's default conditions and uniquely mapped reads were used to calculate DNA methylation level per CpG dinucleotide site. Due to mappability issues, 20 mismatches were allowed for mapping the amplicons over the chimpanzee *ZNF396* gDMR locus. When SNPs were available, reads perfectly matching each parental genome were extracted and aligned separately for allele-specific analyses.

ChIP-sequencing in mouse oocytes. Germinal vesicle oocytes (GVOs) were isolated from 7–10-weeks-old C57BL/6J females. The zona pellucida was dissolved by passaging the oocytes through an acid Tyrode's solution, followed by neutralization in M2 media. The oocytes were re-suspended in nuclear isolation buffer (Sigma), flash-frozen in liquid nitrogen and stored at −80 °C until usage. RNA polymerase II ChIP-seq libraries were prepared from ~200 GVOs using a modified version of ULI-NChIP-seq[77]. Briefly, chromatin was fragmented with MNase (NEB), diluted in native ChIP buffer (20 mM Tris-HCl pH 8.0; 2 mM EDTA; 150 mM NaCl; 0.1% Triton X-100) containing 1 mM PMSF (Sigma) and EDTA-free protease inhibitor cocktail (Roche), and incubated overnight at 4 °C with 0.1 µg of anti-RNA polymerase II monoclonal antibodies (Abcam ab817–8WG16-) and 5 µl of protein A: protein G 1:1 Dynabeads (Thermofisher). Antibody-bond chromatin was washed four times in low salt buffer (20 mM Tris-HCl pH 8.0; 2 mM EDTA; 150 mM NaCl; 1% Triton X-100; 0.1% SDS) and eluted at 65 °C for 1 h in elution buffer (0.1 nM NaCO3; 1% SDS). DNA was then extracted using phenol:chloroform and precipitated in 75% ethanol, followed by paired-end library construction. Libraries were sequenced (75 bp paired-end) on a NextSeq 500 according to manufacturer's protocols.

ERV deletions using CRISPR-Cas9. The ERVs at *Slc38a4* and *Impact* were deleted in mouse ESCs using transient delivery of sgRNAs and Cas9 from an expression vector similar to pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene #42230)[78]. Plasmid phU6-gRNA-CBh-hCas9 was obtained from Rupesh Amin and Mark Groudine and first modified by insertion of the PGK-puro-pA selectable marker from pPGKpuro (Addgene #11349)[79] to obtain pPuro2-hU6-gRNA-CBh-Cas9. Individual sgRNAs were designed from http://crispr.mit.edu/ and incorporated on a single forward oligonucleotide (IDT) including a 5′ SapI site and a 29-bp region of complementarity to a universal reverse primer containing an XbaI site. All primer sequences are shown in Supplementary Data 3. For each sgRNA, 10 pmoles of forward and universal reverse primers were annealed and extended with Q5 high-fidelity DNA polymerase (NEB) according to the recommended conditions and with the following program: 3 min/96 °C, three cycles of 3 s/96 °C-30 s/50 °C-3 min/72 °C, and final extension for 5 min/72 °C. The PCR reaction was cleaned with the QIAquick PCR Purification Kit (QIAGEN), and cloned as a ~110-bp SapI-XbaI fragment into pPuro2-hU6-gRNA-CBh-Cas9. For each deletion, we designed 4 different sgRNAs, 2 targeting each side of the ERV (Supplementary Fig. 9), and tested the efficiency of each individual guide using an endonuclease assay. For these assays, 1 µg of each sgRNA plasmid were transfected in C57BL/6N C2 ESCs[80] by lipofection (Lipofectamine 2000, ThermoFisher). After 16 h, puromycin selection (4 µg/ml) was started and continued for 48 h, after which the cells were grown without selection for 5 days. Whole cell populations were collected and prepared for PCR by HotSHOT[81]. Genomic PCR was performed with primers flanking the sgRNA site (Supplementary Data 3) and purified amplicons were melted, reannealed, and then digested using T4 Endonuclease I (NEB). Cut/uncut ratios were calculated following agarose gel electrophoresis and the most efficient sgRNA pairs were chosen for subsequent use. For each deletion, sgRNA plasmid pairs (total of ~10 µg each) were electroporated into C2 cells as previously described[80]. After 48 h of puromycin selection, cells were grown in ESM for 7 days. Single colonies were picked and expanded, then screened for full-length LTR deletion by PCR, Sanger sequencing, and chromosome contents. Euploid lines containing full-length deletion were selected for aggregation chimera production.

Injection chimera production. Blastocysts were collected at E3.5 from female albino C57Bl/6J-TyrC2J mice following natural mating, and placed in a 200 ul drop of KSOM overlaid with embryo-tested mineral oil and incubated at 37 °C until ready for injection. Pooled clones of *Slc38a4* LTR KO (#1, #2 and #10) or *Impact* LTR KO (#3 and #5) ESCs were suspended in injection media (DMEM with HEPES + 25% KO ES medium without LIF/2i), and blastocysts were injected with 8–10 ES cells per embryo. After injection, 10–20 embryos were implanted into E2.5 pseudo pregnant mice via uterine transfer. In total, 28 and 7 chimeric pups were born for the $Slc38a4^{MT2AKO}$ and $Impact^{MTCKO}$ mutant ESCs, respectively. From these, at least three chimeric males per line transmitted a deletion allele. Transgenic mice were bred and maintained in the Centre for Disease Modelling, Life Sciences Institute, University of British Columbia, under pathogen-free conditions. In all heterozygous genotypes, the maternally inherited allele is always presented first.

Allele-specific expression analysis. E13.5 embryos and placentae were dissected from reciprocal F1 crosses between heterozygous LTR KO animals on the C57BL/6 background and WT CAST mice. Embryonic heads for the *Impact* MTCKO and placentae for the *Slc38a4* MT2AKO crosses were used for RNA extraction using Trizol (ThermoFisher). RNA was DNase I-treated (Fisher) at 37 °C for 1 h, then heat inactivated at 65 °C for 15 min. cDNA was synthesized with MMLV-RT (ThermoFisher) using N15 oligonucleotides. RT-PCR was performed with primers described in Supplementary Data 3, and purified PCR products were subjected to restriction enzyme digestion (PvuII for *Slc38a4* or MluCI for *Impact*) and resolved on a 2% agarose gel. Uncut PCR products were sent for Sanger sequencing, and.ab1 files were analyzed using the PHRED software (http://www.phrap.org/phredphrapconsed.html).

Bisulfite sequencing. E13.5 embryos and placentae were dissected from reciprocal F1 crosses between heterozygous LTR KO animals on the C57BL/6 background and wild-type CAST/EiJ mice obtained from the Jackson Laboratory (stock number 000735). Bodies were minced and gDNA was extracted by proteinase K digestion followed by phenol:chloroform extraction. One hundred nanogram of gDNA was vortexed to shear the DNA, denatured in 0.2 M NaOH, bisulfite converted in 0.225 M NaOH; 0.0125% hydroquinone; 4 M $NaHSO_3$ overnight at 50 °C, cleaned using the Wizard DNA cleanup kit (Promega) and desulfonated in 0.3 M NaOH at 37 °C for 30 min. Oocytes were harvested from females of the appropriate genotype at 21–28 days of age. GVOs were isolated by passing dissected ovaries through a 100 µm filter using a blunt tool, applying the filtrate to a 35 µm filter, then backwashing the trapped GVOs from the filter. GVOs were further purified by manual collection. >200 Oocytes were pooled from multiple females (typically 4–5 females total). Oocyte DNA was harvested by incubation in 0.1% SDS 1 µg/µl Proteinase K in the presence of 1 µg lambda DNA at 37 °C for 60 min followed by 98 °C for 15 min. DNA was converted using EZ DNA Methylation Gold Kit (Zymo) per manufacturer's protocols. Three biological replicates for oocytes and three independent BS conversions for embryos were amplified using semi-nested primers (Supplementary Data 3) and touchdown-PCR conditions. Purified PCR products were TA-cloned into pGEM-T (Promega), and sequenced at the McGill/Genome Quebec Innovation Centre. Sequences were analyzed using BiQ Software (https://biq-analyzer.bioinf.mpi-inf.mpg.de/). Informative SNPs were identified in final sequences and used to identify individual C57BL/6 or CAST strands.

Quantitative RT-PCR. To measure total levels of expression, quantitative PCR (RT-qPCR) was performed on cDNA (described above) using Eva Green (Biotium) on a Step-One Plus Real time PCR system (Applied Biosystems). All reactions were run as follows: 2 min at 95 °C, then 40 cycles of 30 s at 95 °C, 30 s at 60 °C, 30 s at 72 °C, and fluorescence was read at 80 °C. Ct values of six biological replicates from two litters for each stage were averaged and used to calculate relative amounts of transcripts, normalized to levels of the housekeeping gene *Ppia*. Primer sequences are available in Supplementary Data 3.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data generated for this manuscript were deposited at GEO datasets under the accession GSE126363. Tracks for oocyte data analyzed here can be accessed through the data hub https://datahub-d85hei26.udes.genap.ca/NatComm2019/hub.txt. Datasets analyzed for this manuscript are detailed in Supplementary Data 1. The source data underlying Figs. 3c, d, 5d–g, and 6d–g and Supplementary Figs 6 and 9a, b, f, g are provided as a Source Data file.

Received: 31 July 2019; Accepted: 14 November 2019;
Published online: 12 December 2019

References

1. Stewart, K. R. et al. Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes Dev.* **29**, 2449–2462 (2015).

2. Gahurova, L. et al. Transcription and chromatin determinants of de novo DNA methylation timing in oocytes. *Epigenetics Chromatin* **10**, 25 (2017).
3. Veselovska, L. et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol.* **16**, 209 (2015).
4. Shirane, K. et al. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **9**, e1003439 (2013).
5. Kobayashi, H. et al. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8**, e1002440 (2012).
6. Kobayashi, H. et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
7. Tang, W. W. et al. A unique gene regulatory network resets the human germline epigenome for development. *Cell* **161**, 1453–1467 (2015).
8. Guo, F. et al. The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* **161**, 1437–1452 (2015).
9. Okae, H. et al. Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet.* **10**, e1004868 (2014).
10. Xu, Q. et al. SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat. Genet.* **51**, 844–856 (2019).
11. Smallwood, S. A. et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **43**, 811–814 (2011).
12. Hiura, H., Obata, Y., Komiyama, J., Shirai, M. & Kono, T. Oocyte growth-dependent progression of maternal imprinting in mice. *Genes Cells* **11**, 353–361 (2006).
13. Singh, V. B. et al. Blocked transcription through KvDMR1 results in absence of methylation and gene silencing resembling Beckwith-Wiedemann syndrome. *Development* **144**, 1820–1830 (2017).
14. Smith, E. Y., Futtner, C. R., Chamberlain, S. J., Johnstone, K. A. & Resnick, J. L. Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. *PLoS Genet.* **7**, e1002422 (2011).
15. Chotalia, M. et al. Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev.* **23**, 105–117 (2009).
16. Joh, K. et al. Growing oocyte-specific transcription-dependent de novo DNA methylation at the imprinted Zrsr1-DMR. *Epigenetics Chromatin* **11**, 28 (2018).
17. Bretz, C. L. & Kim, J. Transcription-driven DNA methylation setting on the mouse Peg3 locus. *Epigenetics* **12**, 945–952 (2017).
18. Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Developmental Cell* **7**, 597–606 (2004).
19. Franke, V. et al. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res.* **27**, 1384–1394 (2017).
20. Brind'Amour, J. et al. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat. Commun.* **9**, 3331–3344 (2018).
21. Schulz, R. et al. The parental non-equivalence of imprinting control regions during mammalian development and evolution. *PLoS Genet.* **6**, e1001214 (2010).
22. Hanna, C. W. et al. Pervasive polymorphic imprinted methylation in the human placenta. *Genome Res.* **26**, 756–767 (2016).
23. Court, F. et al. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Res.* **24**, 554–569 (2014).
24. Sanchez-Delgado, M. et al. Absence of maternal methylation in biparental hydatidiform moles from women with NLRP7 maternal-effect mutations reveals widespread placenta-specific imprinting. *PLoS Genet.* **11**, e1005644 (2015).
25. Partida, G. et al. Genome-wide survey of parent-of-origin effects on DNA methylation identifies candidate imprinted loci in humans. *Hum. Mol. Genet.* **27**, 2927–2939 (2018).
26. Hamada, H. et al. Allele-specific methylome and transcriptome analysis reveals widespread imprinting in the human placenta. *Am. J. Hum. Genet.* **99**, 1045–1058 (2016).
27. Proudhon, C. et al. Protection against de novo methylation is instrumental in maintaining parent-of-origin methylation inherited from the gametes. *Mol. Cell* **47**, 909–920 (2012).
28. Wang, L. et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
29. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
30. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **487**, 57 (2017).
31. Babaian, A. et al. LIONS: analysis suite for detecting and quantifying transposable element initiated transcription from RNA-seq. *Bioinforma. (Oxf., Engl.)* **7**, 24 (2019).
32. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
33. Mizuno, Y. et al. Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. *Biochem. Biophys. Res. Commun.* **290**, 1499–1505 (2002).
34. Smith, R. J., Dean, W., Konfortova, G. & Kelsey, G. Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res.* **13**, 558–569 (2003).
35. Hagiwara, Y. et al. Screening for imprinted genes by allelic message display: identification of a paternally expressed gene impact on mouse chromosome 18. *Proc. Natl Acad. Sci. USA* **94**, 9249–9254 (1997).
36. Sanchez-Delgado, M., Vidal, E. & Medrano, J. Human oocyte-derived methylation differences persist in the placenta revealing widespread transient imprinting. *PLoS Genet.* **12**, e1006427 (2016).
37. Metsalu, T. et al. Using RNA sequencing for identifying gene imprinting and random monoallelic expression in human placenta. *Epigenetics* **9**, 1397–1409 (2014).
38. Yuen, R., Jiang, R., Peñaherrera, M. S., Mcfadden, D. E. & Robinson, W. P. Genome-wide mapping of imprinted differentially methylated regions by DNA methylation profiling of human placentas from triploidies. *Epigenetics Chromatin* **4**, 10 (2011).
39. Okae, H. et al. Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Hum. Mol. Genet.* **21**, 548–558 (2012).
40. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
41. Ruebel, M. L. et al. Transcriptome analysis of rhesus monkey failed-to-mature oocytes: deficiencies in transcriptional regulation and cytoplasmic maturation of the oocyte mRNA population. *Mol. Hum. Reprod.* **24**, 478–494 (2018).
42. Gao, F. et al. De novo DNA methylation during monkey pre-implantation embryogenesis. *Cell Res.* **27**, 526–539 (2017).
43. Okamura, K., Sakaki, Y. & Ito, T. Comparative genomics approach toward critical determinants for the imprinting of an evolutionarily conserved gene Impact. *Biochem. Biophys. Res. Commun.* **329**, 824–830 (2005).
44. Okae, H. et al. RNA sequencing-based identification of aberrant imprinting in cloned mice. *Hum. Mol. Genet.* **23**, 992–1001 (2014).
45. Matoba, S. et al. Paternal knockout of Slc38a4/SNAT4 causes placental hypoplasia associated with intrauterine growth restriction in mice. *Proc. Natl Acad. Sci. USA* **116**, 21047–21053 (2019).
46. Hanna, C. W. et al. MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. *Nat. Struct. Mol. Biol.* **25**, 73–82 (2018).
47. Wood, A. J. et al. A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS Genet.* **3**, e20 (2007).
48. Cowley, M. & Oakey, R. J. Retrotransposition and genomic imprinting. *Brief. Funct. Genomics* **9**, 340–346 (2010).
49. Inoue, A., Jiang, L., Lu, F., Suzuki, T. & Zhang, Y. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547**, 419–424 (2017).
50. Schmitges, F. W. et al. Histone methylation by PRC2 is inhibited by active chromatin marks. *Mol. Cell* **42**, 330–341 (2011).
51. Matoba, S. et al. Loss of H3K27me3 imprinting in somatic cell nuclear transfer embryos disrupts post-implantation development. *Cell Stem Cell* **23**, 343–354. e5 (2018).
52. Takahashi, N. et al. ZNF445 is a primary regulator of genomic imprinting. *Genes Dev.* **33**, 49–54 (2019).
53. Auclair, G. et al. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome Res.* **26**, 192–202 (2016).
54. Barlow, D. Methylation and imprinting: from host defense to gene regulation? *Science* **260**, 309–310 (1993).
55. Yoder, J., Walsh, C. & Bestor, T. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
56. McDonald, J. F., Matzke, M. A. & Matzke, A. J. Host defenses to transposable elements and the evolution of genomic imprinting. *Cytogenetic Genome Res.* **110**, 242–249 (2005).
57. Pask, A. J. et al. Analysis of the platypus genome suggests a transposon origin for mammalian imprinting. *Genome Biol.* **10**, R1 (2009).
58. Watanabe, T. et al. Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* **332**, 848–852 (2011).
59. Song, Q. et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS ONE* **8**, e81148 (2013).

60. Consortium, E. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
61. Hahn, O. et al. Dietary restriction protects from age-associated DNA methylation and induces epigenetic reprogramming of lipid metabolism. *Genome Biol.* **18**, 56 (2017).
62. Harten, S. K. et al. The recently identified modifier of murine metastable epialleles, Rearranged L-Myc Fusion, is involved in maintaining epigenetic marks at CpG island shores and enhancers. *BMC Biol.* **13**, 21 (2015).
63. Hodges, E. et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* **44**, 17–28 (2011).
64. Hon, G. C. et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* **45**, 1198–1206 (2013).
65. Mendizabal, I. et al. Comparative methylome analyses identify epigenetic regulatory loci of human brain evolution. *Mol. Biol. Evol.* **33**, 2947–2959 (2016).
66. Molaro, A. et al. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**, 1029–1041 (2011).
67. Okae, H. et al. Derivation of human trophoblast stem cells. *Cell Stem Cell* **22**, 50–63.e6 (2018).
68. Consortium, R. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
69. Schroeder, D. I. et al. Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. *PLoS Genet.* **11**, e1005442 (2015).
70. Schroeder, D. I., Lott, P., Korf, I. & LaSalle, J. M. Large-scale methylation domains mark a functional subset of neuronally expressed genes. *Genome Res.* **21**, 1583–1591 (2011).
71. Tung, J. et al. Social environment is associated with gene regulatory variation in the rhesus macaque immune system. *Proc. Natl Acad. Sci. USA* **109**, 6490–6495 (2012).
72. Zeng, J. et al. Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *Am. J. Hum. Genet.* **91**, 455–465 (2012).
73. Haeussler, M. et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).
74. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma. (Oxf., Engl.)* **29**, 15–21 (2012).
75. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
76. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
77. Brind'Amour, J. et al. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.* **6**, 6033 (2015).
78. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
79. Tucker, K. et al. Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.* **10**, 1008–1020 (1996).
80. Gertsenstein, M. et al. Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS ONE* **5**, e11260 (2010).
81. Truett, G. et al. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques* **29**, 52–54 (2000).

Acknowledgements

We thank Rupesh Amin and Mark Groudine for their CRISPR-Cas9 vector, Hiromi Kamura, Prof. Kenichiro Hata (National Center for Child Health and Development), Tomoya Takashima, Dr. Keisuke Tanaka, and Prof. Tomohiro Kono (Tokyo University of Agriculture) for their technical assistance with targeted bisulfite sequencing analysis, Julien Richard Albert and Amanda Ha for technical assistance, and Dixie Mager and Carolyn Brown for helpful discussions. M.C.L. is supported by CIHR Grant PJT-153049 and NSERC Discovery Grant 2015-05228. L.L is supported by CIHR Grant MOP-119357 and NSERC Discovery Grant 386979-12. H.K. is supported by MEXT KAKENHI Grant Numbers 15H05579 and JP18H05553. H.I is supported by MEXT KAKENHI Grant Numbers 15H05242 and 18H04005. A portion of primate samples were provided through the Great Ape Information Network (supported by The Cooperative Research Programs of the Primate Research Institute and Wildlife Research Center, Kyoto University). H.K. and H.I. are supported by The Cooperative Research Programs of the Genome Research for BioResource, NODAI Genome Research Center, Tokyo University of Agriculture, and Primate Research Institute and Wildlife Research Center, Kyoto University.

Author contributions

L.L., M.C.L., H.K., J.B., and A.B.B. conceived the project and designed the experiments. L.L., A.B.B., K.N.J., and J.B. performed murine experiments. H.K., K.N., and H.I. performed primate experiments. A.B.B., J.B., and H.K. performed data analyses. L.L., M.C.L., H.K., J.B., and A.B.B. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-13662-9.

Correspondence and requests for materials should be addressed to M.C.L. or L.L.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 February 2019 | Revised: 10 June 2019 | Accepted: 12 June 2019

DOI: 10.1111/mec.15178

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Monarch butterflies use an environmentally sensitive, internal timer to control overwintering dynamics

Delbert A. Green II[1,2] | **Marcus R. Kronforst**[1]

[1]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA

[2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Correspondence
Delbert A. Green II, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA.
Email: dgreenii@umich.edu

Present address
Delbert A. Green, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Funding information
Division of Integrative Organismal Systems, Grant/Award Number: IOS-1452648; U.S. Fish and Wildlife Service, Grant/Award Number: F17AC01222; National Institute of General Medical Sciences, Grant/Award Number: GM108626; National Science Foundation, Grant/Award Number: PRFB 1523759

Abstract

The monarch butterfly (*Danaus plexippus*) complements its iconic migration with diapause, a hormonally controlled developmental programme that contributes to winter survival at overwintering sites. Although timing is a critical adaptive feature of diapause, how environmental cues are integrated with genetically-determined physiological mechanisms to time diapause development, particularly termination, is not well understood. In a design that subjected western North American monarchs to different environmental chamber conditions over time, we modularized constituent components of an environmentally-controlled, internal diapause termination timer. Using comparative transcriptomics, we identified molecular controllers of these specific diapause termination components. Calcium signalling mediated environmental sensitivity of the diapause timer, and we speculate that it is a key integrator of environmental condition (cold temperature) with downstream hormonal control of diapause. Juvenile hormone (JH) signalling changed spontaneously in diapause-inducing conditions, capacitating response to future environmental condition. Although JH is a major target of the internal timer, it is not itself the timer. Epigenetic mechanisms are implicated to be the proximate timing mechanism. Ecdysteroid, JH, and insulin/insulin-like peptide signalling are major targets of the diapause programme used to control response to permissive environmental conditions. Understanding the environmental and physiological mechanisms of diapause termination sheds light on fundamental properties of biological timing, and also helps inform expectations for how monarch populations may respond to future climate change.

KEYWORDS
diapause, ecdysone, insulin signalling, juvenile hormone, monarch butterfly

1 | INTRODUCTION

Organisms subject to seasonally variable environments have evolved myriad adaptations to translate environmental condition into accurately timed behavioural and physiological responses. The monarch butterfly (*Danaus plexippus*) has evolved migration and diapause to survive inhospitable winters across its temperate North American range. Each fall, millions of monarchs across the US and southern Canada migrate to specific locations in central Mexico if in the eastern North American population, or along the Pacific Coast if in the western North American population, where they overwinter in reproductive diapause. In spring, they mate, remigrate northwards, and repopulate their breeding grounds over three to four successive generations. Previous work has revealed environmental cues and sensory modalities required to accurately interpret environmental condition for both migration and diapause initiation (Gegear, Casselman, Waddell, & Reppert, 2008; Goehring & Oberhauser, 2002, 2004; Guerra, Gegear, & Reppert, 2014; Guerra & Reppert,

2013; Merlin, Gegear, & Reppert, 2009; Zhu et al., 2008), as well as hormonal mechanisms that control monarch diapause development (Barker & Herman, 1973; Dallmann & Herman, 1978; Herman, 1975; Herman & Lessman, 1981). The missing link, however, is how monarchs, and other diapausers, integrate external cues with internal, genetically-controlled responses to achieve specifically timed seasonal responses (Hand, Denlinger, Podrabsky, & Roy, 2016). Here, we use monarch diapause termination as a model to understand molecular control of environmental response and seasonal timing.

Diapause development proceeds through stereotypic ecophysiological phases: initiation, maintenance, and termination (Andrewartha, 1952; Koštál, 2006). Together, initiation and termination define the specific time interval of diapause maintenance, the environmentally insensitive refractory period during which organisms experience suppressed metabolic rate, bolstered stress resistance, and halted reproductive development (Herman, 1973). Upon termination, individuals either immediately resume nondiapause development if prevailing conditions are permissive (e.g., warm), or as is more often the case, enter post-diapause quiescence, an environmentally sensitive dormancy, in adverse conditions (e.g., cold). Monarch diapause, like that of many other temperate species, is primarily induced by low and decreasing photoperiod, and is also enhanced by low temperature and host plant quality (Goehring & Oberhauser, 2002). Termination, on the other hand, is much more variable among insects and is rarely under photoperiodic control in hibernal diapauses (Hodek, 2002; Koštál, 2006; Liedvogel & Lundberg, 2014; Tauber, Tauber, & Masaki, 1986). Eastern and western North American monarch populations terminate diapause before the winter solstice (Herman, 1981; Herman, Brower, & Calvert, 1989), proving that increasing photoperiod is not a relevant termination cue, although not excluding any photoperiodic involvement. Counterintuitively, cold temperature often hastens spontaneous diapause termination in overwintering diapausers (reviewed in Tauber et al., 1986). How cold temperature controls diapause timing in insects is unknown. More generally, how nonphotoperiodically controlled developmental timing occurs is a little explored and open problem.

Hormonal signalling plays critical roles in insect diapauses. In preadult diapauses, 20-hydroxyecdysone (20-HE) is most often recognized as the diapause terminator (reviewed in Denlinger, Yocum, & Rinehart, 2011). The role of 20-HE in adult reproductive diapause is more variable. 20HE plays opposing roles in different organisms, promoting diapause termination and reproduction in fruit flies (*Drosophila melanogaster*), locusts (*Locusta migratoria*), and European firebugs (*Pyrrhocoris apterus*), while being associated with diapause maintenance in Colorado potato beetles (*Leptinotarsa decimlineata*) (reviewed in Denlinger et al., 2011). Insulin/insulin-like peptide signalling (IIS) has been shown to influence adult reproductive diapause phenotypes in fruit flies and mosquitoes (*Culex pipens*) (Sim & Denlinger, 2008; Williams et al., 2006). The consistent controller of adult reproductive diapause across insects is juvenile hormone (JH) (reviewed in Tauber et al., 1986). Increased JH titre is associated with reproductively active monarchs in spring and summer (Lessman et al., 1989). JH signalling is required for reproductive development in nonmigrating monarchs, and exogenous JH analogue application is sufficient to terminate diapause and induce reproductive development in normally nonreproductive migrant monarchs in summer-like conditions (Barker & Herman, 1973; Herman, 1975). While these studies consistently associate JH signalling with reproductive state, they leave open the specific mechanism by which JH controls diapause because they do not follow JH signalling dynamics with sufficient temporal resolution across diapause development. Despite knowing that these hormonal pathways are associated with particular diapause states, it is not clear how these pathways actually function in diapause development. Are they involved in transducing the environmental signal? Do they engender maintenance phenotypes? Do they specify termination? Moreover, JH, ecdysteroid, and IIS pathways interact to achieve coherent responses in different contexts. For example, in *Drosophila*, JH does not control metamorphosis timing as it does in *Manduca sexta* (Nijhout & Williams, 1974; Riddiford, Truman, Mirth, & Shen, 2010). Rather, JH indirectly regulates larval growth rate through controlling ecdysone concentration, versus its timing, whereby influencing IIS pathway activity, which is a well known controller of larval growth rate (Mirth et al., 2014). How do these interactions change in different contexts, and what role, if any, does diapause play in influencing interactions between JH, ecdysteroid, and IIS?

Here, we examined diapause termination in the western North American monarch population in order to understand (a) the relative influence of external environmental cues versus internal physiological mechanisms for terminating diapause; (b) what molecular mechanisms mediate environmental sensitivity during diapause development; and (c) what controls diapause timing. We took a comparative transcriptomics approach to investigate the molecular basis of termination because this allowed us to simultaneously assay genome-wide responses in situ without molecular manipulation of the system.

2 | MATERIALS AND METHODS

2.1 | Experimental design

Herman (1981) previously demonstrated that female monarchs from California undergo true diapause from September to December as measured by decreased sensitivity of reproductive organs to artificial summer-like conditions. We conducted a nearly identical experiment in which we tested responsiveness of wild-caught female monarchs ("Natural" cohort) to simulated summer conditions in an environmental chamber over the course of the overwintering season (November 2015–January 2016). We extended the Herman analysis by adding a treatment in which monarchs that were collected in November were held over the course of the overwintering season in an environmental chamber that approximated conditions at the overwintering site in November ("Chamber" cohort). Female overwintering monarch butterflies were collected from private property (with appropriate permissions granted) near Pismo Beach, CA.

Monarchs were batch collected from roosts and kept in cool, moist conditions until overnight shipment to the University of Chicago (within 48 hr).

Natural cohort females were collected from the overwintering site three times throughout the season and initially acclimated for 4 days in "fall" conditions: 10 hr light at 17°C, 14 hr dark at 10°C (Figure 1a). The initial collection in November was large in order to establish the "Chamber" cohort. Chamber cohort females remained in the chamber for the duration of the experiment. After 4 days in fall conditions, we switched a random sample of Chamber cohort females, as well as newly acclimated Natural cohort females, into "summer" conditions (16 hr day at 25°C, 8 hr night at 18°C) to test for reproductive development after 10 days. At the same acclimation day 4, we evaluated ovary development for a random sample of five females from both Chamber and Natural cohorts without transfer to summer conditions. Samples sizes for each group are indicated in Figure 1b. We set minimum target sample sizes of n = 5 individuals per mature oocytes (MO) count group and n = 3 individuals per group for sequencing. We met these targets for all but the Natural December summer cohort for which we obtained four MO counts and sequenced transcriptomes for two individuals. Individual numbers were largely determined by the number of individuals that survived shipment (Natural cohort), extended containment in the fall chamber (Chamber cohort), and experimental (summer condition) treatment. With these sample sizes, we saw similar variation to that observed by Herman (1981), suggesting that these sizes were sufficient to capture variation in the trait.

Chamber cohort individuals in fall conditions were fed butterfly nectar twice per week. Natural cohort individuals were fed once the day after arrival at the University of Chicago. Chamber cohort individuals were fed on the same day. Therefore, all individuals were fed 3 days prior to removal from fall conditions for dissection or for transfer to summer conditions. All individuals were fed every other day in the summer condition.

We chose to analyze MO in females after 10 days in summer conditions as a measure of reproductive maturity. MOs were defined as fully-chorionated oocytes with ridges appearing along the length. In a small number of individuals (n = 4 total; n = 2 in each December and January Natural cohorts), only a single ovary was significantly developed to the stage of containing MOs. We only recorded MOs for those females in which both ovaries were distinctly visible, indicating normal oocyte development. We chose to measure diapause response in females because Herman (1981) found that male diapause response measurement was less robust and potentially more variable. Female MO number showed the strongest response to diapause of all the reproductive phenotypes, male or female, measured in the study. Female reproductive organs also responded in similar degree to one another, while male reproductive organs showed variability in their responses. Despite this differential response, Herman (1981) did demonstrate that a clear diapause period also exists in male monarchs. This all suggests that the measure of diapause in males is variable and not necessarily the diapause itself.

In no case did a female show significant ovary development (presence of mature or immature oocytes; evidence of vitellogenesis) in fall conditions. Therefore, butterflies transferred to summer conditions throughout the experiment almost certainly contained no MOs when initially placed in summer conditions. All females from all time points showed significant ovary development and contained at least one MO upon analysis at 10 days in summer conditions.

2.2 | Statistics

Comparisons of group central tendencies were made using the corrected Mann–Whitney/Wilcoxon rank-sum W test statistic (to

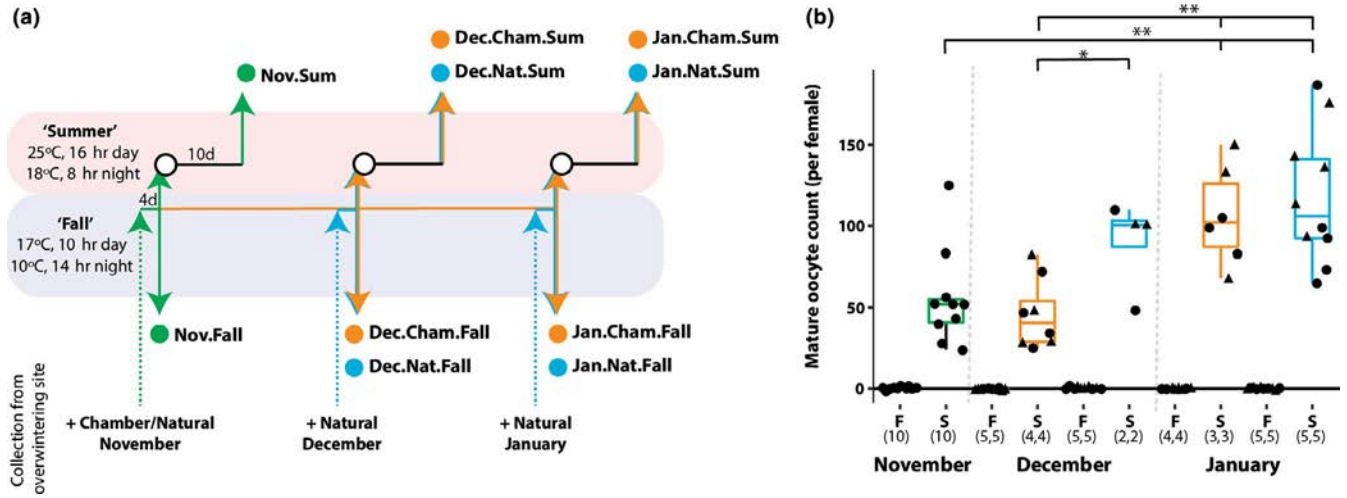


FIGURE 1 Environmental chamber modularizes internal and external components of diapause termination. (a) Overview of the design of the environmental chambers experiment. "Fall" and "summer" conditions are as noted. d, days. (b) Natural (blue) and Chamber (orange) cohorts display different diapause termination dynamics over the course of the overwintering season. Each point/shape represents an individual monarch. Triangles represent individuals included in sequence analysis. Numbers in parentheses present sample sizes. Where two numbers presented, the first is the number of individuals sequenced. F, Fall conditions; S, Summer conditions. *p < .05; **p < .005 [Colour figure can be viewed at wileyonlinelibrary.com]

account to MO count non-normality) unless otherwise noted as implemented in R version 3.5.0 (R Core Team, 2018).

2.3 | RNA extraction, library preparation and sequencing

Half of the butterflies within a cohort, time point, and treatment were selected for sequencing. Butterflies were acclimated at room temperature for approximately 1 hr prior to tissue collection. Heads were collected with antennae and proboscises removed, immediately flash frozen in liquid nitrogen, and stored at −80°C until further processing. Total RNA was extracted from single heads in two separate fractions ("large": >200 nt bases and "small": <200 nt) using the MACHEREY-NAGEL NucleoSpin miRNA kit (Cat #740971). Large RNA fractions (1 μg) were processed for ribosomal RNA depletion using a custom protocol adapted from the RNaseH hybrid degradation method described by Morlan, Qu, and Sinicropi (2012). Single-stranded synthetic DNA probes (Integrated DNA Technologies) were designed to include sequences complementary to the *D. plexippus* 5S, 18S and 28S rRNAs. Probes were 50 bases in length and designed to start 100 bases apart such that every 50 bases of each sequence were covered. Probes were resuspended in water and pooled at a final concentration of 0.5 μM for each probe. With this monarch-specific probe mix, the Morlan et al. (2012) protocol was followed, and depleted RNA was purified with the MACHEREY-NAGEL RNA Clean-up XS kit (Cat #740903).

Libraries were prepared using the KAPA Stranded RNA-Seq kit (Cat #KK8401). Samples were quality controlled by BioAnalyzer and confirmed to have no significant adapter peaks. Samples were individually indexed and sequenced as a single pool on three lanes of the Illumina HiSeq4000 (single end 50 bp reads).

2.4 | Read mapping statistics

RNA sequencing generated 639,995,826 raw reads from 33 monarch head libraries. Quality filtering resulted in 573,340,723 uniquely mapped reads (average 89.6% uniquely mapping reads). Reads were mapped to the *D. plexippus* genome assembly version 3 (Zhan & Reppert, 2012) using STAR version 2.5.2b (Dobin et al., 2012) and gene read counts were generated by htseq2 within STAR (mapped to OGS 2.0). 349,776,743 unique reads (61% of uniquely mapped reads, 54.7% of total reads) mapped to exons. This exon sequence coverage met expected coverage values given the use of an rRNA depletion protocol versus mRNA enrichment (Zhao et al., 2014).

2.5 | Data quality control, differentially expressed gene identification, and functional annotation

We used DESEQ2 version 3.8 (Love, Huber, & Anders, 2014) to identify differentially expressed genes (DEGs). To test the general effect of each experimental variable across all levels of the remaining two, we used likelihood ratio testing to identify DEGs. For individual group comparisons, a grouping variable was created that combined the cohort "source" (Chamber vs. Natural), month, and experimental condition (fall vs. summer). This variable was used to define the linear model of the counts and contrasts of individual groups that were used to identify specific groups of DEGs via the Wald test. We used clustering functions within DESEQ2 and PCAEXPLORER version 2.8.0 (Love et al., 2014; Marini & Binder, 2017) to assess variability within the data set. We used the top 5,000 genes with most variable (normalized) expression across all samples to cluster samples. For visualization and clustering, gene counts were normalized using a variance stabilizing transformation and blinded dispersion estimation as implemented in DESEQ2 by the versus *t*() function. Genes with adjusted *p*-value (padj) < .05, baseMean > 10 and |log2FoldChange| > 0.5 were considered as significantly differentially expressed. Gene set enrichment analysis was conducted in BLAST2GO version 5.2.0 (Conesa & Götz, 2008; Conesa et al., 2005; Götz et al., 2008, 2011) using default parameters. Fisher's exact test was performed using a reference set containing 10,599 genes comprising the head transcriptome (genes with baseMean > 10).

3 | RESULTS

3.1 | Chamber fall conditions modularize environmental versus internal control of diapause timing

To dissociate diapause dynamics that were a response to natural environmental cues experienced at the overwintering site versus dynamics controlled by internal physiological mechanisms, we compared reproductive development in wild-caught females ("Natural" cohort) versus those kept in an environmental growth chamber ("Chamber" cohort; Figure 1a). First, we found that Natural cohort diapause dynamics upon exposure to summer conditions are consistent with those found by Herman (1981). In the Natural cohort, MO count increased from November (mean = 55.5, s.e.m. = 5.8) to December (mean = 89.8, s.e.m. = 5.2; $W = 9$, $p = .1358$) and from December to January (January mean = 117.70, s.e.m. = 13.2; $W = 25$, $p = .5395$) (Figure 1b) in summer conditions. Comparison of November and January counts reveals a significant increase in MO count ($W = 8$, $p = .001679$; Figure 1b), while intermediate comparisons did not. In November, we recorded individuals with 125 and 83 MOs and interpret these points as females that had already terminated diapause and entered post-diapause quiescence before transfer to summer conditions. In December, a single outlier of comparatively low MO count (MO = 48; Figure 1b) was recorded among females with at least 100 MOs, probably indicating that this individual was in diapause when initially transferred to summer conditions. Herman (1981) found similar variation in this population. Altogether, these data show that the majority of individuals in the Natural cohort terminated diapause between November and December.

The Chamber cohort showed different diapause dynamics compared to the Natural cohort. Chamber cohort MO count was unchanged from November to December (December mean = 45.6, s.e.m. = 7.5; $W = 49.5$, $p = .4227$; Figure 1b) and significantly

lower than the Natural cohort in December ($W = 29.5$, $p = .02698$; Figure 1b) in summer conditions. This indicates that Chamber cohort monarchs persist in diapause in December while Natural cohort monarchs terminate diapause over the same time period. This reveals environmental regulation of diapause termination dynamics and suggests that some aspect of natural conditions promotes early diapause termination. We confirm that increasing photoperiod is not the relevant condition because photoperiod declined between the 21 November and 13 December time points, during which the Natural cohort terminated diapause. We can also reject absolute number of short photoperiod days because both Natural and Chamber cohorts experienced the same number of short days. These data leave open the possibility that decreasing photoperiod is an environmental controller because photoperiod was constant in the fall chamber. However, we do not believe decreasing photoperiod controls diapause termination because decreasing photoperiod increases incidence of diapause induction in monarchs (Goehring & Oberhauser, 2002), and variation in day length is small during this interval. Based on day length calculations at Pismo Beach, CA, photoperiod decreases an average of 59 s per day (maximum 86 s), and the average rate of change of photoperiod reduction is 2.6 s per day. Therefore, we conclude that photoperiod

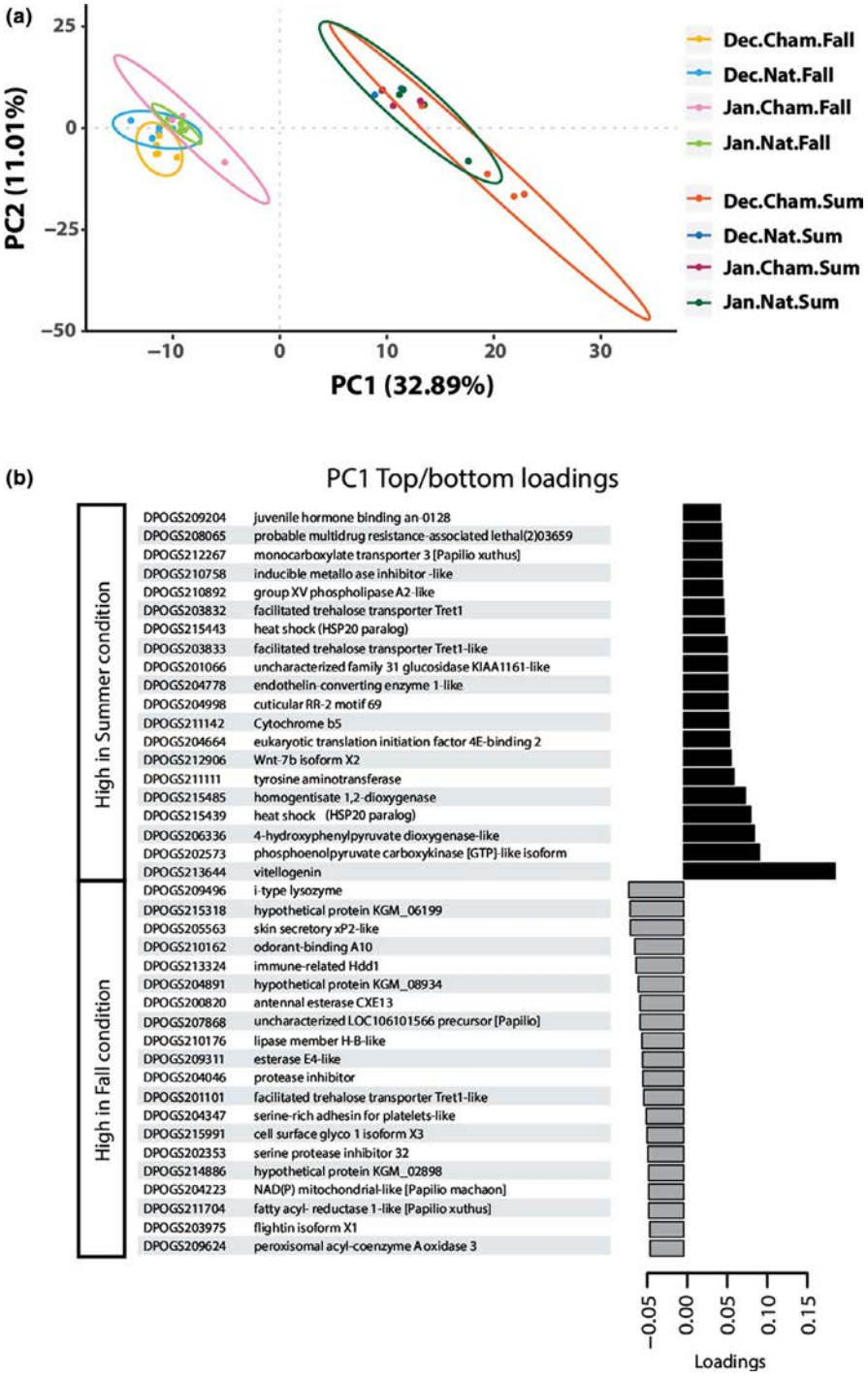


FIGURE 2 PCA analysis of head transcriptional profiles distinguishes reproductive status and diapause development. (a) The results of PCA analysis of head transcriptomes of individuals from eight indicated groups. Ellipses represent 95% confidence intervals. (b) Top 20 loadings (both positive and negative) for PC1 [Colour figure can be viewed at wileyonlinelibrary.com]

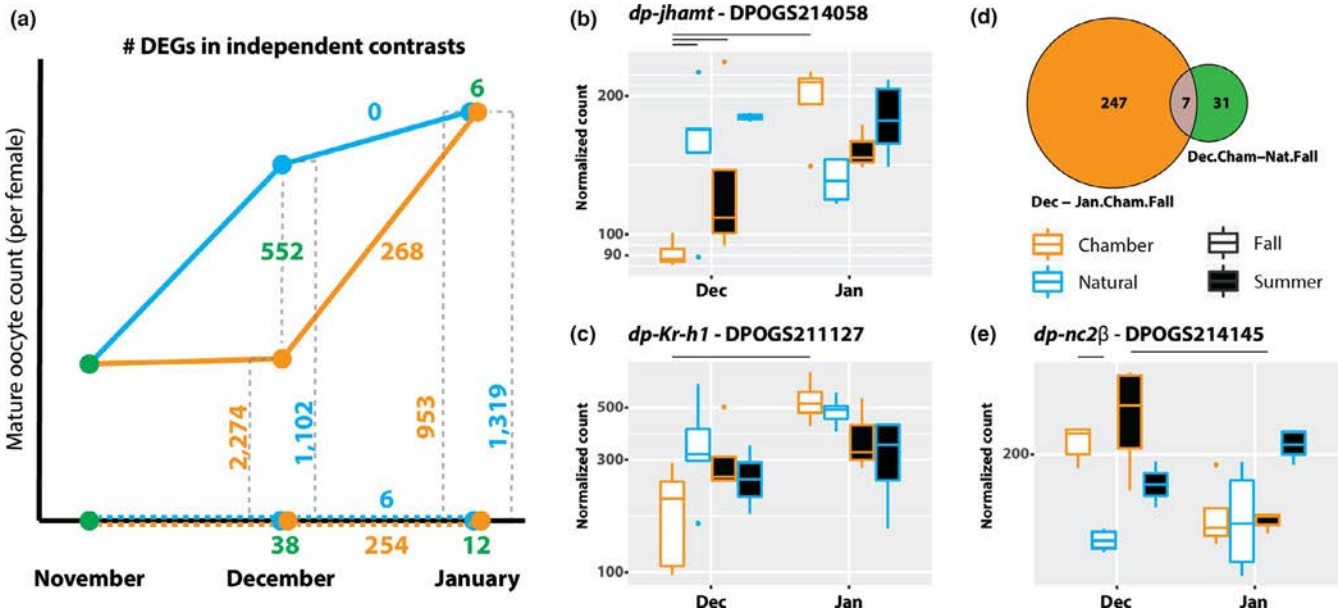


FIGURE 3 Differential termination dynamics in Chamber and Natural cohorts reveal genetic controllers of component termination processes. (a) Summary of DEG numbers across tested groups and indicated comparisons. Number of DEGs in comparisons between Chamber and Natural cohorts, Chamber cohorts, and Natural cohorts for the indicated months and conditions are indicated in green, orange, and blue, respectively. (b and c) Normalized counts of *dp-jhamt* and *dp-Kr-h1*. (d) Venn diagram of DEGs found in fall conditions in two diapause versus nondiapause comparisons (1. December vs. January Chamber cohort and 2. Chamber vs. Natural cohorts in December). (e) Normalized counts of *dp-nc2β* transcription factor. Lines under gene names indicate pairwise comparisons for which normalized counts significantly differ (FDR adjusted *p*-value < .05) [Colour figure can be viewed at wileyonlinelibrary.com]

does not play a significant role in modulating the diapause termination timer in monarchs.

Chamber cohort MO count increased significantly in January in summer conditions (mean = 106.3, s.e.m. = 12.5; W = 46, p = .002664; Figure 1b), showing that this cohort does not persist perpetually in diapause under fall conditions, as do some species (reviewed in Andrewartha, 1952; Tauber et al., 1986). Chamber cohort monarchs eventually terminate diapause between December and January despite being in a constant fall chamber without receiving natural cues. This demonstrates that diapause termination is controlled by a physiological timer. Surprisingly, Chamber cohort MO count was indistinguishable from the Natural cohort in January (W = 26.5, p = .7447; Figure 1b) in summer conditions. Nearly identical reproductive development between Natural and Chamber cohorts in January suggests that the physiological timer controls a spontaneous termination response. Taken together, these results demonstrate that an environmentally-sensitive, internal physiological timer controls spontaneous diapause termination.

3.2 | Head transcriptomes capture molecular profiles of reproductive development and diapause development

To generate hypotheses for molecular determinants of the internal timer and infer mechanisms that mediate the distinct diapause responses observed between the Chamber and Natural cohorts, we generated head transcriptomes for Chamber and Natural cohort individuals in December and January in fall (diapause/post-diapause quiescence) and summer (reproductive) conditions. We focused on heads in order to uncover putative neurohormonal and neuroendocrine mechanisms of diapause development. To get a general description of gene expression variation in the data set, we first assessed the main factors grouping the head transcriptomes via PCA on all samples. Individuals were most strongly differentiated by reproductive development (diapause/post-diapause quiescence vs. reproductive) as indicated by distinct clustering along PC1, which explained 33% of the variance in the data (Figure 2a). The number of genes differentially expressed between fall and summer conditions ranged from 953 to 2,274 (9%–21.5% of the head transcriptome; Figure 3a; Tables S2a–d). These broad transcriptomic shifts are consistent with previous transcriptome profiling of diapause development in other insects (Koštál, Štětina, Poupardin, Korbelová, & Bruce, 2017; Ragland, Denlinger, & Hahn, 2010), even though these studies were conducted with whole animals. The strongest loading on PC1 was vitellogenin (DPOGS213644; Figure 2b), whose expression has been related to diapause status in monarchs (Pan & Wyatt, 1976) and other insects (Okuda & Chinzei, 1988) because it reliably reflects the state of vitellogenesis and oocyte maturation (reviewed in Roy, Saha, Zou, & Raikhel, 2018). More generally, PC1 captures the increased metabolic activity, protein translation, and morphogenetic events related to the nondiapause reproductive state (Figure 2b). PC1 also reflects increased immune-related transcription (e.g., DPOGS209496, a predicted lysozyme, DPOGS213324, a predicted

HDD1-reltaed protein, and DPOGS202353, *Drosophila* Serpin 77Ba homolog; Figure 2b) that is characteristic of insect diapause states (Denlinger, 2002; Ragland et al., 2010). From December to January, there is a slight shift of both Chamber and Natural diapause/post-diapause quiescence cohorts along PC1 towards the reproductive expression profiles. This suggests that post-diapause quiescence is not only a distinct state from diapause, but also may involve the initial steps towards the reproductive transcriptional profile, even in fall conditions.

We next identified differentially expressed genes (DEGs) for each of the three experimental variables (month: December vs. January, source population: Chamber vs. Natural, and condition: Fall vs. Summer) across all levels of the remaining two variables. Relatively few genes were differentially expressed due to month (n = 54 DEGs) and source (n = 33 DEGs) (Tables S1a,b). No GO terms were significantly enriched, even at the highly permissive FDR value of 0.2, in either of these gene sets. In contrast, rearing condition generated the greatest changes in gene expression (n = 1,870 DEGs; Table S1c). Significant GO terms reflect increased metabolic activity (carbohydrate, chitin, amino acid, trehalose, and protein metabolism), redox activity and neurotransmission (Table 1).

3.3 | Diapause termination converges on similar transcriptional profiles in Chamber and Natural cohorts

Diapause termination is often classified as occurring naturally over an extended period of time ("horotelic") or being induced in an accelerated period of time under laboratory settings ("tachytelic"; Hodek, 2002, 2012; Koštál, 2006; Tauber et al., 1986). In our experiment, transfer to summer conditions constitutes tachytelic termination, while both the Chamber and Natural cohorts terminate diapause and enter post-diapause quiescence via horotelic termination in fall conditions. Therefore, we expected transcriptional profiles of the Chamber and Natural cohorts, in both fall and summer conditions, to be similar if termination in the two cohorts reflects the same underlying processes. Indeed, we found that head transcriptomes from Chamber and Natural cohorts in January formed overlapping clusters in the PCA in both fall and summer conditions (Figure 2). Only 12 and six genes (0.11% and 0.06% of head transcriptome, respectively) were differentially expressed between Chamber and Natural cohorts in fall and summer conditions, respectively (Figure 3a; Table 2e–f). This is despite the fact that 254 genes (2.4% of head transcriptome) were differentially expressed in the Chamber cohort between December and January (Figure 3a; Table S2g), consistent with the major transcriptomic change expected of a transition from diapause to post-diapause quiescence (Koštál et al., 2017; Ragland et al., 2010). Natural cohort transcription changed much less between December and January, showing only six DEGs (0.06% of head transcriptome; Figure 3a, Table S2h). This is consistent with the interpretation that most Natural cohort individuals have already exited diapause and entered post-diapause quiescence by December. These results demonstrate that natural and chamber diapause termination mechanisms converge upon common transcriptional profiles both in fall and summer conditions. Furthermore, they indicate that the chamber condition represents naturally-relevant diapause mechanics, revealing modular controls of diapause termination

TABLE 1 "Biological Process" GO terms overrepresented in 1,870 DEGs that differ due to rearing condition among all source populations and months

GO ID	GO Name	FDR adjusted *p*-value
GO:0006022	Aminoglycan metabolic process	3.79E−04
GO:0006030	Chitin metabolic process	9.36E−04
GO:0055114	Oxidation-reduction process	.002
GO:1901071	Glucosamine-containing compound metabolic process	.002
GO:0006040	Amino sugar metabolic process	.003
GO:0006508	Proteolysis	.003
GO:0005975	Carbohydrate metabolic process	.007
GO:0007218	Neuropeptide signalling pathway	.012
GO:0005991	Trehalose metabolic process	.033
GO:0005984	Disaccharide metabolic process	.033
GO:0042133	Neurotransmitter metabolic process	.033
GO:0006720	Isoprenoid metabolic process	.041
GO:0008299	Isoprenoid biosynthetic process	.041
GO:0042391	Regulation of membrane potential	.042
GO:0060078	Regulation of postsynaptic membrane potential	.042
GO:0060079	Excitatory postsynaptic potential	.042
GO:0099565	Chemical synaptic transmission, postsynaptic	.042
GO:0003008	System process	.042
GO:0009123	Nucleoside monophosphate metabolic process	.042
GO:0050877	Nervous system process	.047
GO:0009127	Purine nucleoside monophosphate biosynthetic process	.047
GO:0009168	Purine ribonucleoside monophosphate biosynthetic process	.047
GO:1901605	Alpha-amino acid metabolic process	.049
GO:0044262	Cellular carbohydrate metabolic process	.051
GO:0009124	Nucleoside monophosphate biosynthetic process	.058

TABLE 2 "Biological Process" GO terms overrepresented in 38 DEGs between Chamber and Natural cohorts in fall conditions in December

GO ID	GO Name	FDR adjusted *p*-value
GO:0051282	Regulation of sequestering of calcium ion	.004
GO:0051283	Negative regulation of sequestering of calcium ion	.004
GO:0051208	Sequestering of calcium ion	.004
GO:0051209	Release of sequestered calcium ion into cytosol	.004
GO:0007204	Positive regulation of cytosolic calcium ion concentration	.004
GO:0097553	Calcium ion transmembrane import into cytosol	.004
GO:0051480	Regulation of cytosolic calcium ion concentration	.004
GO:0060402	Calcium ion transport into cytosol	.004
GO:0072503	Cellular divalent inorganic cation homeostasis	.009
GO:0072507	Divalent inorganic cation homeostasis	.009
GO:0006874	Cellular calcium ion homeostasis	.009
GO:0060401	Cytosolic calcium ion transport	.009
GO:0055074	Calcium ion homeostasis	.009

dynamics (i.e., environment vs. internal timing). We propose that in natural conditions, the internal physiological timer exerts ultimate control over diapause termination and environmental signals modulate the timer.

3.4 | Calcium signalling mediates environmental modulation of diapause termination timing

Natural cohort monarchs exit diapause earlier than those of the Chamber cohort, indicating that specific environmental conditions at the overwintering site modulate diapause termination timing. In order to gain insight into what these environmental factors might be, we identified potential functional differences between the Chamber and Natural cohorts in December in fall conditions. A total of 38 DEGs were found in this comparison (Table S2i). Cellular calcium ion regulation is significantly overrepresented among these genes (Table 2). Calcium signalling mediates environmental and stress response in many organisms, and specifically influences cold sensing in insects (Teets, Yi, Lee, & Denlinger, 2013). Chilling advances, and in many cases is required for, diapause termination in many temperate insect species (reviewed in Tauber et al., 1986). Therefore, our data suggest that temperatures in the natural overwintering environment were, at periods, colder than the fall environmental chamber, and this, in turn, led diapause to terminate earlier in the Natural cohort than in the Chamber cohort.

3.5 | JH signalling is a general marker of post-diapause quiescence and a key timer target, but is not itself the timer

Monarchs in diapause are characterized by low haemolymph JH titre, while reproductive monarchs in permissive conditions have high JH titre (Lessman et al., 1989). We found in the Chamber cohort that JH signalling increased between December and January, during the diapause to post-diapause quiescence transition, in fall conditions. Two key genes regulating JH biosynthesis and downstream activity were significantly upregulated in January: juvenile hormone acid O-methyltransferase (*dp-jhamt*; DPOGS214058) and Krüppel homolog 1 (*dp-Kr-h1*; DPOGS211127) (Figure 3b–c; Table S2g). JHAMT is expressed in the corpora allata in insects and has been shown to be the rate-limiting enzyme in JH biosynthesis in *Bombyx mori* (Shinoda & Itoyama, 2003). Kr-h1 is expressed in peripheral tissues and is an early-response transcription factor target and effector of JH signalling (Kayukawa et al., 2012; Minakuchi, Zhou, & Riddiford, 2008). These results show that in January, under fall conditions, JH biosynthesis is upregulated, leading to increased JH activity and *dp-Kr-h1* transcription. Furthermore, of the 38 genes differentially expressed between Natural and Chamber cohorts in fall conditions in December (Figure 3a; Table S2i), seven genes were found in the intersection of the two diapause versus post-diapause quiescence comparisons (December vs. January in Chamber cohort in Table S2l). Among these seven was *dp-jhamt*, whose expression is significantly upregulated in the Natural cohort. This suggests that increased JH biosynthesis is a general marker of post-diapause quiescence and is a key controller of environmental response.

JH promotes diapause termination through 20-HE signalling in some insects (Denlinger, 1979; Zdárek & Denlinger, 1975). Surprisingly, no regulators or downstream transcriptional targets of ecdysteroid signalling were found to be differentially expressed from December to January in the Chamber cohort in fall conditions (Table S2g). The lack of 20-HE-dependent transcriptional changes indicates that JH signalling does not activate ecdysteroid signalling in fall conditions, and further suggests that 20-HE does not control diapause termination timing in fall conditions in monarchs.

Administration of high amounts of JH promotes reproductive development in permissive conditions in diapause monarchs (Herman, 1981). However, development is significantly reduced in the JH-induced diapause monarchs compared to post-diapause monarchs. Herman (1981) hypothesized that this dampened response was a result of reduced effective JH titre. We did not find evidence that JH clearing (via esterases or epoxide hydrolases) or sequestration (via hexamerins) are increased in diapause butterflies compared to nondiapause cohorts (Table S2g,i). This implies that additional mechanisms beyond low JH titre act to suppress response to permissive environmental conditions. Furthermore, it indicates that JH signalling, in and of itself, is not the timer. This leaves two open questions: what additional mechanisms during diapause control response to inductive conditions and what is the primary internal timing mechanism?

The monarch homolog of the *Negative Cofactor 2β* (dp-NC2β; also referred to as "Dr1") transcription factor presents a compelling candidate for a controller of environmental response during diapause and the proximate diapause timer. Although only a modest number of DEGs differentiated Chamber and Natural cohorts in fall conditions in December (38 DEGs; Figure 3a), the Chamber cohort transcriptome showed substantially greater change in summer conditions than did the Natural cohort (2,274 vs. 1,102 DEGs; Figure 3a). This suggests that one or more of the 38 DEGs may be a factor with broad control of transcription. We found that *dp-nc2β* (DPOGS214145) was significantly upregulated in the Chamber cohort compared to the Natural cohort in fall conditions in December (Table S2i). *dp-nc2β* was also upregulated in December compared to January in the Chamber cohort in summer conditions (Table S2j). NC2β is a conserved protein that contributes to histone acetyltransferase activity and activates downstream promoter element binding motifs and represses TATA promoters (Willy, Kobayashi, & Kadonaga, 2000). It influences global transcriptional dynamics in response to heat or stress in yeast, *Drosophila*, and humans (de Graaf et al., 2010; Honjo, Mauthner, Wang, Skene, & Tracey, 2016; Spedale et al., 2011; Wang, Faiola, Xu, Pan, & Martinez, 2008). These results are consistent with the hypothesis that *dp-nc2β* expression renders monarchs insensitive to inductive summer conditions and may also directly control diapause termination timing via histone acetylation.

3.6 | Diapause state controls environmental sensitivity through regulating hormonal and neuroendocrine pathways

Genes that differ in expression in the inductive summer environmental conditions as a consequence of diapause state are the effective targets of diapause. To identify these genetic pathways that mediate environmental response with respect to diapause state, we identified DEGs between December and January Chamber cohorts in summer conditions. Hundreds of genes are differentially expressed between December and January in the Chamber cohort in summer conditions (n = 268 DEGs; Table S2j). Redox chemistry functions differed significantly between these months as evidenced by the single GO term (GO:0055114, oxidation-reduction process) to reach the

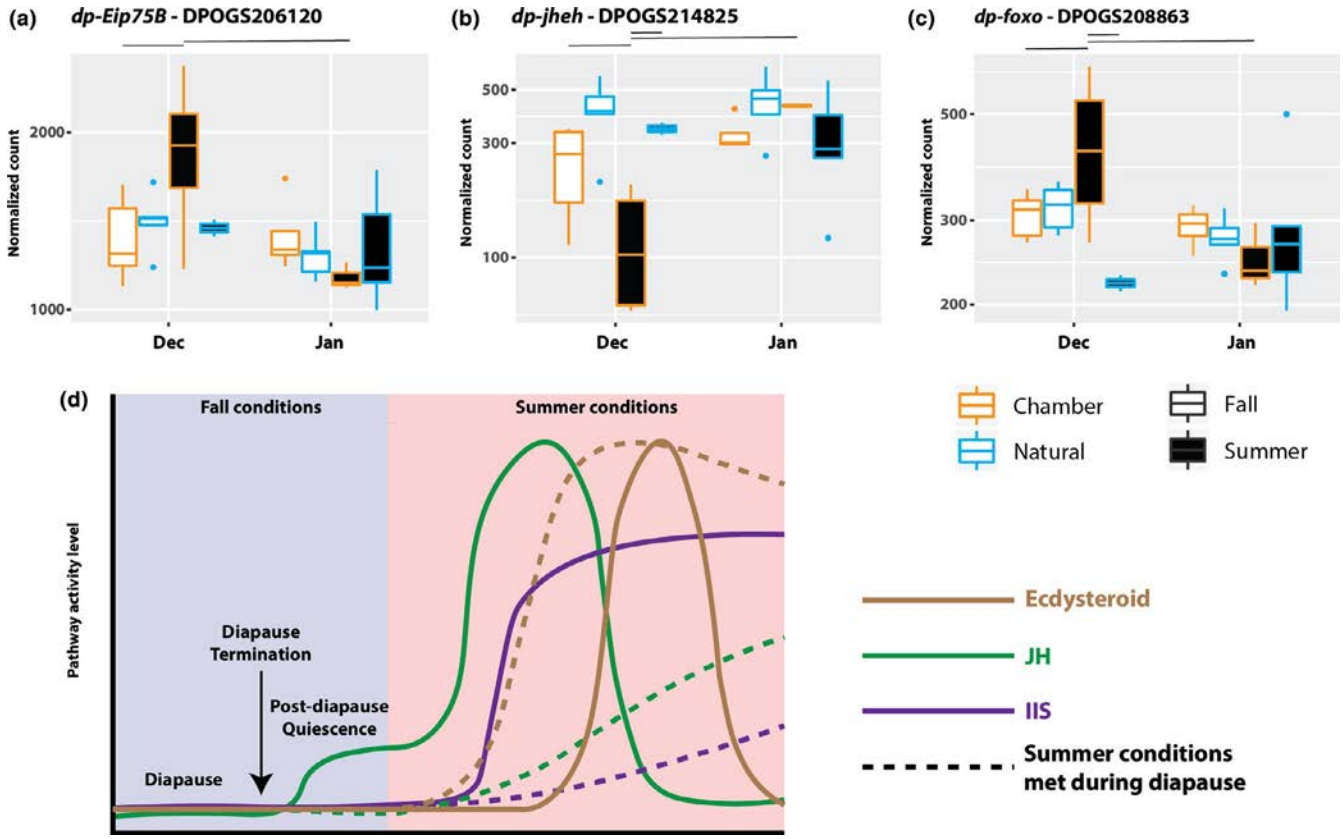


FIGURE 4 Hormonal gene expression is targeted by diapause programme. Normalized counts of (a) *dp-e75* transcription factor (DPOGS206120), (b) *dp-jheh* homolog (DPOGS214825), and (c) *dp-foxo* (DPOGS208863) in ecdysteroid, JH and IIS pathways, respectively. Lines under gene names indicate pairwise comparisons for which normalized counts significantly differ (FDR adjusted p-value < .05). (d) Putative model of ecdysteroid (brown), JH (green), and IIS (purple) pathway activity in fall conditions (blue background) and summer conditions (red background). Solid lines indicate pathway activity under normal diapause development in which case individuals experience summer conditions after having transitioned to post-diapause quiescence. Dashed lines indicate pathway activity in the case when individuals are still in diapause upon experiencing summer conditions. Diapause development labels apply to solid lines, only. See main text (Discussion) for detailed explanation of model interpretation [Colour figure can be viewed at wileyonlinelibrary.com]

significance threshold (FDR = 0.052). These results suggest that energy metabolism operates differently depending on diapause state at the time that summer conditions are entered.

Hormonal pathways previously associated with insect diapause were changed. Ecdysteroid signalling, which involves a cascade of oxidation-reduction reactions mediated by cytochrome P450 enzymes, plays a role in differential response. Several components related to ecdysone and 20-HE metabolism and transcriptional regulation are differentially expressed (Table S2j; *dp-Eip75B* (DPOGS206120) shown as example in Figure 4a), demonstrating that diapause alters ecdysteroid response in summer conditions. A strongly expressed JH epoxide hydrolase (*dp-jheh*) isoform (DPOGS214825), which degrades JH, shows significantly low expression in December in the Chamber cohort in summer conditions (Figure 4b). This result suggests that diapause alters JH clearance dynamics in summer conditions. Finally, the transcription factor *dp-foxo* (DPOGS208863), a central effector and negative transcriptional regulator of the IIS pathway, is expressed at significantly high levels in the Chamber cohort in December compared to January. In fact, the Chamber cohort in December presents the single exception where *dp-foxo* expression in summer conditions remains high, mirroring relative reproductive development among groups (Figure 4c). This suggests that diapause causes repression of IIS activity upon experiencing summer conditions. It appears that diapause forces specific regulation of these pathways in order to control environmental response.

4 | DISCUSSION

4.1 | Diapause termination as a model of long-term timekeeping

Diapause has been studied for over a century in scores of species across class Insecta (Andrewartha, 1952) not least of which because it exemplifies fundamental properties of biological systems including the recording and storage of historical information over long, specific time periods. Although resolution of both of these properties requires elucidation of a diapause termination mechanism, termination remains poorly understood. Lehmann et al. (2018) outlined particular challenges of studying termination that are primarily a result of the complex interaction of environment and internal physiological control. By studying a combination of different source populations in different environmental chamber conditions over time, we were able to modularize individual components of termination and identify molecular controllers of specific termination processes. While the focus of this work was on head transcription in order to understand neurohormonal controllers of diapause timing, other tissues, including the reproductive organs and abdominal fat body, are associated with the diapause response. In subsequent work it will be interesting to investigate how these organs communicate with the diapause timer to coordinate overall diapause response.

Our data suggest that epigenetic mechanisms play a critical role in the recording and storage properties of biological timing. Epigenetic regulation of diapause development in other insects has previously been hypothesized (reviewed in Reynolds, 2017). Moreover, seasonal photoperiodic timing is controlled epigenetically through histone modification in plants and mammals. In a classic example of seasonal timing, histone modifications at FLOWERING LOCUS C control vernalization and flowering time in *Arabidopsis thaliana* (reviewed in Whittaker & Dean, 2017). Histone methylation controls hormonal regulation of a winter refractory period in Siberian hamsters (Stevenson & Prendergast, 2013). Interestingly, both of these cases involve timing of cold-mediated refractory periods such as in monarch adult reproductive diapause. Validation of this mechanism in monarchs would add to the mounting evidence that epigenetics is a common way across diverse biological systems to achieve long-term timekeeping (e.g., Stevenson & Prendergast, 2013; Whittaker & Dean, 2017).

Although we found that canonical neurohormonal (ecdysteroid, JH, and IIS) pathways are important for mediating environment-dependent response, our work suggests that histone modification and JH signalling, specifically, work in concert to control the specific timing of diapause dynamics. Interestingly, in the few cases in which epigenetic regulation of JH signalling has been studied, histone acetylation, in particular, has been shown to control transcription of JH-response genes (Roy & Palli, 2018; Xu, Roy, & Palli, 2018). We hypothesize that JH signalling, via *dp-jhamt*, is one of several direct transcriptional targets of *dp-nc2β* that are required for a full termination of diapause.

Furthermore, we show that environmental conditions (we predict cold temperature in this case) modulate the timer and propose that this may occur through cryoprotectant accumulation. Accumulation of small molecule cryoprotectants (e.g., sorbitol, alanine, etc) in response to cold exposure has been observed in diapause and implicated in diapause termination in other insects (e.g., Leal et al., 2018; Lehmann et al., 2018; Michaud & Denlinger, 2007; Wang, Egi, Takeda, Oishi, & Sakamoto, 2014). Calcium signalling mediates osmotic stress due to small molecule cryoprotectant accumulation in insects (Storey, 1990). Interestingly, Ca^{2+} stimulates JH III production from the corpora allata in in vitro culture in *Gryllus bimaculatus* (Woodring & Hoffman, 1994). This leads us to speculate that cold hastens monarch diapause through promoting accumulation of small molecule cryoprotectants that increase calcium signalling, and in turn prompts JH biosynthesis, whereby promoting diapause termination. It is possible that this mechanism may operate more generally across insects to control diapause dynamics. While well documented that cold mediates epigenetic regulation of transcriptional states in plants (reviewed in Banerjee, Wani, & Roychoudhury, 2017), this relationship is not well established in animals. It will be interesting to determine if in monarchs cold-mediated effects on diapause through cryoprotectant accumulation and calcium signalling, and effects due to epigenetic regulation, are integrated or independent responses.

Altogether, this work illuminates molecular mechanisms underlying diapause termination and provides important insights into the mechanisms of broader biological properties represented by the termination model. We believe this represents one of the first accessible models to study the molecular basis of long-term photoperiod-independent timekeeping, which is pervasive (e.g., obligate, low-latitude, or underground diapauses, Denlinger, Hahn, Merlin, Holzapfel, & Bradshaw, 2017).

4.2 | Model for the evolution of hormonal pathway interactions

The defining feature of diapause is that it fundamentally changes organismal response to environment, rendering individuals insensitive to normally permissive environmental cues. Much previous work has been dedicated to understanding the molecular mechanisms behind the maintenance of environmental insensitivity (e.g., Lehmann et al., 2016), or rather how gene expression differs between diapause and either nondiapause or post-diapause quiescence states (e.g., Poelchau, Reynolds, Elsik, Denlinger, & Armbruster, 2013; Ragland & Keep, 2017). Here we are able to address an equally important but open problem: through what genetic mechanisms is diapause acting to change response to environmental conditions? Our evidence points to canonical hormonal pathways (ecdysteroid, JH, and insulin) as playing significant roles. These pathways mediate environment-dependent growth, development, and reproduction in many different cases such as seasonal polyphenism in *Bicyclus anynana* (Koch, Brakefield, & Kesbeke, 1996), colouration in *M. sexta* (Suzuki & Nijhout, 2006), and reproductive division of labour in ants (Chandra et al., 2018).

Importantly, our approach allows us to make hypotheses about how regulation of and interaction between these pathways changes as a result of diapause (Figure 4d). We propose that the critical change that controls subsequent pathway dynamics is reduced JH signalling in fall conditions (*dp-jhamt*; Figure 3b). Given this change, our data suggest key hypotheses about monarchs that experience summer conditions while still in diapause. First, JH pathway activity increases slowly in summer conditions (*dp-jhamt*; Figure 3b) while being less efficiently cleared (*dp-jheh*; Figure 4b). Second, while 20-HE is generally expressed as a pulse, ecdysteroid signalling activity in summer conditions rises earlier due to initially reduced and only slowly increasing JH levels. Ecdysteroid signalling remains high after extended time in summer conditions because less efficient clearance of JH leads to less efficient clearance of 20-HE, hence an elevated *dp-Eip75B* signal (Figure 4a). JH clearance has been shown to be necessary for full ecdysone response in some cases (e.g., in *G. bimaculatus*, Espig & Hoffmann, 1985). Third, IIS activity remains low in summer conditions due to increased *dp-foxo* transcription (Figure 4c). Previous work exploring the relationship between JH signalling and IIS in other insects provides clues into how JH may control IIS dynamics. JH promotes IIS in *Tribolium castaneum* (Xu, Sheng, & Palli, 2013), which may provide a mechanism by which IIS is reduced via *dp-foxo*. FOXO has been shown to regulate JH degradation in *B. mori* (Zeng et al., 2017), which suggests an additional, orthologous mechanism by which JH accumulates slowly in summer conditions when met during diapause. This model highlights how extensive crosstalk between these developmental hormones can render this integrated network vulnerable to a single perturbation, leading to a completely reorganized hormonal landscape that changes organismal function.

These hormonal pathways have been repeatedly coopted to generate a diversity of phenotypes by controlling major life history traits and coordinating them with environment (Finch & Rose, 1995; Flatt, Tu, & Tatar, 2005). Monarch diapause presents a compelling model to study the evolution of regulatory connections among these pathways. Global monarch populations show a range of variation in seasonally-controlled traits, presenting an attractive model for pursuing evolutionary studies. Some traits show a certain degree of conservation among populations (Freedman et al., 2018). In contrast, Australian monarchs do not show a full diapause but rather undergo a seasonally-controlled oligopause (James, 1982; James & Hales, 1983).

4.3 | Diapause timing and monarch overwintering dynamics

This work highlights the importance of the overwintering period for monarch biology. Diapause timing has important implications for monarch survival (Herman & Tatar, 2001) and mating at overwintering sites. Interestingly, according to Herman (1981), male monarch diapause lasts for shorter duration (September–November) than does female diapause (September–December), potentially suggesting that diapause timing in females may have a stronger impact on population dynamics. In subsequent studies, it will be interesting to determine if similar or different molecular mechanisms control male monarch diapause timing and determine what mechanisms lead to sex differences in diapause timing.

Environmental sensitivity of diapause dynamics means that monarchs will act as an important sentinel species for monitoring environmental change and disturbance at overwintering sites. Spatial structure and temporal dynamics of the overwintering colonies change within and across years (Vidal & Rendón-Salinas, 2014). It is possible that these movements at least partially reflect a strategy to optimize diapause timing. Changes in diapause dynamics may be considered alongside ecological modelling (e.g., Oberhauser & Peterson, 2003) to understand distribution at overwintering sites. Indeed, understanding how diapause dynamics are affected by environmental and anthropogenic factors at their overwintering sites may be critical for understanding North American monarch population decline (Agrawal & Inamine, 2018) and guiding future conservation efforts, a point highlighted by the record low number of monarchs recorded in the western North American monarch population in 2018 (The Xerxes Society Western Monarch Thanksgiving Count).

ACKNOWLEDGEMENTS

We thank Francis Villablanca (California Polytechnic State University, San Luis Obispo) for help with butterfly collection as well as valuable discussion and experience at the overwintering site. We are grateful to the University of Chicago Genomics Core Facility for expert sequencing services. This work was supported by an NSF PRFB award 1523759 (to D.A.G.), US Fish & Wildlife award F17AC01222 (to M.R.K.), NSF grant IOS-1452648 (to M.R.K.), and NIH grant GM108626 (to M.R.K.). Finally, we thank three anonymous reviewers for thoughtful comments on improving this manuscript.

AUTHOR CONTRIBUTIONS

D.A.G. conceived the project, D.A.G. and M.R.K. designed the analysis, D.A.G. collected and analyzed the data, and D.A.G. wrote the manuscript with input from MRK.

DATA AVAILABILITY STATEMENT

All raw sequence reads from this study have been uploaded to NCBI Sequence Read Archive to BioProjectID PRJNA548105 and mature oocyte count data uploaded to Dryad (https://doi.org/10.5061/dryad.fd517kq).

REFERENCES

Agrawal, A. A., & Inamine, H. (2018). Mechanisms behind the monarch's decline. *Science*, *360*, 1294–1296. https://doi.org/10.1126/science.aat5066

Andrewartha, H. G. (1952). Diapause in relation to the ecology of insects. *Biological Reviews*, *27*, 50–107. https://doi.org/10.1111/j.1469-185X.1952.tb01363.x

Banerjee, A., Wani, S. H., & Roychoudhury, A. (2017). Epigenetic control of plant cold responses. *Frontiers in Plant Science*, *8*, 1643. https://doi.org/10.3389/fpls.2017.01643

Barker, J. F., & Herman, W. S. (1973). On the neuroendocrine control of ovarian development in the Monarch butterfly. *Journal of Experimental Zoology*, *183*, 1–9. https://doi.org/10.1002/jez.1401830102

Chandra, V., Fetter-Pruneda, I., Oxley, P. R., Ritger, A. L., McKenzie, S. K., Libbrecht, R., & Kronauer, D. J. C. (2018). Social regulation of insulin signaling and the evolution of eusociality in ants. *Science*, *361*, 398–402. https://doi.org/10.1126/science.aar5723

Conesa, A., & Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, *2008*, 1–12. https://doi.org/10.1155/2008/619832

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Dallmann, S. H., & Herman, W. S. (1978). Hormonal regulation of hemolymph lipid concentration in the monarch butterfly, *Danaus plexippus*. *General and Comparative Endocrinology*, *36*, 142–150. https://doi.org/10.1016/0016-6480(78)90059-X

de Graaf, P., Mousson, F., Geverts, B., Scheer, E., Tora, L., Houtsmuller, A. B., & Timmers, M. (2010). Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. *Journal of Cell Science*, *123*, 2663–2671. https://doi.org/10.1242/jcs.064097

Denlinger, D. L. (1979). Pupal diapause in tropical flesh flies: Environmental and endocrine regulation, metabolic rate and genetic selection. *Biological Bulletin*, *156*, 31–46. https://doi.org/10.2307/1541001

Denlinger, D. L. (2002). Regulation of diapause. *Annual Review of Entomology*, *47*, 93–122. https://doi.org/10.1146/annurev.ento.47.091201.145137

Denlinger, D. L., Hahn, D. A., Merlin, C., Holzapfel, C. M., & Bradshaw, W. E. (2017). Keeping time without a spine: What can the insect clock teach us about seasonal adaptation? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *372*, 20160257–20160259.

Denlinger, D. L., Yocum, G. D., & Rinehart, J. P. (2011). Hormonal control of diapause. In L. I. Gilbert (Ed.), *Insect Endocrinology* (pp. 430–463). Amsterdam, the Netherlands: Elsevier.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2012). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Espig, W., & Hoffmann, K. H. (1985). Juvenile hormone and reproduction in the cricket. II. Effect of rearing temperature on corpus allatum activity (in vitro) in adult females. *Experientia*, *41*, 758–759. https://doi.org/10.1007/BF02012583

Finch, C. E., & Rose, M. R. (1995). Hormones and the physiological architecture of life history evolution. *The Quarterly Review of Biology*, *70*, 1–52. https://doi.org/10.1086/418864

Flatt, T., Tu, M.-P., & Tatar, M. (2005). Hormonal pleiotropy and the juvenile hormone regulation of Drosophila development and life history. *BioEssays*, *27*, 999–1010. https://doi.org/10.1002/bies.20290

Freedman, M. G., Dingle, H., Tabuloc, C. A., Chiu, J. C., Yang, L. H., & Zalucki, M. P. (2018). Non-migratory monarch butterflies, *Danaus plexippus* (L.), retain developmental plasticity and a navigational mechanism associated with migration. *Biological Journal of the Linnean Society*, *123*, 265–278.

Gegear, R. J., Casselman, A., Waddell, S., & Reppert, S. M. (2008). Cryptochrome mediates light-dependent magnetosensitivity in Drosophila. *Nature*, *454*, 1014–1018. https://doi.org/10.1038/nature07183

Goehring, L., & Oberhauser, K. S. (2002). Effects of photoperiod, temperature, and host plant age on induction of reproductive diapause and development time in *Danaus plexippus*. *Ecological Entomology*, *27*, 674–685. https://doi.org/10.1046/j.1365-2311.2002.00454.x

Goehring, L., & Oberhauser, K. S. (2004). In M. J. Solensky, & K. S. Oberhauser (Eds.), *The monarch butterfly ecology and conservation* (pp. 187–196).

Götz, S., Arnold, R., Sebastián-León, P., Martín-Rodríguez, S., Tischler, P., Jehl, M.-A., … Conesa, A. (2011). B2G-FAR, a species-centered GO annotation repository. *Bioinformatics*, *27*, 919–924. https://doi.org/10.1093/bioinformatics/btr059

Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., … Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, *36*, 3420–3435. https://doi.org/10.1093/nar/gkn176

Guerra, P. A., Gegear, R. J., & Reppert, S. M. (2014). A magnetic compass aids monarch butterfly migration. *Nature Communications*, *5*, 4164. https://doi.org/10.1038/ncomms5164

Guerra, P. A., & Reppert, S. M. (2013). Coldness triggers northward flight in remigrant monarch butterflies. *Current Biology*, *23*, 419–423. https://doi.org/10.1016/j.cub.2013.01.052

Hand, S. C., Denlinger, D. L., Podrabsky, J. E., & Roy, R. (2016). Mechanisms of animal diapause: Recent developments from nematodes, crustaceans, insects, and fish. *American Journal of Physiology – Regulatory, Integrative and Comparative Physiology*, *310*, R1193–R1211. https://doi.org/10.1152/ajpregu.00250.2015

Herman, W. S. (1973). The endocrine basis of reproductive inactivity in monarch butterflies overwintering in central California. *Journal of Insect Physiology*, *19*, 1883–1887. https://doi.org/10.1016/0022-1910(73)90056-5

Herman, W. S. (1975). Endocrine regulation of posteclosion enlargement of the male and female reproductive glands in Monarch butterflies. *General and Comparative Endocrinology*, *26*, 534–540. https://doi.org/10.1016/0016-6480(75)90176-8

Herman, W. S. (1981). Studies on the adult reproductive diapause of the monarch butterfly, *Danaus plexippus*. *The Biological Bulletin*, *160*, 89–106.

Herman, W. S., Brower, L. P., & Calvert, W. H. (1989). Reproductive tract development in monarch butterflies overwintering in California and Mexico. *Journal of the Lepidopterists' Society*, *43*, 50–58.

Herman, W. S., Lessman, C. A., & Johnson, G. D. (1981). Correlation of juvenile hormone titer changes with reproductive tract development in the posteclosion monarch butterfly. *Journal of Experimental Zoology*, *218*, 387–395. https://doi.org/10.1002/jez.1402180310

Herman, W. S., & Tatar, M. (2001). Juvenile hormone regulation of longevity in the migratory monarch butterfly. *Proceedings of the Royal Society B: Biological Sciences*, *268*, 2509–2514. https://doi.org/10.1098/rspb.2001.1765

Hodek, I. (2002). Controversial aspects of diapause development. *European Journal of Entomology*, *99*, 163–173. https://doi.org/10.14411/eje.2002.024

Hodek, I. (2012). Adult diapause in coleoptera. *Psyche: A Journal of Entomology*, *2012*, 1–10.

Honjo, K., Mauthner, S. E., Wang, Y., Skene, J. H. P., & Tracey, W. D. Jr (2016). Nociceptor-enriched genes required for normal thermal nociception. *Cell Reports*, *16*, 295–303. https://doi.org/10.1016/j.celrep.2016.06.003

James, D. G. (1982). Ovarian dormancy in *Danaus plexippus* (L.) (Lepidoptera: Nymphalidae)—oligopause not diapause. *Australian Journal of Entomology*, *21*, 31–35. https://doi.org/10.1111/j.1440-6055.1982.tb01761.x

James, D. G., & Hales, D. F. (1983). Sensitivity to juvenile hormone is not reduced in clustering monarch butterflies, *Danaus plexippus*, in Australia. *Physiological Entomology*, *8*, 273–276. https://doi.org/10.1111/j.1365-3032.1983.tb00359.x

Kayukawa, T., Minakuchi, C., Namiki, T., Togawa, T., Yoshivama, M., Kamimura, M., … Shinoda, T. (2012). Transcriptional regulation of juvenile hormone-mediated induction of Krüppel homolog 1, a repressor of insect metamorphosis. *Proceedings of the National Academy of Sciences of the USA*, *109*, 11729–11734.

Koch, P. B., Brakefield, P. M., & Kesbeke, F. (1996). Ecdysteroids control eyespot size and wing color pattern in the polyphenic butterfly *Bicyclus anynana* (Lepidoptera: Satyridae). *Journal of Insect Physiology*, *42*, 223–230. https://doi.org/10.1016/0022-1910(95)00103-4

Koštál, V. (2006). Eco-physiological phases of insect diapause. *Journal of Insect Physiology*, *52*, 113–127. https://doi.org/10.1016/j.jinsphys.2005.09.008

Koštál, V., Štětina, T., Poupardin, R., Korbelová, J., & Bruce, A. W. (2017). Conceptual framework of the eco-physiological phases of insect diapause development justified by transcriptomic profiling. *Proceedings of the National Academy of Sciences of the USA*, *1*, 201707281–201707313. https://doi.org/10.1073/pnas.1707281114

Leal, L., Talla, V., Källman, T., Friberg, M., Wiklund, C., Dincă, V., … Backström, N. (2018). Gene expression profiling across ontogenetic stages in the wood white (*Leptidea sinapis*) reveals pathways linked to butterfly diapause regulation. *Molecular Ecology*, *27*, 935–948.

Lehmann, P., Pruisscher, P., Koštál, V., Moos, M., Šimek, P., Nylin, S., … Gotthard, K. (2018). Metabolome dynamics of diapause in the butterfly *Pieris napi*: Distinguishing maintenance, termination and post-diapause phases. *The Journal of Experimental Biology*, *221*, jeb169508.

Lehmann, P., Pruisscher, P., Posledovich, D., Carlsson, M., Käkelä, R., Tang, P., … Gotthard, K. (2016). Energy and lipid metabolism during direct and diapause development in a pierid butterfly. *Journal of Experimental Biology*, *219*, 3049–3060. https://doi.org/10.1242/jeb.142687

Lessman, C. A., Herman, W. S., Schooley, D. A., Tsai, L. W., Bergot, B. J., & Baker, F. C. (1989). Detection of juvenile hormone I, II and III in adult monarch butterflies (*Danaus plexippus*). *Insect Biochemistry*, *19*, 431–433. https://doi.org/10.1016/0020-1790(89)90049-8

Liedvogel, M., & Lundberg, M. (2014). The genetics of migration. In L. A. Hansson, & S. Åkesson (Eds.), *Animal movement across scales* (pp. 219–231). Oxford, UK: Oxford University Press.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8

Marini, F., & Binder, H. (2017). Development of applications for interactive and reproducible research: A case study. *Genomics and Computational Biology*, *3*, e39.

Merlin, C., Gegear, R. J., & Reppert, S. M. (2009). Antennal circadian clocks coordinate sun compass orientation in migratory monarch butterflies. *Science*, *325*, 1700–1704. https://doi.org/10.1126/science.1176221

Michaud, M. R., & Denlinger, D. L. (2007). Shifts in the carbohydrate, polyol, and amino acid pools during rapid cold-hardening and diapause-associated cold-hardening in flesh flies (*Sarcophaga crassipalpis*): A metabolomic comparison. *Journal of Comparative Physiology B*, *177*, 753–763. https://doi.org/10.1007/s00360-007-0172-5

Minakuchi, C., Zhou, X., & Riddiford, L. M. (2008). Krüppel homolog 1 (Kr-h1) mediates juvenile hormone action during metamorphosis of *Drosophila melanogaster*. *Mechanisms of Development*, *125*, 91–105. https://doi.org/10.1016/j.mod.2007.10.002

Mirth, C. K., Tang, H. Y., Makohon-Moore, S. C., Salhadar, S., Gokhale, R. H., Warner, R. D., … Shingleton, A. W. (2014). Juvenile hormone regulates body size and perturbs insulin signaling in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA*, *111*, 7018–7023. https://doi.org/10.1073/pnas.1313058111

Morlan, J. D., Qu, K., & Sinicropi, D. V. (2012). Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS ONE*, *7*, e42882. https://doi.org/10.1371/journal.pone.0042882

Nijhout, H. F., & Williams, C. M. (1974). Control of moulting and metamorphosis in the tobacco hornworm, *Manduca sexta* (L.): Cessation of juvenile hormone secretion as a trigger for pupation. *Journal of Experimental Biology*, *61*, 493–501.

Oberhauser, K., & Peterson, A. T. (2003). Modeling current and future potential wintering distributions of eastern North American monarch butterflies. *Proceedings of the National Academy of Sciences of the USA*, *100*, 14063–14068. https://doi.org/10.1073/pnas.2331584100

Okuda, T., & Chinzei, Y. (1988). Vitellogenesis in a lady-beetle, *Coccinella septempunctata* in relation to the aestivation-diapause. *Journal of Insect Physiology*, *34*, 393–401. https://doi.org/10.1016/0022-1910(88)90109-6

Pan, M. L., & Wyatt, G. R. (1976). Control of vitellogenin synthesis in the Monarch butterfly by juvenile hormone. *Developmental Biology*, *54*, 127–134. https://doi.org/10.1016/0012-1606(76)90291-8

Poelchau, M. F., Reynolds, J. A., Elsik, C. G., Denlinger, D. L., & Armbruster, P. A. (2013). RNA-Seq reveals early distinctions and late convergence of gene expression between diapause and quiescence in the Asian tiger mosquito, *Aedes albopictus*. *Journal of Experimental Biology*, *216*, 4082–4090. https://doi.org/10.1242/jeb.089508

R Core Team (2018). *R: A language and environment for statistical computing*.

Ragland, G. J., Denlinger, D. L., & Hahn, D. A. (2010). Mechanisms of suspended animation are revealed by transcript profiling of diapause in the flesh fly. *Proceedings of the National Academy of Sciences of the USA*, *107*, 14909–14914. https://doi.org/10.1073/pnas.1007075107

Ragland, G. J., & Keep, E. (2017). Comparative transcriptomics support evolutionary convergence of diapause responses across Insecta. *Physiological Entomology*, *42*, 246–256. https://doi.org/10.1111/phen.12193

Reynolds, J. A. (2017). Epigenetic influences on diapause. In H. Verlinden (Ed.), *Advances in insect physiology*, Vol. *24* (pp. 115–144). London, UK: Elsevier.

Riddiford, L. M., Truman, J. W., Mirth, C. K., & Shen, Y. C. (2010). A role for juvenile hormone in the prepupal development of *Drosophila melanogaster*. *Development*, *137*, 1117–1126. https://doi.org/10.1242/dev.037218

Roy, A., & Palli, S. R. (2018). Epigenetic modifications acetylation and deacetylation play important roles in juvenile hormone action. *BMC Genomics*, *19*, 1–15. https://doi.org/10.1186/s12864-018-5323-4

Roy, S., Saha, T. T., Zou, Z., & Raikhel, A. S. (2018). Regulatory pathways controlling female insect reproduction. *Annual Review of Entomology*, *63*, 489–511. https://doi.org/10.1146/annurev-ento-020117-043258

Shinoda, T., & Itoyama, K. (2003). Juvenile hormone acid methyltransferase: A key regulatory enzyme for insect metamorphosis. *Proceedings of the National Academy of Sciences of the USA*, *100*, 11986–11991. https://doi.org/10.1073/pnas.2134232100

Sim, C., & Denlinger, D. L. (2008). Insulin signaling and FOXO regulate the overwintering diapause of the mosquito *Culex pipiens*. *Proceedings of the National Academy of Sciences of the USA*, *105*, 6777–6781. https://doi.org/10.1073/pnas.0802067105

Spedale, G., Meddens, C. A., Koster, M. J. E., Ko, C. W., van Hooff, S. R., Holstege, F. C. P., … Pijnappel, W. W. M. P. (2011). Tight cooperation between Mot1p and NC2β in regulating genome-wide transcription, repression of transcription following heat shock induction and genetic interaction with SAGA. *Nucleic Acids Research*, *40*, 996–1008. https://doi.org/10.1093/nar/gkr784

Stevenson, T. J., & Prendergast, B. J. (2013). Reversible DNA methylation regulates seasonal photoperiodic time measurement. *Proceedings of the National Academy of Sciences of the USA*, *110*, 16651–16656. https://doi.org/10.1073/pnas.1310643110

Storey, K. B. (1990). Biochemical Adaptation for Cold Hardiness in Insects. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *326*, 635–654.

Suzuki, Y., & Nijhout, H. F. (2006). Evolution of a polyphenism by genetic accommodation. *Science*, *311*, 650–652. https://doi.org/10.1126/science.1118888

Tauber, M. J., Tauber, C. A., & Masaki, S. (1986). *Seasonal adaptations of insects*. New York, NY: Oxford University Press.

Teets, N. M., Yi, S.-X., Lee, R. E., & Denlinger, D. L. (2013). Calcium signaling mediates cold sensing in insect tissues. *Proceedings of the National Academy of Sciences of the USA*, *110*, 9154–9159. https://doi.org/10.1073/pnas.1306705110

Vidal, O., & Rendón-Salinas, E. (2014). Dynamics and trends of overwintering colonies of the monarch butterfly in Mexico. *Biological Conservation*, *180*, 165–175.

Wang, Q., Egi, Y., Takeda, M., Oishi, K., & Sakamoto, K. (2014). Melatonin pathway transmits information to terminate pupal diapause in the Chinese oak silkmoth *Antheraea pernyiand* through reciprocated inhibition of dopamine pathway functions as a photoperiodic counter. *Entomological Science*, *18*, 74–84.

Wang, Y.-L., Faiola, F., Xu, M., Pan, S., & Martinez, E. (2008). Human ATAC Is a GCN5/PCAF-containing acetylase complex with a novel NC2-like histone fold module that interacts with the TATA-binding protein. *Journal of Biological Chemistry*, *283*, 33808–33815. https://doi.org/10.1074/jbc.M806936200

Whittaker, C., & Dean, C. (2017). The FLCLocus: A platform for discoveries in epigenetics and adaptation. *Annual Review of Cell and Developmental Biology*, *33*, 555–575.

Williams, K. D., Busto, M., Suster, M. L., So, A. K., Ben-Shahar, Y., Leevers, S. J., & Sokolowski, M. B. (2006). Natural variation in *Drosophila melanogaster* diapause due to the insulin-regulated PI3-kinase. *Proceedings of the National Academy of Sciences of the USA*, *103*, 15911–15915.

Willy, P. J., Kobayashi, R., & Kadonaga, J. T. (2000). A basal transcription factor that activates or represses transcription. *Science*, *290*, 982–985. https://doi.org/10.1126/science.290.5493.982

Woodring, J., & Hoffman, K. (1994). The effects of octopamine, dopamine and serotonin on juvenile hormone synthesis, in vitro, in the cricket, *Gryllus bimaculatus*. *Journal of Insect Physiology*, *40*, 797–802. https://doi.org/10.1016/0022-1910(94)90009-4

Xu, J., Roy, A., & Palli, S. R. (2018). CREB-binding protein plays key roles in juvenile hormone action in the red flour beetle, *Tribolium castaneum*. *Scientific Reports*, *8*, 11855.

Xu, J., Sheng, Z., & Palli, S. R. (2013). Juvenile hormone and insulin regulate trehalose homeostasis in the red flour beetle, *Tribolium castaneum*. *PLOS Genetics*, *9*, e1003535. https://doi.org/10.1371/journal.pgen.1003535

Zdárek, J., & Denlinger, D. L. (1975). Action of ecdysoids, juvenoids, and non-hormonal agents on termination of pupal diapause in the flesh fly. *Journal of Insect Physiology*, *21*, 1193–1202. https://doi.org/10.1016/0022-1910(75)90087-6

Zeng, B., Huang, Y., Xu, J., Shiotsuki, T., Bai, H., Palli, S. R., … Tan, A. (2017). The FOXO transcription factor controls insect growth and development by regulating juvenile hormone degradation in the silkworm, *Bombyx mori*. *Journal of Biological Chemistry*, *292*, 11659–11669.

Zhan, S., & Reppert, S. M. (2012). MonarchBase: The monarch butterfly genome database. *Nucleic Acids Research*, *41*, D758–D763. https://doi.org/10.1093/nar/gks1057

Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D. N., & Perous, C. M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*, *15*, 419. https://doi.org/10.1186/1471-2164-15-419

Zhu, H., Sauman, I., Yuan, Q., Casselman, A., Emery-Lee, M., Emery, P., & Reppert, S. M. (2008). Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. *PLOS Biology*, *6*, e4. https://doi.org/10.1371/journal.pbio.0060004

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Green DA II, Kronforst MR. Monarch butterflies use an environmentally sensitive, internal timer to control overwintering dynamics. *Mol Ecol*. 2019;28:3642–3655. https://doi.org/10.1111/mec.15178

Epigenetics and the Evolution of Darwin's Finches

Michael K. Skinner[1,*], Carlos Gurerrero-Bosagna[1,3], M. Muksitul Haque[1], Eric E. Nilsson[1], Jennifer A.H. Koop[2,4], Sarah A. Knutie[2], and Dale H. Clayton[2]

[1]Center for Reproductive Biology, School of Biological Sciences, Washington State University

[2]Department of Biology, University of Utah

[3]Present address: Department of Physics, Biology and Chemistry (IFM), Linköping University, Sweden

[4]Present address: Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ

*Corresponding author: E-mail: skinner@wsu.edu.

Accepted: July 18, 2014

Data deposition: All DMR and CNV genomic data obtained in this study have been deposited in the NCBI public GEO database under the accession (GEO #: GSE58334).

Abstract

The prevailing theory for the molecular basis of evolution involves genetic mutations that ultimately generate the heritable phenotypic variation on which natural selection acts. However, epigenetic transgenerational inheritance of phenotypic variation may also play an important role in evolutionary change. A growing number of studies have demonstrated the presence of epigenetic inheritance in a variety of different organisms that can persist for hundreds of generations. The possibility that epigenetic changes can accumulate over longer periods of evolutionary time has seldom been tested empirically. This study was designed to compare epigenetic changes among several closely related species of Darwin's finches, a well-known example of adaptive radiation. Erythrocyte DNA was obtained from five species of sympatric Darwin's finches that vary in phylogenetic relatedness. Genome-wide alterations in genetic mutations using copy number variation (CNV) were compared with epigenetic alterations associated with differential DNA methylation regions (epimutations). Epimutations were more common than genetic CNV mutations among the five species; furthermore, the number of epimutations increased monotonically with phylogenetic distance. Interestingly, the number of genetic CNV mutations did not consistently increase with phylogenetic distance. The number, chromosomal locations, regional clustering, and lack of overlap of epimutations and genetic mutations suggest that epigenetic changes are distinct and that they correlate with the evolutionary history of Darwin's finches. The potential functional significance of the epimutations was explored by comparing their locations on the genome to the location of evolutionarily important genes and cellular pathways in birds. Specific epimutations were associated with genes related to the bone morphogenic protein, toll receptor, and melanogenesis signaling pathways. Species-specific epimutations were significantly overrepresented in these pathways. As environmental factors are known to result in heritable changes in the epigenome, it is possible that epigenetic changes contribute to the molecular basis of the evolution of Darwin's finches.

Key words: epimutations, DNA methylation, copy number variation, phylogeny, adaptive radiation, BMP, toll, melanogenesis.

Introduction

Epigenetic change has been postulated to play a role in the ecology and evolution of natural populations (Richards et al. 2010; Holeski et al. 2012; Liebl et al. 2013). Epigenetic changes are broadly defined as "molecular processes around DNA that regulate genome activity independent of DNA sequence and are mitotically stable" (Skinner et al. 2010). Some epigenetic processes are also meiotically stable and are transmitted through the germline (Anway et al. 2005; Jirtle and Skinner 2007). These epigenetic mechanisms, such as DNA methylation, can become programmed (e.g., imprinted) and inherited over generations with potential evolutionary impacts. Environmental factors have been shown to promote the epigenetic transgenerational inheritance of phenotypic variants (Skinner et al. 2010). In recent years, the importance of environmental cues in the induction of such variation has been widely acknowledged (Bonduriansky 2012). Thus, like genetic change (Greenspan 2009), epigenetic change may also play an important role in evolution (Guerrero-Bosagna et al. 2005; Day and Bonduriansky 2011; Geoghegan and Spencer 2012, 2013a, 2013b, 2013c; Klironomos et al. 2013).

In order for inherited epigenetic changes to play a significant role in microevolution, they must persist for tens of generations, or longer (Slatkin 2009). It is conceivable that epigenetic changes may also accumulate over longer periods of evolutionary time, contributing to processes such as adaptive radiation (Rebollo et al. 2010; Flatscher et al. 2012). This hypothesis assumes that epigenetic changes persist over thousands of generations. An initial step in testing this hypothesis would be to compare epigenetic differences among closely related species, and whether such changes accumulate over short spans of macroevolutionary time. For example, do epigenetic changes accumulate with phylogenetic distance? Addressing this question was the primary goal of this study.

The study was designed to explore the relationship between epigenetic changes and the evolutionary history of several species of Darwin's finches in the Galapagos Islands. This group of birds has been central to work on a variety of important topics in evolutionary biology, including adaptive radiation, character displacement, rapid evolution, hybridization between species, evolutionary developmental mechanisms, and the effect of invasive pathogens and parasites (Grant and Grant 2008; Huber et al. 2010; Donohue 2011). The adaptive radiation of Darwin's finches over a period of 2–3 Myr resulted in 14 extant species that fill distinct ecological niches. These species show striking variation in body size and the size and shape of their beaks (Grant and Grant 2008). Darwin's finches were selected for study because they are a well-studied example of the evolution of closely related species into different ecological niches (Grant and Grant 2008; Donohue 2011).

Natural selection is a process in which environmental factors influence the survival and reproductive success of individuals bearing different phenotypes. Only selection on phenotypic traits with a heritable basis can lead to evolutionary change (Endler 1986). Observations indicate that epigenetic mechanisms have a role in influencing genomic variability (Huttley 2004; Ying and Huttley 2011). As epigenetic changes are also influenced by environmental factors, and can be heritable across generations (Skinner et al. 2010), they provide another molecular mechanism that can influence evolutionary change. Although Lamarck (1802) proposed that environmental factors can influence inheritance directly, his mechanism has not been widely recognized as a component of modern evolutionary theory (Day and Bonduriansky 2011). Recent work in epigenetics shows that epigenetic changes can, in fact, increase the heritable phenotypic variation available to natural selection (Holeski et al. 2012; Liebl et al. 2013). Thus, epigenetics appears to provide a molecular mechanism that can increase phenotypic variation on which selection acts (Skinner 2011). The integration of genetic and epigenetic mechanisms has the potential to significantly expand our understanding of the origins of phenotypic variation and how environment can influence evolution.

For example, Crews et al. (2007) investigated the ability of an environmental factor (toxicant) to promote the epigenetic transgenerational inheritance of alterations in the mate preferences of rats, with consequences for sexual selection. An F0 generation gestating female rat was exposed to the agricultural fungicide vinclozolin transiently. A dramatic alteration in the mate preferences of the F3 generation was observed (Crews et al. 2007) along with epigenetic alterations (termed epimutations) in the germline (sperm) (Guerrero-Bosagna et al. 2010). Transgenerational transcriptome changes in brain regions correlated with these alterations in mate preference behavior were also observed (Skinner et al. 2008, 2014). Thus, an environmental factor that altered mate preference was found to promote a transgenerational alteration in the sperm epigenome in an imprinted-like manner that was inherited for multiple generations (Crews et al. 2007; Skinner et al. 2010). Studies such as these suggest that environmental epigenetics may play a role in evolutionary changes through processes, such as sexual selection.

Recent reviews suggest a pervasive role for epigenetics in evolution (Rebollo et al. 2010; Day and Bonduriansky 2011; Kuzawa and Thayer 2011; Flatscher et al. 2012; Klironomos et al. 2013). The primary goal of this study was to test whether epigenetic changes accumulate over the long periods of evolutionary time required for speciation with adaptive radiation. Genome wide analyses were used to investigate changes in genetic and epigenetic variation among five species of Darwin's finches. The measure of genetic variation was copy number variation (CNV), which has been shown to provide useful and stable genetic markers with potentially more phenotypic functional links than point mutations such as single nucleotide polymorphisms (SNPs) (Lupski 2007; Sudmant et al. 2013). CNVs involve an increase or decrease in the number of copies of a repeat element at a specific genomic location. Recently, CNV changes in primates and other species have been shown to be very useful genetic measures for comparing evolutionary events (Nozawa et al. 2007; Gazave et al. 2011; Poptsova et al. 2013). CNV changes are involved in gene duplication and deletion phenomena, as well as repeat element phenomenon such as translocation events and can be influenced by DNA methylation (Skinner et al. 2010; Macia et al. 2011; Tang et al. 2012). The measure of epigenetic variation used was differential DNA methylation sites, which are known to be stable and heritable (Skinner et al. 2010). Comparing data for both genetic mutations (i.e., CNV) and epimutations (i.e., DNA methylation) allowed the relative magnitudes of these sources of variation to be compared across the five species included in the study.

Materials and Methods

Finch Field Work and Collection of Blood

Blood samples were collected from birds captured January–April 2009 at El Garrapatero, a lowland arid site on Santa Cruz Island, Galapagos Archipelago, Ecuador (Koop et al. 2011).

Birds were captured with mist nests and banded with numbered Monel bands to track recaptures. Birds were identified, aged, and sexed using size and plumage characteristics. A small blood sample (90 μl) from each bird was collected in a microcapillary tube through brachial venipuncture. Samples were stored on wet ice in the field, then erythrocytes purified by centrifugation and cells stored in a −20 °C freezer at a field station. Following the field season, samples were placed in a −80 °C freezer for longer term storage. All procedures were approved by the University of Utah Institutional Animal Care and Use Committee (protocol #07-08004) and by the Galápagos National Park (PC-04-10: #0054411).

DNA Processing

Erythrocyte DNA was isolated with DNAeasy Blood and Tissue Kit (Qiagen, Valencia, CA) and then stored at −80 °C prior to analysis. DNA was sonicated following a previously described protocol (without protease inhibitors) (Tateno et al. 2000) and then purified using a series of washes and centrifugations (Ward et al. 1999) from variable number of animals per species analyzed. The same concentrations of DNA from individual blood samples were then used to produce pools of DNA material. Two DNA pools were produced in total per species, each one containing the same amount of DNA from different animals. The number of individuals used per pool is shown in supplementary table S6, Supplementary Material online. These DNA pools were then used for chromosomal genomic hybridization (CGH) arrays or chromatin immunoprecipitation of methylated DNA fragments (MeDIP).

CNV Analysis

The array used for the CNV analysis was a CGH custom design by Roche Nimblegen that consisted of a whole-genome tiling array of zebra finch (*Taeniopygia guttata*) with 720,000 probes per array. The probe size ranged from 50 to 75 mer in length with median probe spacing of 1,395 bp. Two different comparative (CNV vs. CNV) hybridization experiments were performed (two subarrays) for each species in query (*Geospiza fuliginosa* [FUL], *G. scandens* [SCA], *Camarhynchus parvulus* [PAR], and *Platyspiza crassirostris* [CRA]) versus control *G. fortis* (FOR), with each subarray including hybridizations from DNA pools from these different species. Two DNA pools were built for each species (supplementary table S6, Supplementary Material online). For one subarray of each species, DNA samples from the experimental groups were labeled with Cy5 and DNA samples from the control lineage were labeled with Cy3. For the other subarray of each species, a dye swap was performed so that DNA samples from the experimental groups were labeled with Cy3 and DNA samples from the control lineage were labeled with Cy5.

For the CNV experiment raw data from the Cy3 and Cy5 channels were imported into R (R Development Core Team 2010), checked for quality, and converted to *MA* values ($M = \text{Cy5} - \text{Cy3}$; $A = [\text{Cy5} + \text{Cy3}]/2$). Within array and between array normalizations were performed as previously described (Manikkam et al. 2012). Following normalization, the average value of each probe was calculated and three different CNV algorithms were used on each of these probes including circular binary segmentation from the DNA copy (Olshen et al. 2004), CGHseg (Picard et al. 2005) and cghFlasso (Tibshirani and Wang 2008). These three algorithms were used with the default parameters. The average values from the output of these algorithms were obtained. A threshold of 0.04 as a cutoff was used on the summary (average of the log-ratio from the three algorithms) where gains are probes above the positive threshold and losses are probes below the negative threshold. Consecutive probes (≥ 3) of gains and losses were used to identify separate CNV regions. A cutoff of three-probe minimum was used and those regions were considered a valid CNV. The statistically significant CNVs were identified and *P* values associated with each region presented. A cutoff of $P < 10^{-5}$ was used to select the final regions of gains and losses.

Differential DNA Methylation Regions Analysis

MeDIP was performed as previously described (Guerrero-Bosagna et al. 2010) as follows: 6 μg of genomic DNA was subjected to series of three 20-pulse sonications at 20% amplitude and the appropriate fragment size (200–1,000 ng) was verified through 2% agarose gels; the sonicated genomic DNA was resuspended in 350 μl TE buffer and denatured for 10 min at 95 °C and then immediately placed on ice for 5 min; 100 μl of 5× IP buffer (50 mM Na-phosphate pH 7, 700 mM NaCl (PBS), 0.25% Triton X-100) was added to the sonicated and denatured DNA. An overnight incubation of the DNA was performed with 5 μg of antibody anti-5-methylCytidine monoclonal from Diagenode (Denville, NJ) at 4 °C on a rotating platform. Protein A/G beads from Santa Cruz were prewashed on PBS–BSA (bovine serum albumin) 0.1% and resuspended in 40 μl 1× IP (immunoprecipitation) buffer. Beads were then added to the DNA-antibody complex and incubated 2 h at 4 °C on a rotating platform. Beads bound to DNA-antibody complex were washed three times with 1 ml 1× IP buffer; washes included incubation for 5 min at 4 °C on a rotating platform and then centrifugation at 6,000 rpm for 2 min. Beads DNA-antibody complex were then resuspended in 250 μl digestion buffer (50 mM Tris–HCl pH 8, 10 mM ethylenediaminetetraacetic acid, 0.5% SDS (sodium dodecyl sulfate) and 3.5 μl of proteinase K (20 mg/ml) was added to each sample and then incubated overnight at 55 °C on a rotating platform. DNA purification was performed first with phenol and then with chloroform:isoamyl alcohol. Two washes were then performed with 70% ethanol, 1 M NaCl, and glycogen. MeDIP-selected DNA was then resuspended in 30 μl TE buffer.

The array used for the differential methylation analysis was a DNA-methylated custom array by Roche Nimblegen that

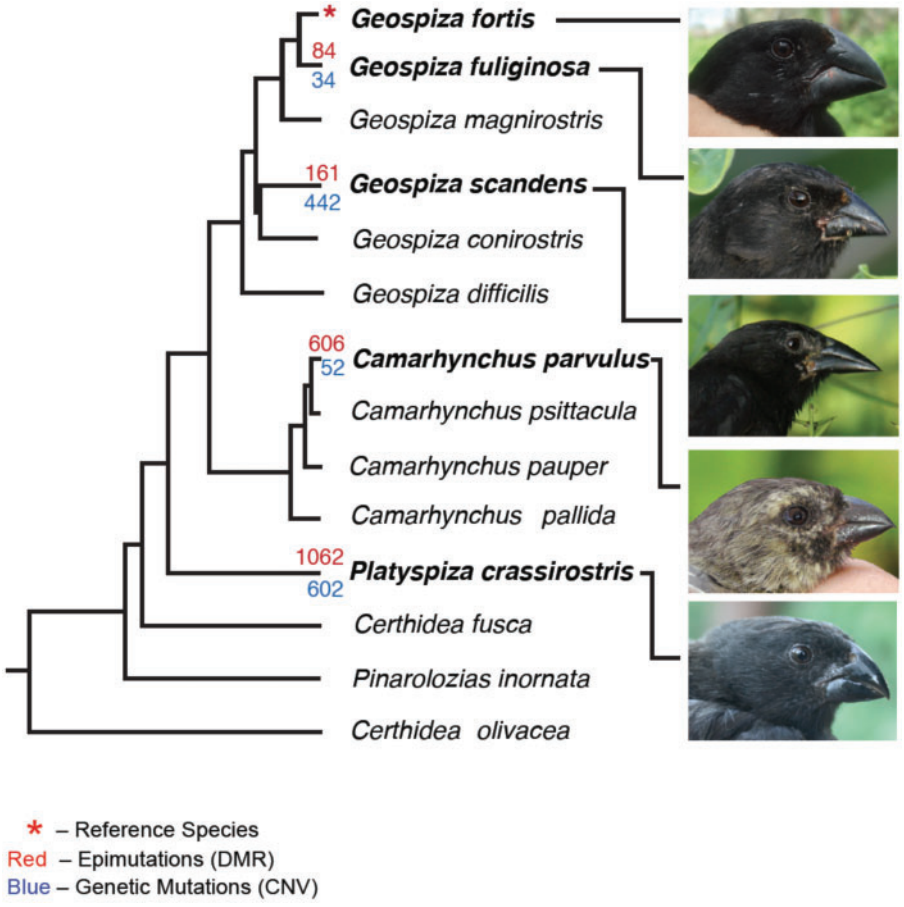


Fig. 1.—Number of epimutations and genetic mutations in relation to the phylogenetic relationships of five species of Darwin's finches. Photographs (by J.A.H.K. or S.A.K.) show variation in bill size and shape. Numbers on branches are the number of differences (three or more probes; table 1) in epimutations (DMR; in red) .and genetic mutations (CNV; in blue) for each of four species, compared with a single reference species FOR (asterisk). The phylogram is based on allele length variation at 16 polymorphic microsatellite loci (from Petren et al. 1999). The topology of the tree is similar to that proposed by Lack (1947) on the basis of morphological traits.

consisted of a whole-genome tiling array of zebra finch (*Taeniopygia guttata*) made of four 2.1M and one 3x720k array with 8,539,570 probes per array. Probe sizes were 50–75 mer in length and median probe spacing was 200 bp. Two different comparative (MeDIP vs. MeDIP) hybridization experiments were performed (two subarrays) for each experimental species (FUL, SCA, PAR, CRA) versus control FOR, with each subarray including hybridizations from MeDIP DNA from DNA pools from these different species (supplementary table S6, Supplementary Material online). For one subarray of each species, MeDIP DNA samples from the experimental groups were labeled with Cy5 and MeDIP DNA samples from the control lineage were labeled with Cy3. For the other subarray of each species, a dye swap was performed so that MeDIP DNA samples from the experimental groups were labeled with Cy3 and MeDIP DNA samples from the control lineage were labeled with Cy5.

For each comparative hybridization experiment, raw data from both the Cy3 and Cy5 channels were imported into R, checked for quality, and converted into *MA* values. The normalization procedure is as previously described (Guerrero-Bosagna et al. 2010). Following normalization each adjacent $\geq$3 probe set value represents the median intensity difference between FUL, SCA, PAR and CRA and control FOR of a 600-bp window. Significance was assigned to probe differences between experimental species samples and reference FOR samples by calculating the median value of the intensity differences as compared with a normal distribution scaled to the experimental mean and standard deviation of the normalized data. A *Z* score and *P* value were computed for each probe from that distribution. The statistically significant differential DNA methylation regions (DMR) were identified and *P* values associated with each region represented, as previously described (Guerrero-Bosagna et al. 2010).

A

Differential DNA Methylation Regions (DMR) (Epimutations)							
All probes ($p<10^{-5}$)				3 or more probes ($p<10^{-5}$)			
	(Up)	(Down)	Total		(Up)	(Down)	Total
FUL	116	398	514	FUL	76	8	84
SCA	211	679	890	SCA	17	144	161
PAR	191	1438	1629	PAR	28	578	606
CRA	361	2406	2767	CRA	61	1001	1062
Total	**Up**	**Down**	**Total Sites**	**Total**	**Up**	**Down**	**Total Sites**
	879	4921	5800		182	1731	1913

Copy Number Variation (CNV)							
All probes ($p<10^{-5}$)				3 or more probes ($p<10^{-5}$)			
	Gains	Loss	Total		Gains	Loss	Total
FUL	59	12	71	FUL	28	6	34
SCA	567	22	589	SCA	440	2	442
PAR	78	217	295	PAR	15	37	52
CRA	621	194	815	CRA	541	61	602
Total	**Gains**	**Loss**	**Total Sites**	**Total**	**Gains**	**Loss**	**Total Sites**
	1325	445	1770		1024	106	1130

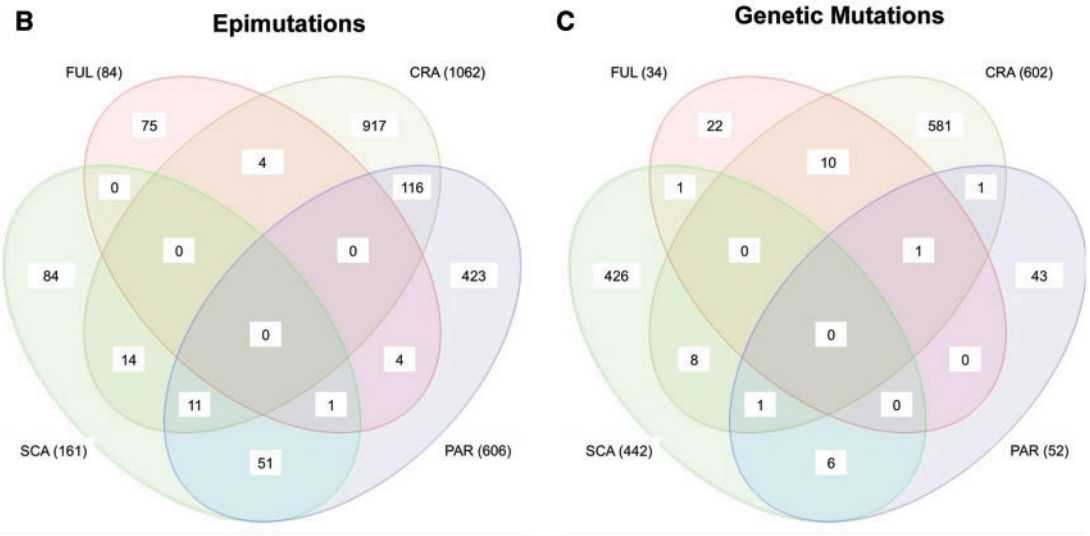


Fig. 2.—Number of epimutations and genetic mutations associated with Darwin's finches. The number of differential DMR epimutations and CNV genetic mutations (*A*). DMR and CNV that differ significantly ($P < 10^{-5}$) from the reference species (FOR) are presented for all oligonucleotide probes, compared with peaks of three or more adjacent probes. The epimutations with an increase (Up) or decrease (Down) in DNA methylation are indicated. Those genetic mutations with an increase (Gain) or decrease (Loss) in CNV are indicated. Venn diagrams for epimutations (*B*) and genetic mutations (*C*) show overlaps between epimutations (DMR) and genetic mutations (CNV) among species. The species and total number of sites compared are listed on the outside of each colored elliptical.

Additional Bioinformatics and Statistics

The July 2008 assembly of the zebra finch genome (taeGut1, WUSTL v3.2.4) produced by the Genome Sequencing Center at the Washington University in St Louis (WUSTL) School of Medicine was retrieved (WUSTL 2008). A seed file was constructed and a BSgenome package was forged for using the Finch DNA sequence in the R code (Herve Pages BSgenome: Infrastructure for Biostrings-based genome data packages. R

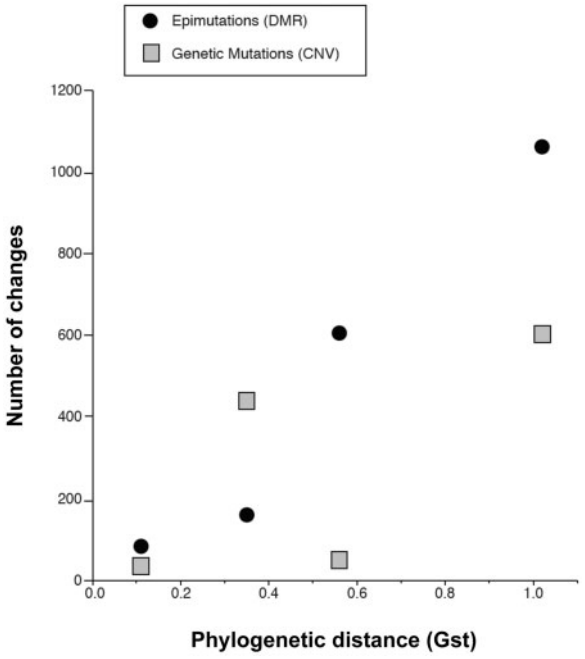


Fig. 3.—Phylogenetic distance is correlated with epigenetic changes, but not genetic changes. Branch lengths in figure 1 were used as measures of phylogenetic distance. The number of epimutations increased with phylogenetic distance (Spearman Rho = 1.0, $P < 0.0001$). In contrast, the number of genetic mutations did not increase with phylogenetic distance (Spearman Rho = 0.8, $P = 0.2$).

package version 1.24.0). This sequence was used to design the custom tiling arrays and to perform the bioinformatics.

The chromosomal location of CNV and DMR clusters used an R-code developed to find chromosomal locations of clusters (Skinner et al. 2012). A 2-Mb sliding window with 50,000 base intervals was used to find the associated CNV and DMR in each window. A *Z*-test statistical analysis with $P < 0.05$ was used on these windows to find the ones with overrepresented CNV and DMR were merged together to form clusters. A typical cluster region averaged approximately 3 Mb in size.

The DMR and CNV association with specific zebra finch genes and genome locations used the Gene NCBI database for zebra finch gene locations and correlated the epimutations associated (overlapped) with the genes. The three adjacent probes constituted approximately a 200-bp homology search. The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway associations were identified as previously described (Skinner et al. 2012). Statistically significant overrepresentation uses a Fisher's exact analysis.

Spearman Rank correlation coefficients were used to test for a relationship between phylogenetic distance and epigenetic and genetic changes (Whitlock and Schluter 2009).

Results

Phylogenetic relationships of the five finch species in this study are shown in figure 1. The taxa chosen for this study included: Two species of ground finches, FOR and FUL, which have crushing beaks with relatively deep bases; the cactus finch SCA, which has a long thin beak used for probing flowers; the small tree finch PAR, which has curved mandibles used for applying force at the tips; and the vegetarian finch CRA, which has a relatively short stubby bill used for crushing food along its entire length (Grant and Grant 2008; Donohue 2011; Rands et al. 2013). FOR was selected as a reference species for comparing genetic and epigenetic alterations among the remaining four species. Branch lengths in figure 1 were used as measures of phylogenetic distance.

The experimental design used purified erythrocytes from the different species. Although DNA sequences are the same for all cell types of an organism, the epigenome is distinct for each cell type, providing a molecular mechanism for the genome activity and functions that differ among different cell types (Skinner et al. 2010). Therefore, to investigate the overall epigenome requires a purified cell type. As birds have erythrocytes (red blood cells) that contain nuclei, samples of purified erythrocytes were collected from each of the Darwin's finch species to obtain DNA for molecular analysis.

The epigenetic alterations termed epimutations were assessed through the identification of differential DMR. The DMR were identified with the use of MeDIP with a methyl cytosine antibody, followed by a genome wide tiling array (Chip) for an MeDIP-Chip protocol (Guerrero-Bosagna et al. 2010). Although other epigenetic processes such as histone modifications, chromatin structure, and noncoding RNA are also important, DNA methylation is the best known epigenetic process associated with germline-mediated heritability and environmental manipulations (Skinner et al. 2010). Genetic variation was assessed using CNVs (i.e., amplifications and deletions of repeat elements) in the DNA using a CGH protocol (Pinkel and Albertson 2005; Gazave et al. 2011).

The reference genome used for the analysis was that of the zebra finch (*Taeniopygia guttata*) (Clayton et al. 2009), which had a preliminary estimate of greater than 83% similarity with a partial shotgun sequence of a Darwin's finch genome (Rands et al. 2013). This study actually suggests a much higher degree of identity. The zebra finch genome was tiled in a genome wide array with a 200-bp resolution and for a CGH array with a 1,500-bp resolution. These arrays were used in a competitive hybridization protocol between FOR (reference species) and the other four species (Guerrero-Bosagna et al. 2010). Differential hybridization using two different fluorescent DNA labeling tags identified the CNV with CGH using genomic DNA and the epimutation DMR with a MeDIP-Chip protocol. A statistical significance threshold of $P < 10^{-5}$ was set for the CNV or epimutation to be identified as a gain or loss compared with the reference species (fig. 2 and supplementary tables S1 and S2, Supplementary Material online). The data for all probes (oligonucleotides on the arrays) are presented. However, the criteria used to identify the CNV and DMR required the involvement of three or more adjacent

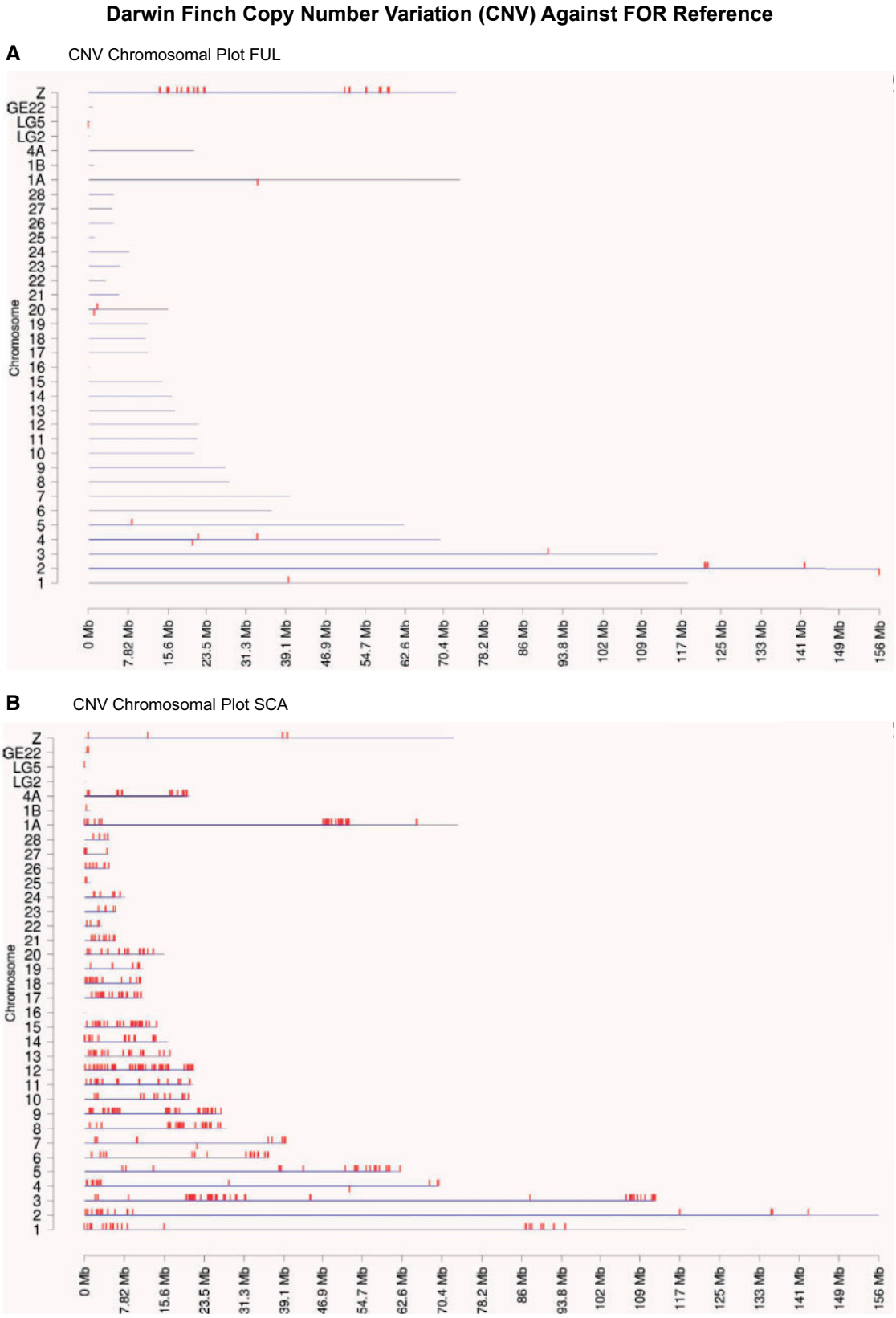


Fig. 4.—Chromosomal locations of the CNVs for each species. The chromosome number and size are presented in reference to the zebra finch genome. The chromosomal location of each CNV is marked with a red tick for FUL (*A*), SCA (*B*), PAR (*C*), and CRA (*D*).

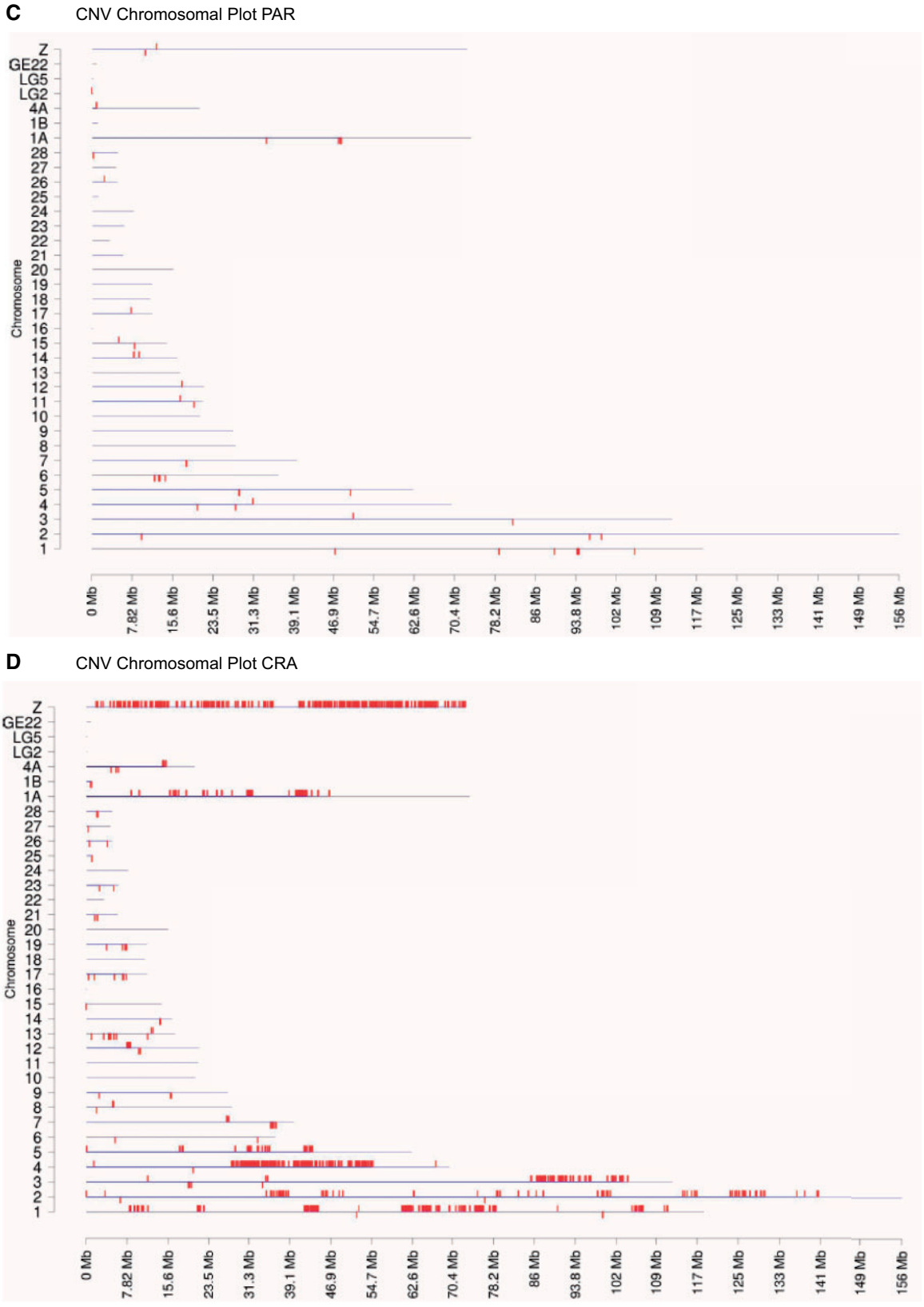


Fig. 4.—Continued.

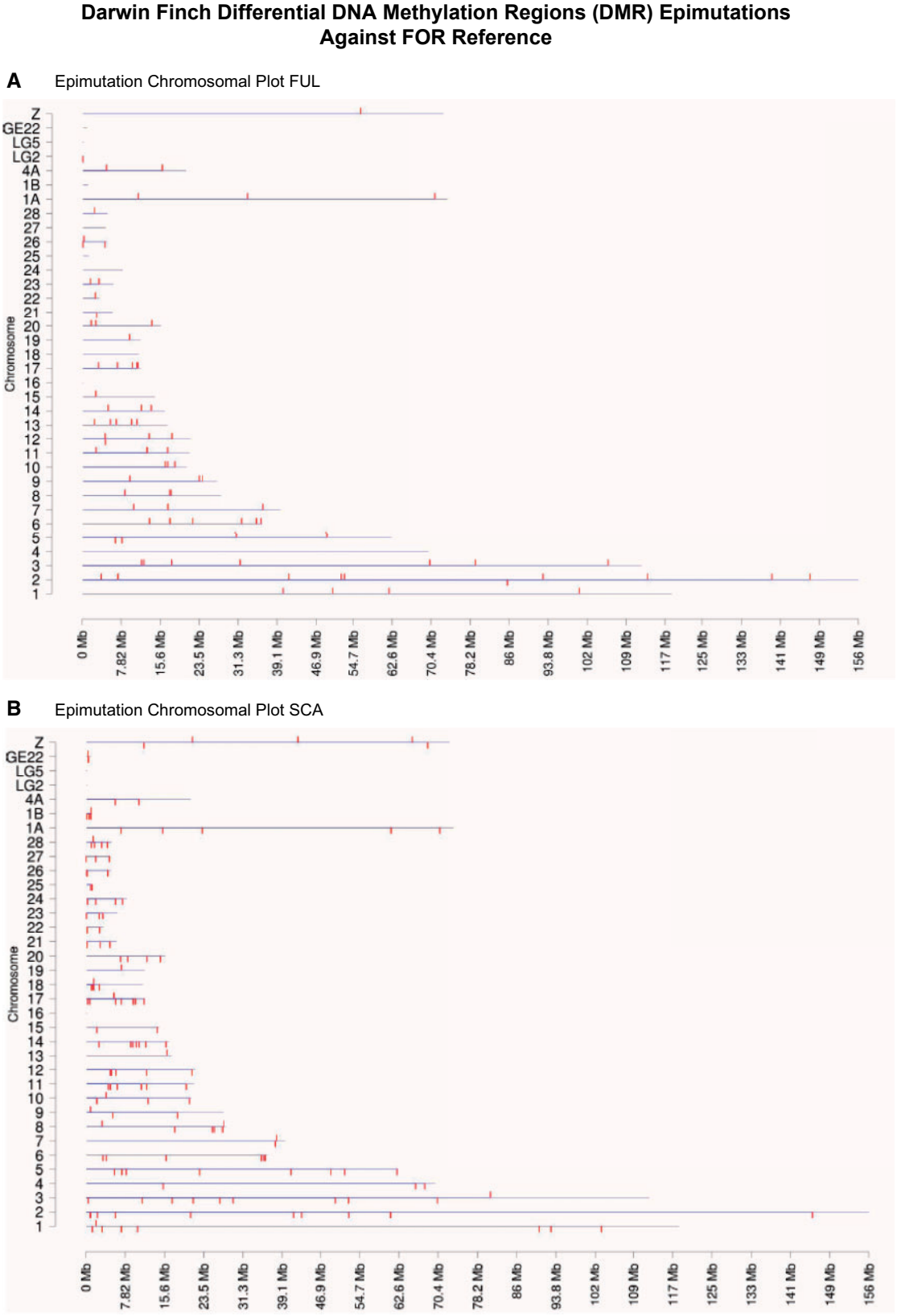
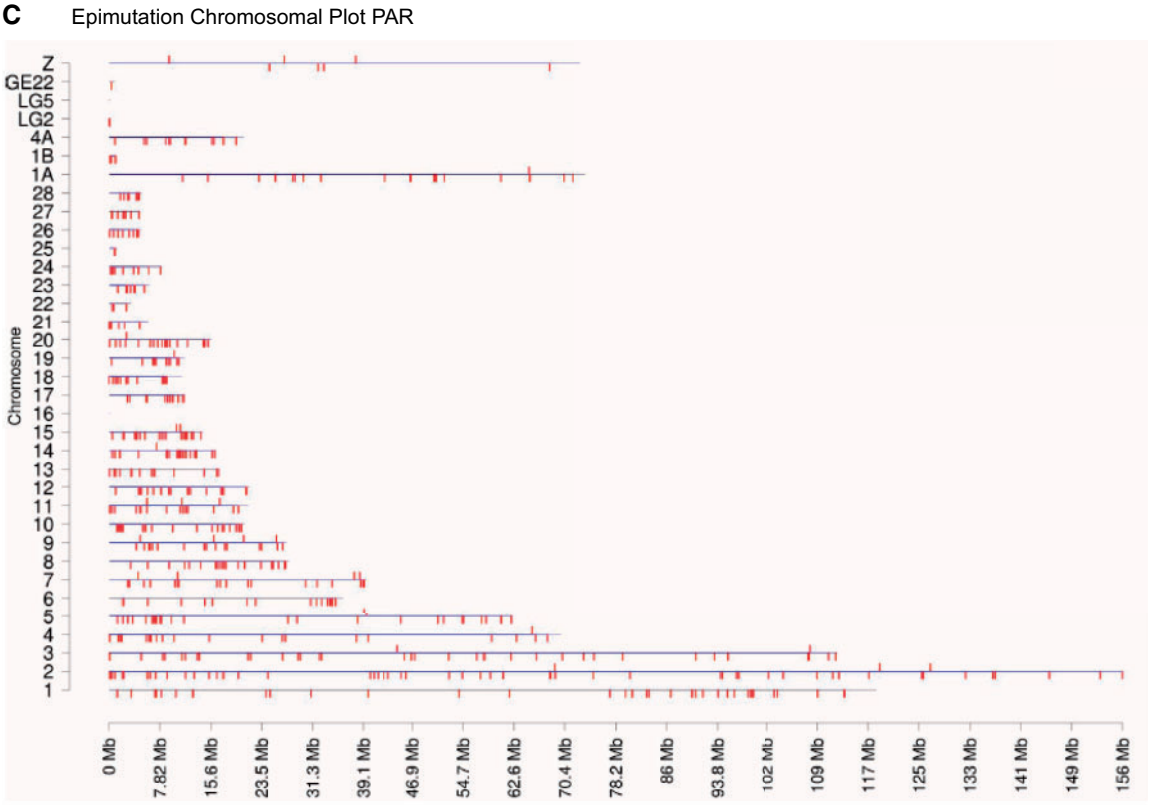


Fig. 5.—Chromosomal locations of the epimutations for each species. The chromosome number and size are presented in reference to the zebra finch genome. The chromosomal location of each DMR is marked with a red tick for FUL (*A*), SCA (*B*), PAR (*C*), and CRA (*D*).



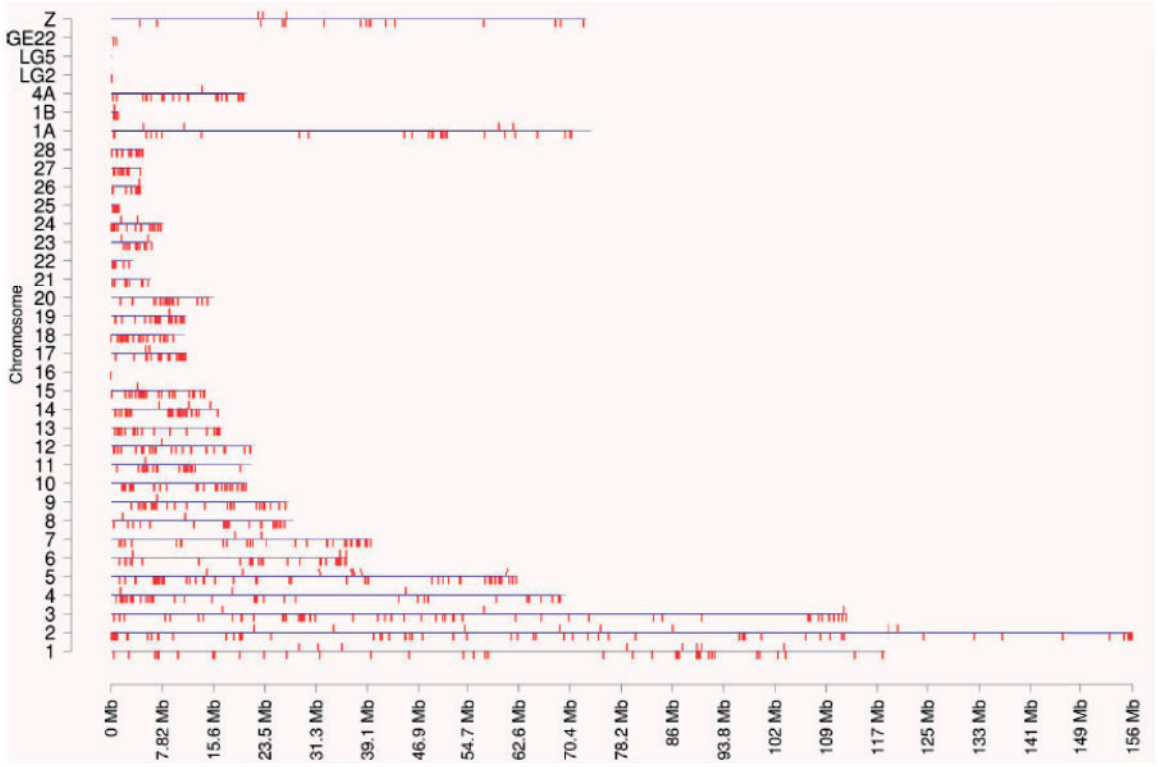


Fig. 5.—Continued.

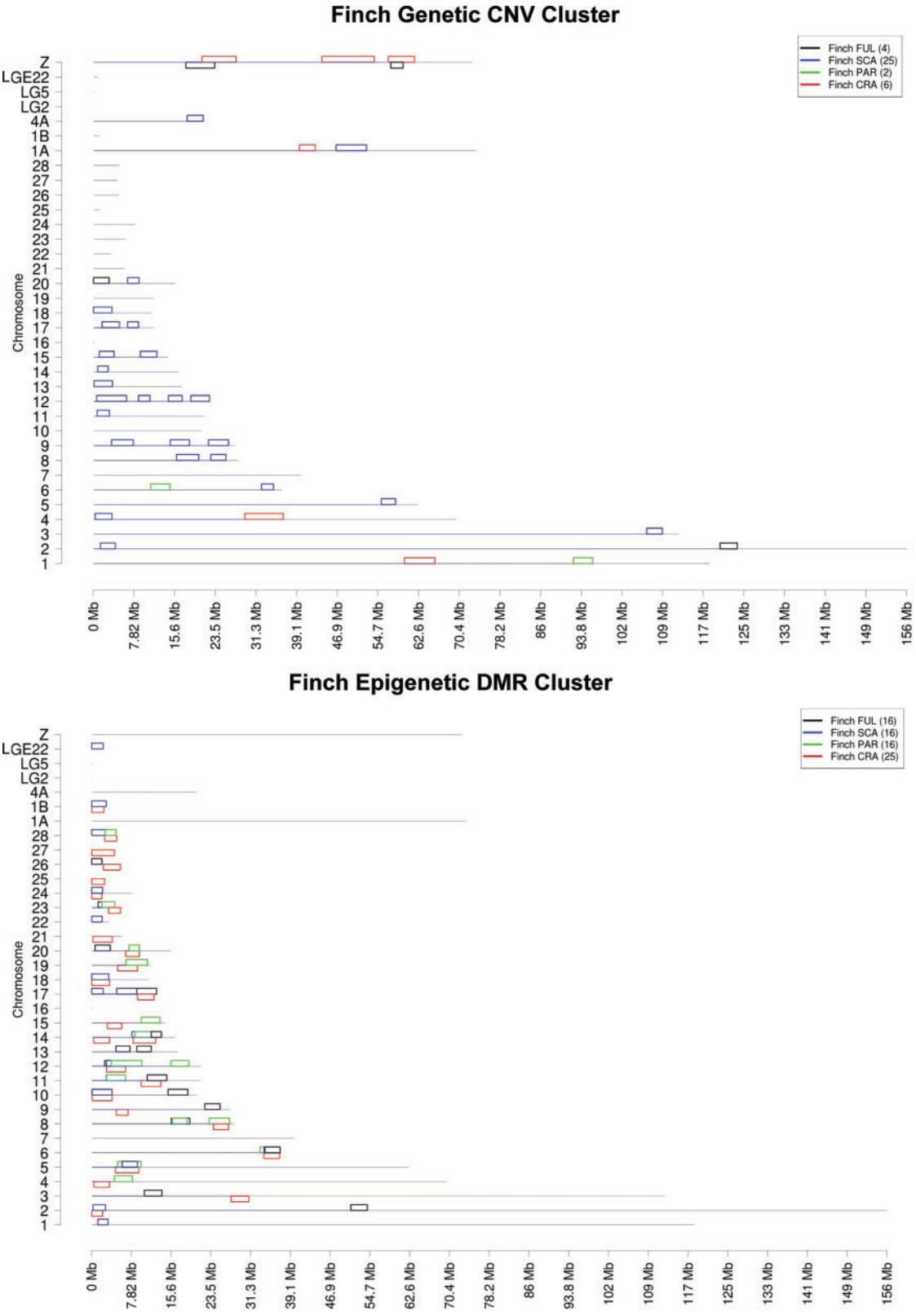


Fig. 6.—Chromosomal locations for clusters of CNV and DMR. The chromosome number and size are presented in reference to the zebra finch genome. The chromosomal location of statistically significant ($P < 10^{-5}$) overrepresented clusters of CNV (*A*) and DMR (*B*). The legend shows species and total number of clusters.

probes in the genome sequence having significant differential hybridization. These selection criteria reduce the number of false positives and provide a more reliable comparison (fig. 2). Therefore, the data presented used stringent criteria and represent the most reproducible epimutations and genetic CNV mutations among all three different experiments.

The increases or decreases in DNA methylation for the DMR are presented, along with the total number of epimutations in figure 2. The majority of epimutations for all species but FUL involves a decrease in DNA methylation (fig. 2*A*). The gains or losses in CNV are also presented, along with the total number of genetic alterations. The majority of genetic mutations for all species but PAR involves an increase in CNV number. Interestingly, the number of epimutations observed was generally higher, using the criteria selected, than the number of genetic alterations (fig. 2). However, the overall magnitude of epigenetic change was comparable to that of genetic change. Data for the five different species are shown in figure 1 for both epimutations (red) and genetic alterations (blue). The number of epimutations was significantly correlated with phylogenetic distance, whereas the number of genetic mutations was not (fig. 3).

The chromosomal locations of the CNV for the different finch species are shown in figure 4. CNVs were found on most chromosomes, with FUL having the least and CRA having the most. The chromosomal locations of the DMR epimutations for the different finch species are shown in figure 5. All chromosomes were found to have epimutations, with CRA having the highest number. These chromosomal plots suggested that some of the species might have clusters of CNV and/or DMR on some of the chromosomes (figs. 3 and 4). Therefore, a cluster analysis previously described (Skinner et al. 2012) was used to examine 50-kb regions throughout the genome to test for statistically significant ($P < 10^{-5}$) overrepresentation of CNV or DMR (fig. 6). Clusters, which have an average size of 3 Mb, are shown as species-specific boxes for CNV (fig. 6*A*) and for DMR (fig. 6*B*). Cluster characteristics and overlap are presented in supplementary table S3, Supplementary Material online. Clusters were obtained for all species, with a higher number of DMR clusters than CNV clusters. The highest number of CNV clusters was in SCA, with more than a 4-fold increase over CRA (fig. 6). Therefore, in addition to having more CNV than expected (assuming an increasing number with phylogenetic distance), SCA showed more CNV clusters than other species (fig. 2). Genome instability in these cluster regions may influence the increased numbers of CNV in SCA, which increases the presence of CNV clusters. In contrast, SCA did not show more DMR numbers or clusters than expected, assuming an increasing number with phylogenetic distance. Epimutation cluster overlap was more common among species (fig. 6 and table 1), suggesting that specific regions of the chromosomes were more susceptible to epigenetic alterations. Altered DNA methylation states have been experimentally shown to be stable for hundreds of generations (Cubas et al. 1999; Akimoto et al. 2007; Skinner et al. 2010).

Table 1
Cluster Overlap between Species

CNVs				
	CNV			
	FUL	**SCA**	**PAR**	**CRA**
FUL	4	0	0	2
SCA	0	25	0	0
PAR	0	0	2	0
CRA	2	0	0	6
Epimutations				
	DMR			
	FUL	**SCA**	**PAR**	**CRA**
FUL	16	5	6	7
SCA	5	16	8	11
PAR	6	8	16	11
CRA	7	11	11	25

NOTE.—The overlap of CNV or DMR clusters between species is presented for the CNVs and epimutations.

The potential overlaps in specific CNV or DMR sites among species were examined. The overlap in genetic mutations among the four species is shown in a Venn diagram in figure 2*C*, whereas the overlap in epimutations is shown in figure 2*B*. No overlap in specific CNV or DMR sites was observed among all species, and less than 10% overlap was generally observed between any two species. Interestingly, the CNV overlap between FUL and CRA was higher than for the other species (fig. 2*C*). Generally, genetic and epigenetic alterations were distinct between species, with the majority being species specific. The epimutations showed more overlap between species than the genetic CNV mutations (fig. 2*B* and table 1). In considering within species overlap between the CNV and epimutations, less than 3% had common genomic locations. Therefore, the epimutations do not appear to be linked to the genetic CNV mutations, but are distinct.

The final analysis examined the potential functional significance of the epimutations by examining DMR and genes known to be associated with avian evolution. Several gene families and cellular signaling pathways have previously been shown to be involved in bird evolution, including the bone morphogenic protein (BMP) family and pathway (Abzhanov et al. 2004; Badyaev et al. 2008), the toll receptor family and signaling pathway (Alcaide and Edwards 2011), and the melanins family and pathway (Mundy 2005). All the genes associated with these signaling pathways were localized on the finch genome and compared with the genomic locations of the epimutations and CNV. Epimutation-associated genes within the BMP pathway (fig. 7), toll pathway (fig. 8), and

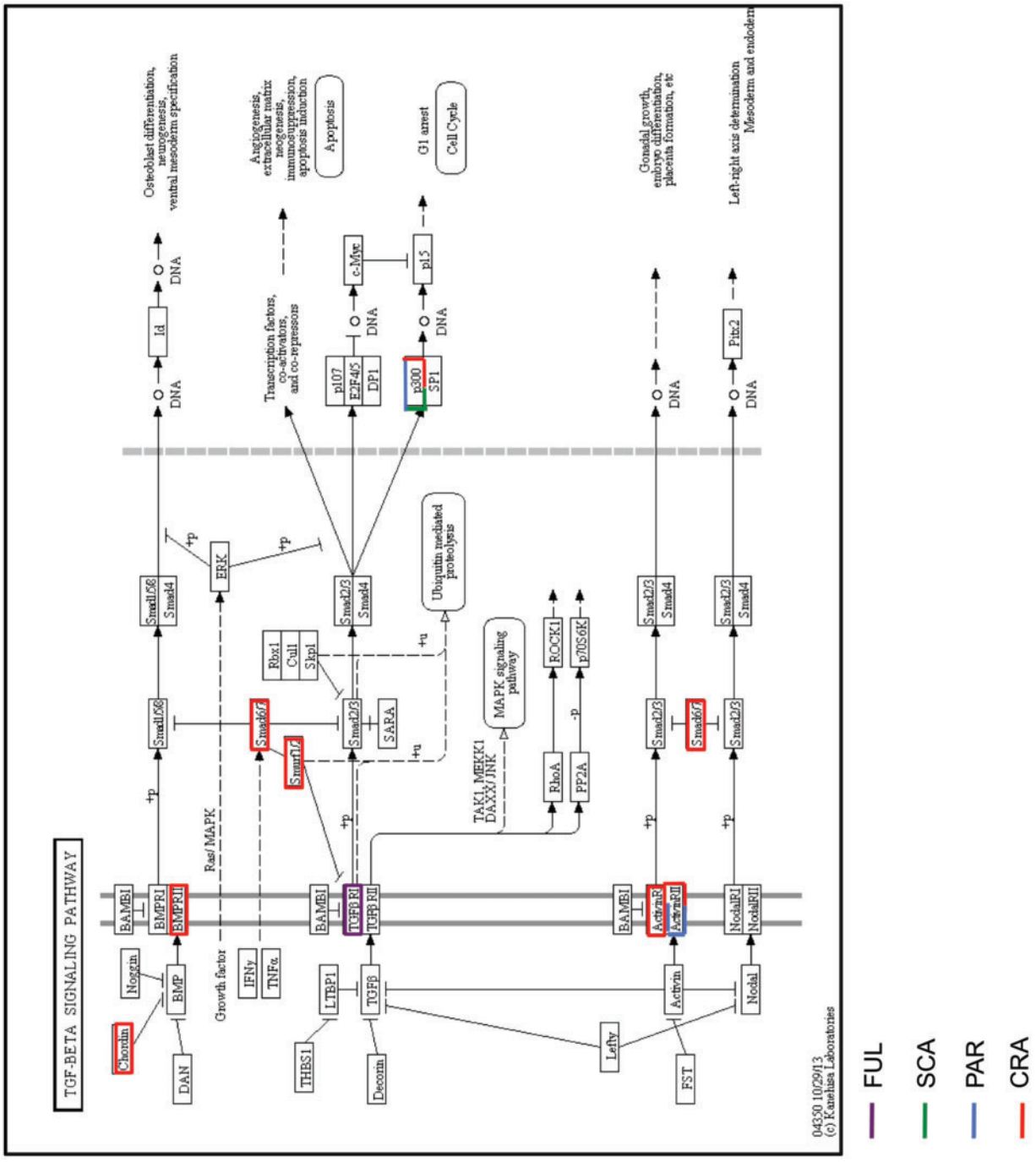


Fig. 7.—Epimutation-associated genes and correlated BMP pathway. The genes having associated epimutations in the signaling pathway presented for the different species are identified as FUL (purple), SCA (green), PAR (blue), and CRA (red) colored boxed genes.

melanin's pathway (fig. 9) are shown. Epimutations were overrepresented in all of these pathways (Fisher's exact test: BMP/TGFbeta (transforming growth factor) pathway, $P < 1 \times 10^{-6}$; toll pathway, $P < 5.7 \times 10^{-4}$; melanogenesis pathway, $P < 2.5 \times 10^{-13}$). Interestingly, the BMP pathway involved in beak development and shape had a statistically significant overrepresentation of CRA-associated epimutations when examined independently ($P < 2.7 \times 10^{-5}$) (fig. 7). In addition, the toll receptor pathway involved in immune response had a statistically significant overrepresentation of PAR-associated epimutations when examined independently ($P < 7.7 \times 10^{-4}$) (fig. 8). The melanogenesis pathway involved in color had a mixture of epimutations from most of the species when examined independently ($P < 7 \times 10^{-5}$) (fig. 9).

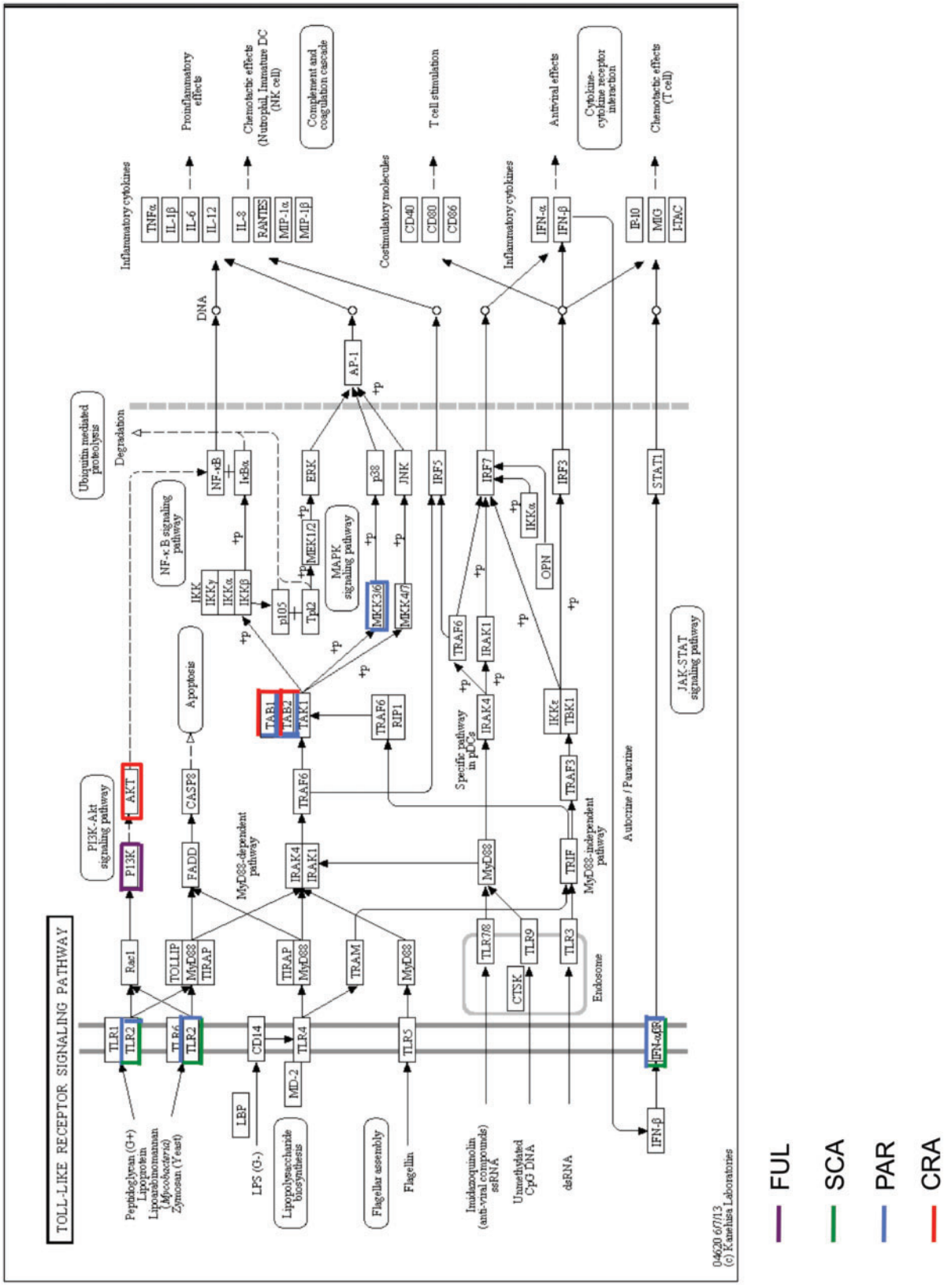


Fig. 8.—Epimutation-associated genes and correlated toll receptor pathway. The genes having associated epimutations in the signaling pathway presented for the different species are identified as FUL (purple), SCA (green), PAR (blue), and CRA (red) colored boxed genes.

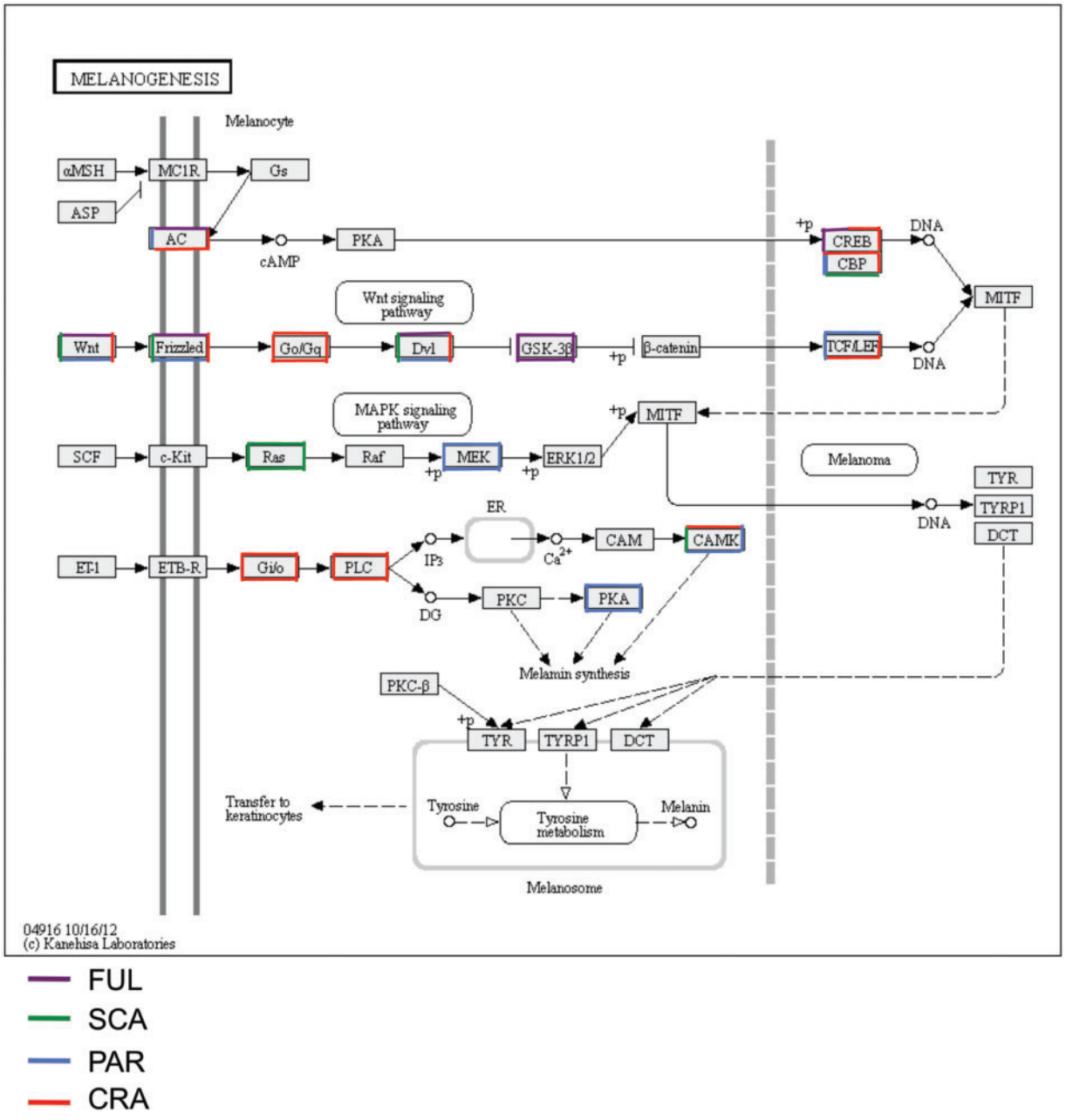


Fig. 9.—Epimutation-associated genes and correlated melanogenesis pathway. The genes having associated epimutations in the signaling pathway presented for the different species are identified as FUL (purple), SCA (green), PAR (blue), and CRA (red) colored boxed genes.

In addition to the pathway-specific genes, the total number of epimutations and CNV associated with genes are presented in table 2, with full lists in supplementary tables S4 and S5, Supplementary Material online. The epimutations and CNV for single probe and ≥3 probe identification are presented in table 2. Observations indicate that approximately half of the epimutations and CNV identified were associated with genes. Therefore, a high percentage of the epimutations and CNV identified were associated with genes and were statistically overrepresented in several gene pathways previously shown to be involved in particular aspects of avian evolution. Although this gene association analysis demonstrates that epimutations correlate with genes and important pathways, the functional or causal link to specific evolutionary processes remains to be investigated.

Discussion

This study provides one of the first genome-wide comparisons of genetic and epigenetic mutations among related species of

Table 2
Epimutation and CNV Gene Associations

CNVs				
	Total CNV 1+ Probes	Total CNV 3+ Probes	CNV Association with 14K Genes 1+ Probes	CNV Association with 14K Genes 3+ Probes
FUL	71	34	40	24
SCA	589	442	363	350
PAR	295	52	136	37
CRA	815	602	437	345
Epimutations				
	Total Epimutations 1+ Probes	Total Epimutations 3+ Probes	Epimutation Association with 14K Genes 1+ Probes	Epimutation Association with 14K Genes 3+ Probes
FUL	514	84	295	48
SCA	890	161	558	115
PAR	1,629	606	996	407
CRA	2,767	1,062	1,611	639

NOTE.—The 14,000 zebra finch genes annotated having epimutation or CNV associations are presented for the total number of associations (overlaps) for both regions identified with single (1+ probes) and adjacent (3+ probes) data sets.

organisms. There were relatively more epimutations than genetic CNV mutations among the five species of Darwin's finches, which suggests that epimutations are a major component of genome variation during evolutionary change. There was also a statistically significant correlation between the number of epigenetic differences and phylogenetic distance between finches (figs. 1 and 3), indicating that the number of epigenetic changes continues to accumulate over long periods of evolutionary time (2–3 Myr). In contrast, there was no significant relationship between the number of genetic CNV changes and phylogenetic distance.

The zebra finch genome was used as a reference for this study because a complete Darwin's finch genome is not yet available. The zebra finch genome showed hybridization with all probes on the array for each of the Darwin's finch species, suggesting that the genomes appear to be extremely similar. Loss of heterozygosity (absence of genomic regions, resulting in lack of probe hybridization) was not identified in any of the analyses. This suggests a high level of conservation and identity between the species' genomes. In the event the Darwin's finch genome has additional DNA sequence that is not present in the zebra finch genome, we would not have detected this DNA. Therefore, our data may be an underestimate of the Darwin's finch genome. Another technical limitation of our study was that we only considered genetic CNV (amplifications and deletions of repeat elements), but not other genetic variants such as point mutations or translocations. Although CNV frequency is higher than other mutations (e.g., SNPs) and stable in the genome (Gazave et al. 2011), this study's focus on CNV should kept in mind. The epimutations examined are differential DMR that have previously been shown to be frequent and transgenerationally stable (Anway et al. 2005; Guerrero-Bosagna et al. 2010; Skinner et al. 2010). Although other epigenetic processes such as histone modification, altered chromatin structure, and noncoding RNA may also be important, DNA methylation is the most established heritable epigenetic mark. This aspect of the experimental design should be kept in mind.

Among the five species of finches there were fewer genetic mutations (CNV) than epigenetic mutations. However, the cactus finch SCA showed a surprisingly large number of genetic CNV mutations than expected when compared with the reference species (FOR). The SCA mutations also clustered to similar locations on the genome to a greater extent than in the other species (fig. 6*A*). The reason for the disproportionately large number of CNV in the SCA comparison is unclear.

In contrast to the genetic mutation (CNV) analysis, the number of epimutations increased monotonically with phylogenetic distance (figs. 1 and 3). Overlap of specific epigenetic sites among species was minimal, including those for SCA (fig. 2*B*). An interesting possibility is that the epigenome may alter genome stability and generate genetic variation within species. A similar phenomenon has been shown for cancer, in which epigenetic alterations may precede genetic changes and alter genomic stability (Feinberg 2004). A decrease in the DNA methylation of specific repeat elements has previously been shown to correlate with an increase in CNV (Macia et al. 2011; Tang et al. 2012). Therefore, environmentally induced abnormal epigenetic shifts may influence genetic

mutations, such that a combination of epigenetics and genetics promotes phenotypic variation. Our observations demonstrate a relationship between the number of epigenetic changes and phylogenetic distance.

A comparison of the positions of epimutations and known gene families was also carried out. These gene families included those involved in the BMP pathway, which is related to beak shape (Badyaev et al. 2008), the toll receptor pathway, which is involved in immunological function (Alcaide and Edwards 2011), and the melanogenesis pathway, which affects color (Mundy 2005). Genes in all three of these families and signaling pathways were found to have species-specific epimutations (figs. 7–9). Future studies should focus on the causal relationship between epigenetic alterations and phenotypic traits.

Genetic mutations are postulated to provide much of the variation upon which natural selection acts (Gazave et al. 2011; Stoltzfus 2012). However, genetic changes alone are limited in their ability to explain phenomena ranging from the molecular basis of disease etiology to aspects of evolution (Skinner et al. 2010; Day and Bonduriansky 2011; Longo et al. 2012; Klironomos et al. 2013). Therefore, genetic mutations may not be the only molecular factors to consider (Richards 2006, 2009). Indeed, epigenetic and genetic changes may jointly regulate genome activity and evolution, as recent evolutionary biology modeling suggests (Day and Bonduriansky 2011; Klironomos et al. 2013). This integration of genetics and epigenetics may improve our understanding of the molecular control of many aspects of biology, including evolution.

Supplementary Material

Supplementary tables S1–S6 are available at *Genome Biology and Evolution* online (http://www.gbe. oxfordjournals.org/).

Acknowledgments

The authors acknowledge the advice and critical reviews of Dr Jeb Owen (WSU), Dr Hubert Schwabl (WSU), Dr David Crews (U Texas Austin), Dr Kevin P. Johnson (U Illinois) and Dr Sarah Bush (Utah). They thank Ms Sean Leonard, Ms Shelby Weeks, Dr C. Le Bohec, Mr O. Tiselma and Mr R. Clayton for technical assistance and Ms Heather Johnson for assistance in preparation of the manuscript. The research was supported by the National Institute of Health grants to M.K.S. and National Science Foundation grants to D.H.C. M.K.S. conceived the study. M.K.S. and D.H.C. designed the study. C.G.B., M.M.H., E.E.N., J.A.H.K., S.A.K., and D.H.C. performed the experiments and acquired the data. All authors analyzed the data. M.K.S. and D.H.C. wrote the manuscript. All authors edited and approved the manuscript. The authors declare no competing financial interests.

Literature Cited

Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ. 2004. Bmp4 and morphological variation of beaks in Darwin's finches. Science 305(5689):1462–1465.

Akimoto K, et al. 2007. Epigenetic inheritance in rice plants. Ann Bot. 100(2):205–217.

Alcaide M, Edwards SV. 2011. Molecular evolution of the toll-like receptor multigene family in birds. Mol Biol Evol. 28(5):1703–1715.

Anway MD, Cupp AS, Uzumcu M, Skinner MK. 2005. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 308(5727):1466–1469.

Badyaev AV, Young RL, Oh KP, Addison C. 2008. Evolution on a local scale: developmental, functional, and genetic bases of divergence in bill form and associated changes in song structure between adjacent habitats. Evolution 62(8):1951–1964.

Bonduriansky R. 2012. Rethinking heredity, again. Trends Ecol Evol. 27(6): 330–336.

Clayton DF, Balakrishnan CN, London SE. 2009. Integrating genomes, brain and behavior in the study of songbirds. Curr Biol. 19(18): R865–R873.

Crews D, et al. 2007. Transgenerational epigenetic imprints on mate preference. Proc Natl Acad Sci U S A. 104(14):5942–5946.

Cubas P, Vincent C, Coen E. 1999. An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401(6749):157–161.

Day T, Bonduriansky R. 2011. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. Am Nat. 178(2):E18–E36.

Donohue K. 2011. Darwin's finches: readings in the evolution of a scientific paradigm. Chicago (IL): University of Chicago Press, p. 492.

Endler J. 1986. Natural selection in the wild. Princeton (NJ): Princeton University Press.

Feinberg AP. 2004. The epigenetics of cancer etiology. Semin Cancer Biol. 14(6):427–432.

Flatscher R, Frajman B, Schönswetter P, Paun O. 2012. Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation? Genet Res Int. 2012:698421.

Gazave E, et al. 2011. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. Genome Res. 21(10):1626–1639.

Geoghegan JL, Spencer HG. 2012. Population-epigenetic models of selection. Theor Popul Biol. 81(3):232–242.

Geoghegan JL, Spencer HG. 2013a. Exploring epiallele stability in a population-epigenetic model. Theor Popul Biol. 83:136–144.

Geoghegan JL, Spencer HG. 2013b. The adaptive invasion of epialleles in a heterogeneous environment. Theor Popul Biol. 88:1–8.

Geoghegan JL, Spencer HG. 2013c. The evolutionary potential of paramutation: a population-epigenetic model. Theor Popul Biol. 88:9–19.

Grant P, Grant R. 2008. How and why species multiply: the radiation of Darwin's finches. Princeton (NJ): Princeton University Press.

Greenspan RJ. 2009. Selection, gene interaction, and flexible gene networks. Cold Spring Harb Symp Quant Biol. 74:131–138.

Guerrero-Bosagna C, Sabat P, Valladares L. 2005. Environmental signaling and evolutionary change: can exposure of pregnant mammals to environmental estrogens lead to epigenetically induced evolutionary changes in embryos? Evol Dev. 7(4):341–350.

Guerrero-Bosagna C, Settles M, Lucker B, Skinner MK. 2010. Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. PLoS One 5(9):e13100.

Holeski LM, Jander G, Agrawal AA. 2012. Transgenerational defense induction and epigenetic inheritance in plants. Trends Ecol Evol. 27: 618–626.

Huber SK, et al. 2010. Ecoimmunity in Darwin's finches: invasive parasites trigger acquired immunity in the medium ground finch (*Geospiza fortis*). PLoS One 5(1):e8605.

Huttley GA. 2004. Modeling the impact of DNA methylation on the evolution of BRCA1 in mammals. Mol Biol Evol. 21(9):1760–1768.
Jirtle RL, Skinner MK. 2007. Environmental epigenomics and disease susceptibility. Nat Rev Genet. 8(4):253–262.
Klironomos FD, Berg J, Collins S. 2013. How epigenetic mutations can affect genetic evolution: model and mechanism. Bioessays 35(6): 571–578.
Koop JA, Huber SK, Laverty SM, Clayton DH. 2011. Experimental demonstration of the fitness consequences of an introduced parasite of Darwin's finches. PLoS One 6(5):e19706.
Kuzawa CW, Thayer ZM. 2011. Timescales of human adaptation: the role of epigenetic processes. Epigenomics 3(2):221–234.
Lack D. 1947. Darwin's finches. Cambridge University Press.
Lamarck JB. 1802. Recherches sur l'organisation des corps vivans. Paris: Chez L'auteur, Maillard.
Liebl AL, Schrey AW, Richards CL, Martin LB. 2013. Patterns of DNA methylation throughout a range expansion of an introduced songbird. Integr Comp Biol. 53(2):351–358.
Longo G, Miquel PA, Sonnenschein C, Soto AM. 2012. Is information a proper observable for biological organization? Prog Biophys Mol Biol. 109(3):108–114.
Lupski JR. 2007. An evolution revolution provides further revelation. Bioessays 29(12):1182–1184.
Macia A, et al. 2011. Epigenetic control of retrotransposon expression in human embryonic stem cells. Mol Cell Biol. 31(2):300–316.
Manikkam M, Guerrero-Bosagna C, Tracey R, Haque MM, Skinner MK. 2012. Transgenerational actions of environmental compounds on reproductive disease and epigenetic biomarkers of ancestral exposures. PLoS One 7(2):e31901.
Mundy NI. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc Biol Sci. 272(1573):1633–1640.
Nozawa M, Kawahara Y, Nei M. 2007. Genomic drift and copy number variation of sensory receptor genes in humans. Proc Natl Acad Sci U S A. 104(51):20421–20426.
Olshen AB, Venkatraman ES, Lucito R, Wigler M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5(4):557–572.
Petren K, Grand BR, Grant PR. 1999. A phylogeny of Darwin's finches based on microsatellite DNA lenght variation. Proc R Soc Lond B. 266(1417):321–329.
Picard F, Robin S, Lavielle M, Vaisse C, Daudin J-J. 2005. A statistical approach for array CGH data analysis. BMC Bioinformatics 6:27.
Pinkel D, Albertson DG. 2005. Comparative genomic hybridization. Annu Rev Genomics Hum Genet. 6:331–354.
Poptsova M, Banerjee S, Gokcumen O, Rubin MA, Demichelis F. 2013. Impact of constitutional copy number variants on biological pathway evolution. BMC Evol Biol. 13:19.
R Development Core Team. 2010. R: a language for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://www.R-project.org.
Rands CM, et al. 2013. Insights into the evolution of Darwin's finches from comparative analysis of the *Geospiza magnirostris* genome sequence. BMC Genomics 14:95.
Rebollo R, Horard B, Hubert B, Vieira C. 2010. Jumping genes and epigenetics: towards new species. Gene 454(1–2):1–7.
Richards CL, Bossdorf O, Pigliucci M. 2010. What role does heritable epigenetic variation play in phenotypic evolution? BioScience 60: 232–237.
Richards EJ. 2006. Inherited epigenetic variation—revisiting soft inheritance. Nat Rev Genet. 7(5):395–401.
Richards EJ. 2009. Quantitative epigenetics: DNA sequence variation need not apply. Genes Dev. 23(14):1601–1605.
Skinner MK. 2011. Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. Epigenetics 6(7): 838–842.
Skinner MK, Anway MD, Savenkova MI, Gore AC, Crews D. 2008. Transgenerational epigenetic programming of the brain transcriptome and anxiety behavior. PLoS One 3(11):e3745.
Skinner MK, Manikkam M, Guerrero-Bosagna C. 2010. Epigenetic transgenerational actions of environmental factors in disease etiology. Trends Endocrinol Metab. 21(4):214–222.
Skinner MK, Mohan M, Haque MM, Zhang B, Savenkova MI. 2012. Epigenetic transgenerational inheritance of somatic transcriptomes and epigenetic control regions. Genome Biol. 13(10):R91.
Skinner MK, Savenkova MI, Zhang B, Gore AC, Crews D. 2014. Gene bionetworks involved in epigenetic transgenerational inheritance of altered mate preference: environmental epigenetics and evolutionary biology. BMC Genomics 15:377.
Slatkin M. 2009. Epigenetic inheritance and the missing heritability problem. Genetics 182(3):845–850.
Stoltzfus A. 2012. Constructive neutral evolution: exploring evolutionary theory's curious disconnect. Biol Direct. 7:35.
Sudmant PH, et al. 2013. Evolution and diversity of copy number variation in the great ape lineage. Genome Res. 23:1373–1382.
Tang MH, et al. 2012. Major chromosomal breakpoint intervals in breast cancer co-localize with differentially methylated regions. Front Oncol. 2:197.
Tateno H, Kimura Y, Yanagimachi R. 2000. Sonication per se is not as deleterious to sperm chromosomes as previously inferred. Biol Reprod. 63(1):341–346.
Tibshirani R, Wang P. 2008. Spatial smoothing and hot spot detection for CGH data using the fused lasso. Biostatistics 9(1):18–29.
Ward WS, Kimura Y, Yanagimachi R. 1999. An intact sperm nuclear matrix may be necessary for the mouse paternal genome to participate in embryonic development. Biol Reprod. 60(3):702–706.
Whitlock MC, Schluter D. 2009. The analysis of biological data. Greenwood Village (CO): Roberts and Company Publishers.
WUSTL. 2008. Jul. 2008 assembly of the zebra finch genome (taeGut1, WUSTL v3.2.4), as well as repeat annotations and GenBank sequences, Database Provider, NCBI.
Ying H, Huttley G. 2011. Exploiting CpG hypermutability to identify phenotypically significant variation within human protein-coding genes. Genome Biol Evol. 3:938–949.

Associate editor: Bill Martin