# Weak parent-of-origin expression bias: Is this imprinting?

Carol A. Edwards[1*], William M. D. Watkinson[1], Stephanie B. Telerman[1], Lisa C. Hülsmann[1], Russell S. Hamilton[1] and Anne C. Ferguson-Smith[1*]

1 Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

*Corresponding authors: Anne C. Ferguson-Smith, Carol A. Edwards

afsmith@gen.cam.ac.uk, cae28@cam.ac.uk

## Abstract

In mouse and human, genes subjected to genomic imprinting have been shown to function in development, behaviour, and post-natal adaptations. Failure to correctly imprint genes in human is associated with developmental syndromes, adaptive and metabolic disorders during life as well as numerous forms of cancer. In recent years researchers have turned to RNA-seq technologies applied to reciprocal hybrid strains of mice to identify novel imprinted genes, causing a 3-fold increase in genes reported as having a parental origin specific expression bias. The functional relevance of parental origin-specific expression bias is not fully appreciated especially since many are reported with only minimal parental bias (e.g. 51:49). Here we present an in-depth meta-analysis of previously generated RNA-seq data and show that the methods used to generate and analyse libraries greatly influence the calling of allele-specific expression. Validation experiments show that most novel genes called with parental-origin specific allelic bias are artefactual, with the mouse strain contributing a larger effect on expression biases than parental origin. Of the weak novel genes that do validate, most are located at the periphery of known imprinted domains, suggesting they may be affected by local allele- and tissue-specific conformation. Together these findings highlight the need for robust tools, definitions, and validation of putative imprinted genes to provide meaningful information within imprinting databases and to understand the functional and mechanistic implications of the process.

## Introduction

Genomic imprinting is a mammalian-specific epigenetic process causing some genes to be expressed in a parent-of-origin specific manner, leading to the functional inequality of parental genomes. In the 35 years since imprinting was first discovered in mammals much has been learned about the mechanisms governing this process and the role epigenetic mechanisms as a whole play in genome function[1]. To date, approximately 150 imprinted genes have been identified in mice and humans where they have been shown to have vital roles in development, behaviour and post-natal adaptations[1–3]. Failure to correctly imprint genes in human is associated with developmental syndromes, adaptive and metabolic disorders during life as well as numerous forms of cancer[4,5].

Canonical imprinting is established during gametogenesis, when certain regions of the genome become differentially DNA methylated in the two germlines. Germline differentially methylated regions (gDMRs) associated with imprinted genes differ from others because they are protected from the global demethylation that occurs in the zygote; a process requiring the KRAB-zinc finger proteins Zfp57 and Zfp445[6,7]. Imprinted genes are frequently organised into clusters or pairs in the genome and a single gDMR acts as an imprinting control region (ICR) for the entire domain[8]. Secondary or somatic DMRs are also associated with imprinted genes: these DMRs are established after fertilisation, under the control of the ICR and are generally not bound by Zfp57 and or Zfp445. More recently Inoue *et al.* identified a germline-derived histone 3 lysine 27 tri-methylation (H3K27me3) mediated mechanism that also can confer parental origin specific expression. These "non-canonically" imprinted genes show paternally biased expression in the preimplantation embryo which persists in extra-embryonic tissue but is lost in the embryo proper[9].

Advances in RNA-sequencing technology and analysis pipelines have enabled quantification of allele specific expression and the identification of imprinted genes in reciprocal hybrids between distantly related strains of mice. Early studies using this method highlighted the challenges of analysing data derived from reciprocal hybrids[10–14]. For example, in 2010 two related works reported approximately

1300 new imprinted genes in the brain, which would have increased the number of known imprinted genes in the mouse by an order of magnitude[12,13]. These datasets were subsequently shown to include multiple false positives emphasising the need for validation of putative imprinted genes via a complementary method.[15].

More recent studies applied improved analytical tools, as well as additional controls, statistical filtering and validation[16–20]. The tissues interrogated in these studies vary and each identify a distinct set of genes with significant parental-origin-specific expression biases raising the possibility that the imprinting status of these new candidates might be cell or tissue-specific and hence overlooked to date. Particular attention has been paid to neurological tissues where a much higher proportion of putative imprinted genes were identified. Interestingly, most of the candidates have significant yet weak allelic biases including those with only 51-60% expression originating from the preferred allele. Here, applying the same analysis tools and validation approach to all datasets, we investigate genes with weak parental origin-specific bias further by focusing on four studies: Babak *et al.* 2015 (Dataset A), Bonthuis *et al.* 2015 (Dataset B), Perez *et al.* 2015 (Dataset C), Crowley *et al.* 2015 (Dataset D)[16–19]. It is noteworthy that the different studies have employed different analysis pipelines.

## Results

### Limited overlap between the sets of novel imprinting candidates from RNA-seq

The four studies used for our analysis generated whole transcriptomes from reciprocal hybrid mouse tissues to identify novel imprinted genes. Reciprocal mouse crosses take advantage of sequence specific polymorphisms to distinguish parental-origin effects from those caused by the genetic background of different mouse strains. A summary comparing the four studies is described in Supplementary information and summarised in Supplementary Table S1. We first assessed the overlap between the studies (Figure 1A and B). For this, all genes identified were compiled and those with alternative names between studies, merged. This produced a list of 313 genes of which only 36 (11.5%) were identified with allelic biases in all 4 studies (Supplementary Table S2). Genes were then divided into "Known" (had previously been identified as imprinted or validated using a non-RNA-seq method - Supplementary Table S2) or "Novel" (first identified in these studies or previously un-validated RNA-seq experiments). By our classification, Dataset A identified 99 genes with ASE, of which 26 were novel and 73 were known. Dataset B identified 209 genes, 151 novel and 58 known. Dataset C identified 112 genes, 51 novel and 61 known, and Dataset D identified 95 genes, 57 novel and 38 known. When the overlap between studies was assessed, 30 out of 87 (35%) known genes were identified in all 4 studies compared with six out of 226 novel genes (2.7%) (Figure 1A and B and Supplementary Table S3).

As imprinted genes tend to be clustered within the genome, novel genes were subdivided into three categories. Class-1 are novel singletons that are more than 1 Mb from another gene identified in any of the studies. Class-2 are novel clusters where two or more novel genes are within 1 Mb of each other but over 1 Mb from a known imprinted gene. Finally, Class-3 novel genes are within 1 Mb of a known imprinted gene. The degree of overlap of each of these classes of novel genes varies greatly with a much higher proportion of Class-3 genes overlapping. Indeed, the only six novel genes that came up in all four studies belong to Class-3 (Figure 1C) indicating that boundaries of some imprinting clusters may not have been fully defined previously.

The vast majority (82.7%) of novel parental-origin-specific biased genes were unique to one study, whereas only 28.7% of known genes fell in the same category (Figure 1D). The lack of overlap in novel genes suggests that the genes identified in these studies may be subjected to tissue specific imprinting. Most known genes only identified in one study come from Dataset A. This is not surprising since more tissues were analysed including four extra-embryonic tissues and six of the 14 unique known genes from this study have previously been shown to be specifically imprinted in placenta or yolk sac[14,21–25]. To investigate the effect of tissue specificity further, gene sets identified in different studies, but in the same tissue, were compared. Once again minimal overlap was observed between novel genes compared with known genes in the same tissues (Figure 1E and F). Thus, despite the careful analytical measures that appear to have been taken in all four studies to reduce false positives, the minimal overlap between the sets of novel genes suggests that some of these may indeed be false positives.

**Reanalysis of data from previous studies indicates differences between studies are due to experimental design and analysis.**

The lack of overlap between the four studies could be due to the different methods used, and we therefore decided to run the sequence data generated from three of the studies (which all used C57BL/6xCastEiJ reciprocal crosses) through the same analysis pipeline and see if the overlap was improved. We utilized the more recently established ISoLDE (Integrative Statistics of alleLe Dependent Expression) package[26] that had not been employed in any of the four studies to call allele-specific expression (ASE). ISoLDE uses a nonparametric statistical method to infer ASE in RNA-seq data from reciprocal crosses. It was benchmarked by the authors on six RNA-seq datasets including Datasets B and C used in this study and has been used by others to study imprinted gene expression in the mammary gland[27].

We chose hypothalamus, cerebellum, liver, muscle, and whole adult brain from Dataset A, the arcuate nucleus (ARN), the dorsal raphe nucleus (DRN), liver and muscle from Dataset B, and P8 and P60 cerebellum from Dataset C. First, we assessed the overlap between our calls and the calls from the original studies. 204 genes were identified across the five different tissues from Dataset A (Supplementary Table S4). 66% of total known genes and 17% of Class-3 genes overlap using their approach and ISoLDE (Figure 2A). For these data ISoLDE identified a high proportion of Class-1+2 genes (65.2%). No overlap is observed in Class-1+2 genes as none were identified in the original study. This high number of novel genes called by ISoLDE is most likely due to the absence of biological replicates in Dataset A as other datasets without biological replicates also have increased novel calls (data not shown).

For Datasets B and C this identified 86 and 63 total ASE genes respectively (Supplementary Table S4). For Dataset B, 49 known genes overlap which is 73% of all genes identified using both methods (Figure 2B). This is consistent across all four tissues analysed with between 50-71% of known imprinted genes overlapping in the tissues (Supplementary Table S5). Fewer novel genes show overlap between the two methods: 11% of Class-3 genes and 1% of Class-1+2 overlap (Figure2B and Supplementary Table S4). A similar pattern is observed in the Dataset C with 67% of known genes overlapping, 14% of Class-3 and 8% of Class-1+2 (Figure 2C). For both datasets fewer novel genes were identified using the common pipeline than in the original studies indicating that the individual methods used to identify allele specific genes in the four studies greatly influences the calling of novel imprinted genes. Only one Class-1+2 gene was found in each study: *Gm11410* in Dataset B and *BC034090* in Dataset C. This is a much lower proportion of Class-1+2 genes than was called in Dataset A (2.3% and 3.2% in datasets B and C respectively vs 65.2%) and is most likely due to the higher number of biological replicates sampled in Datasets B and C.

To gain further insight into how methodology affects calling, datasets derived from the same tissue from different studies, but analysed using our pipeline, were compared. We were able to compare data for postnatal liver, skeletal muscle and hypothalamus (ARN) between Datasets A and B and for cerebellum between the Datasets A and C. Nine out of 20 known imprinted genes identified in the two liver datasets overlapped (45% - Figure 4D), 20 of 29 in muscle (69% - Figure 2E), 37 of 53 in hypothalamus (70% - Figure 2F) and 24 of 49 in the cerebellum (49% - Figure 2G). Fewer overlaps were found for Class-3 genes: Two out of 5 genes overlapped between the muscle datasets (40% - Figure 2E), 14 of 36 in the hypothalamus (39% - Figure 2F) and 8 of 24 in the cerebellum (33% - Figure 2G). Only one gene was identified in the liver datasets (Figure 2D). No overlapping Class-1+2 genes were identified in any tissue (Figures 2D-G).

To see if a particular dataset caused the lack of overlap, we next incorporated data from another study (Andergassen *et al.* 2017 – Dataset E) that investigated imprinting across multiple tissues and time points in CastEiJ and FVB reciprocal crosses[20]. Liver and muscle data were run through our pipeline (Supplementary Table S4). The three-way comparison shows high overlap between the known genes identified in all three datasets (36-57%), variable overlap in Class-3 genes (0-17%) and no overlap between Class-1+2 called genes (Figure 2H and I). Together, these data show that the methods used to generate libraries, the number of biological replicates included, and methods used to call ASE all greatly influence the genes called.

We next investigated whether strain biases influenced novel gene calling and the impact of this on calling parental-origin specific expression bias. The overlap between strain specific genes called by ISoLDE and novel genes called by Datasets B and C was compared. Twenty-five genes called as ASE in Dataset B are called as strain specific by ISoLDE of which 20 are Class-1+2 (Supplementary Table S6) compared with only one overlapping parent-of-origin called gene, suggesting they were mis-called as imprinted in the original study. Five overlapping strain specific genes were called in the Dataset C: three of them in known imprinted regions and two known imprinted genes – highlighting that strain specific expression can also act on imprinted genes and needs to be considered when calling allele specific expression.

**Weakly biased Class-3 genes are peripherally located and preferentially expressed from the chromosome carrying the germline methylation mark**

One of the most surprising findings of the cited studies is that many novel imprinting candidates show only a weak bias in the expression of the parental alleles, in some cases only slightly different from an unbiased 50:50 expression ratio. Perez and colleagues grouped their genes according to the percentage expression from the preferred allele[19] and we used and extended this grouping to the gene sets from the other studies (Figure 3A and Supplementary Table S2). Overall, among the 311 known and novel candidate genes for which bias data was available, we observed a bimodal distribution with 169 genes in the 50-60% bias group and 95 genes in the 90-100% group (Figure 3A). This bimodal distribution had already been described by Perez and colleagues[19]. Notably, two thirds of known imprinted genes (77.3%) were in the 90-100% group while only 12.1% of novel candidates show this high expression bias. Most novel genes (73.1%) were found to have only weak biases with 50-60% expression coming from the preferred allele. 88.7% of Class-1 novel singletons and 83.9% of Class-2 genes in novel clusters fall within this weakly biased group (Figure 3A). Interestingly, Class-3 genes, close to known imprinted regions, also show a bimodal distribution: 53.7% fall in the 50-60% group and 23.2% in the 90-100% group but, only 9 genes (9.5%) show a 70-90% expression bias. The presence of Class-3 genes displaying strong imprinted expression again indicates that the full extent of imprinted expression at some imprinted clusters has not been fully established.

Next, we compared the direction of the parental bias among our four classes of genes. Both Class-1 and Class-2 genes had a relatively even split between preferentially paternal or maternal expression in both weak and strongly bias genes (Figure 3B). Interestingly, although the known imprinted genes consist of about equal proportions of maternally and paternally expressed genes (42 paternal, 33 maternal, 13 genes with both directions depending on tissue), paternally expressed genes tend to have higher expression biases (37 have 90-100% bias), while all eleven known genes with a low expression bias (50-70%) are maternally expressed (Figure 3B). Novel genes in known clusters (Class-3) follow the same trend of parental bias and direction (Figure 3D) with 70.3% weakly biased genes preferentially expressed from the maternal chromosome and 64.5% of highly biased genes expressed from the paternally inherited chromosome. Most imprinted regions are controlled by maternal, allele-specific methylation at the ICR acting as a repressed promoter for maternally repressed imprinted alleles. As a consequence of this direct repression, these genes show very strong paternal expression bias whereas paternally repressed genes within these same regions tend to rely on more indirect repression mechanisms such transcription of a long ncRNA and differential histone modifications[28,29]. To see if the direction of ICR methylation is causing the trend towards strong paternal and weak maternal biases, we categorised the genes in regions with known ICRs according to whether the preferentially expressed allele was on the chromosome with a methylated (Meth-ICR) or unmethylated ICR (Un-ICR) regardless of parental origin (Figure 3C). The direction of methylation does influence the bias as weakly biased genes tend to be preferentially expressed from the chromosome with the Meth-ICR (81% of Class-3 and 100% known <70% biased) whereas strongly biased genes tend to be preferentially expressed from the chromosome with the Un-ICR (80% of Class-3 and 54% known >70% biased -Figure 3C).

To investigate this trend further, we assessed the DNA methylation and H3K27me3 at the promoter regions of all the Class-3 and known genes at canonically imprinted clusters using previously published embryonic and postnatal data[30–32] (Figure 3D). Strongly biased known genes on the Un-ICR chromosome were associated with promoter methylation suggesting direct regulation by DNA methylation, whilst strongly biased known genes on the Meth-ICR chromosome tend to have unmethylated promoters but higher levels of H3K27me3 indicating repression of genes on the Un-ICR chromosome is controlled by differential histone modifications rather than differential promoter methylation. Interestingly, strongly biased Class-3 genes on the Un-ICR chromosome follow the same trend as the known genes and have methylated promoters indicating a shared mechanism, whereas strongly biased genes preferentially expressed from the Meth-ICR show low promoter methylation and H3K27me3 occupancy. Weakly biased genes tend to be associated with unmethylated promoters and low H3K27me3 regardless of which allele is preferentially expressed. This is unsurprising as the biases are so weak, however the trend for low biased genes to be preferentially expressed from the Meth-ICR allele implies the genic environment is more repressive on the Un-ICR chromosome perhaps due to ectopic spreading of repressive marks in some cells.

Next, Class-3 genes were classified by whether they were flanked by previously known imprinted genes or were peripherally located within known imprinted clusters. Weakly biased novel genes are predominantly found at the periphery of annotated imprinted domains: 83% of peripheral Class-3 genes (60/72) have a bias below the 70:30 canonical imprinting threshold set by Andergassen et al 2017 (Figure 4A). Conversely, highly biased Class-3 genes tend to be flanked by known imprinted genes: only 13% (3/23) of novel genes flanked by known imprinted genes have a ratio below 70:30

(Figure 4A). Most of the twenty flanked, highly biased Class-3 genes belong to just two imprinted domains. Fourteen map between *Ndn* and *Snrpn* on chromosome 7 (Figure 4C) and five map downstream of *Meg3* on chromosome 12 (Figure 4D). In both cases these highly biased genes follow the direction of imprinting of the long non-coding gene in the region that is repressed on the meth-ICR chromosome suggesting that these may be poorly annotated RNAs arising from known poly-cistronic imprinted transcripts.

We next integrated the position and strength of bias with the ICR status on the chromosome preferentially expressing the gene and found over half of all Class-3 genes are weakly biased, at the edge of the known cluster and preferentially expressed from the chromosome that carries the methylated copy of the ICR (Figure 4B). This suggests that secondary repressive mechanisms acting at imprinted genes on the unmethylated chromosome are exerting a small effect on the expression levels of a gene at the periphery of the cluster. Perez et al. previously showed that the degree of bias is reduced as a function of distance from the most strongly bias gene in the cluster indicating that the influence of ICRs diminish over distance. To test whether the ICRs control these weak biases we used the Dlk1/Dio3 region as a model since it has strongly biased Class-3 genes in the centre of the cluster and nine weakly biased genes were called at the periphery (Figure 4D). Utilizing a previously described mouse model with a deletion of its ICR (IG-DMR) that causes a maternal-to-paternal epigenotype switch when maternally inherited[33], we looked to see if the biased expression of peripheral genes was lost when the ICR was removed. Female mice heterozygous for the deletion were crossed with castaneus males to allow allele-specific expression to be determined by pyrosequencing. *Dync1h1* which is located on the distal side of the region and was reported to be paternally biased in two of the original studies shows fully biallelic expression in E15.5 brain from maternal heterozygotes and wildtype littermates (Figure 4E). Conversely, two genes proximal to the defined cluster, *Wars* and *Wdr25*, both show a weak paternal expression bias in wildtype brains that is significantly reduced in mice with the IG-DMR deletion (Figure 4E). This indicates that the weak biases seen at the edges of imprinted clusters are regulated by the ICRs and may be innocent bystanders of the different epigenetic environments established by the ICR on each chromosome.

### Weakly biased genes close to known imprinted domains are more likely to experimentally validate than novel genes elsewhere in the genome

The lack of overlap between studies in the same tissue (Figure1F and Figure 2D-I) implies a high number of false positives. To test if these parent-of-origin biases are real we performed quantitative RT-PCR and pyrosequencing validation on eight Class-1, fifteen Class-2 genes (representing five novel clusters) and twenty Class-3 genes identified close to seven known imprinted domains. These genes included those that overlapped between the original studies, and those called as ASE or 'Undetermined' by ISoLDE that overlapped with an original study (Figure 2A-C, Supplementary Figure S1 and Supplementary Tables S3 and S7). Six known imprinted genes were tested as controls: the robustly imprinted *Dlk1* and *Peg3*, the more moderately biased *Mcts2* and *Herc3* and the extra-embryonic tissue-specific imprinted genes *Gab1* and *Ampd3* [22,34]. Thirteen tissues were analysed from three different timepoints e16.5, P7 and P60. Expression was called as biallelic if the mean of the paternal expression from both the C57BL/6 x CastEiJ and CastEiJ x C57BL/6 crosses was between 45 and 55% (Table 1).

**Class-1:** Of the eight novel singletons (Class-1) tested, four were biallelically expressed in all tissues and indeed three were expressed at very low levels. One gene, *Nhlrc1*, showed a consistent paternal bias across all postnatal brain tissues (Table 1 and Figure5A). It is noteworthy that this was the only Class-1 gene that was called in three of the original studies (Figure 1C and Supplementary Table S3). Nhlrc1 lies 1.3kb downstream from a known germline DMR and a Zfp57 binding site suggesting a plausible mechanism for biased gene expression at the region[35,36]. To see if this DMR persists postnatally, bisulfite pyrosequencing was performed on DNA from P7 cerebellum and liver. In cerebellum the DMR is partially retained: the maternal allele is hypermethylated compared with the paternal allele in both BxC and CxB crosses (Figure 5B). However, in liver where *Nhlrc1* shows no parental bias (Supplementary Figure 2A), both alleles are hypermethylated (Figure 5C), indicating *Nhlrc1* expression is parentally biased only in tissues where the DMR is retained. This suggests a mechanism whereby a gDMR can influence tissue specific imprinting postnatally.

**Class-2:** Four of the fifteen Class-2 genes tested fell below the expression threshold (Table 1). Of the other eleven, six were biallelic in all tissues tested, two showed biased expression in postnatal brain (*Pcdhb12* and *Wnk4*) and three had weak maternal bias in the placenta (*Vat1, Rtn3* and *Pla2g16*) (Table 1). *Pcdhb12* shows preferential expression from the maternal allele in all postnatal tissues which is consistent with the Datasets B and C (Figure 5D). This gene encodes protocadherin beta-12 and is in a cluster of protocadherin genes on chromosome 18 including *Pcdhb10* and *Pcdhb20* which were also called as biased in the original studies. Both of these genes were also tested: *Pcdhb10* was only expressed at very low levels and *Pcdhb20* showed biallelic expression (Table 1).

The other Class-2 gene which validated in postnatal brain is *Wnk4*. This gene forms a novel cluster with *Vat1*, *Tmem106a* and *Rdm1* on chromosome 11 that spans approximately 375 kb. Both validating tissues exhibit a weak bias. Indeed, *Wnk4* has a much stronger strain bias in both tissues in C57BL/6xCastEiJ hybrids which may be confounding the data (Supplementary Figure S3). Further analysis of this gene in other reciprocal hybrid strains is necessary to confirm the nature of its bias. One of the other genes in the *Wnk4* novel cluster, *Vat1*, was one of three Class-2 genes that validated with maternal expression bias in the placenta (Figure 5E). The other two genes *Rtn3* and *Pla2g16* form a novel cluster with *Prdx5* (which is biallelic in all tissues tested*)* on Chromosome 19. All three genes with the maternal placental bias were originally called as being biased in neural tissue where no evidence for biased expression was found.

**Class-3:** The twenty peripheral Class-3 genes assessed by allele specific cDNA pyrosequencing had all been called as biased in brain tissues in the original studies and had maximum biases below 70%, except for *Ago2* which has a maximal maternal bias of 79.3%. In contrast with the putative biased genes identified elsewhere in the genome, we found those peripheral to known imprinted clusters were more likely to validate. Nine of the 20 tested genes validated in somatic tissues (*Adam23, Cox4i2, Bcl2l2, Tpx2, Smim17, Ifitm10, Wars, Wdr25* and *Ago2* - Table 1). Of these *Bcl2l1* and *Ago2* showed a bias in all neural tissues tested (Supplementary Figures S4 and S6). *Bcl2l1*, along with *Cox4i2* and *Tpx2* is located close to the known imprinted gene *Mcts2.* All three genes showed a paternal bias in at least five neural tissues. Interestingly, *Tpx2* validated in all P60 tissues but not in the E16.5 or P7 material (Table 1 and Supplementary Figure S4) suggesting the bias strengthens over time postnatally.

Of the six genes tested that are located close to the *Peg3* cluster, four could not be tested due to low expression. In contrast, *Smim17* (*Gm16532*) is preferentially expressed from the maternal allele in five neural tissues (Supplementary Figure S5). *Smim17* bias is strongest in the P7 hypothalamus (64.7%) but is reduced to 56.7% by P60. A maternal bias is also detected in P7 hippocampus and brain stem but is lost by P60, together indicating the *Smim17* bias reduces over time (Table 1).

Six genes were tested that are located at the periphery of the *Dlk1*/*Dio3* cluster on Chromosome 12. *Wdr25* was biallelically expressed in all tissues except the P60 hypothalamus where 59.7% of expression is from the paternal allele (Figure 5D). Unlike *Wdr25*, *Wars* expression was consistently higher from the paternally inherited allele in all somatic tissues however, the bias was only above the 55% cut-off in two tissues: e16.5 brain (55.1%) and P7 hypothalamus (56.2%) (Figure 5C). The most peripheral genes tested on both the proximal (*Evl*) and distal (*Ppp2r5c* and *Dync1h1*) side of the cluster were called as paternally biased in the original studies but biallelic in all tissues we assessed Figure 5A and F and Table 1). *Slc25a29* was called as paternally biased in ARN and DRN in Dataset B, but we found it to be biallelic in all neural tissues. Taken together our data suggest the weak biases observed at the periphery of known imprinted domains are tissue- and stage-specific.

Interestingly*, Slc25a29* showed a very strong maternal bias in the placenta (81.1% - Figure 5B). To determine if this biased expression is regulated by the *Dlk1*/*Dio3* imprinting control region, we again made use of the IG-DMR knockout mouse model[33], Male and female mice heterozygous for the IG-

DMR deletion were crossed with CastEiJ mice and placentas were collected at E15.5. Allele specific pyrosequencing revealed maternal and paternal heterozygotes to both the same degree of maternal bias as wildtype litter mates indicating that Slc25a29 imprinting in the placenta is not under the control of the IG-DMR (Figure 5G).

## Discussion

In order to discuss weak parent-of-origin expression bias, it is first necessary to define canonical imprinting. Historically, imprinted expression was defined as monoallelic but as the sensitivity of methods quantifying ASE have improved it has become apparent that many imprinted genes show low level expression from the repressed allele. Within a single canonically imprinted cluster, the extent of bias can vary greatly between genes. For example, in the *Dlk1/Dio3* region >98% of *Meg3/Gtl2* expression arises from the maternal allele and 84-99% of *Dlk1* expression arises from the paternal chromosome, but for *Dio3*, 80% of expression is from the paternal allele in the embryo [37]. However, *Dio3* is still considered to be an imprinted gene as this bias is established by the ICR for this region and when the ICR is deleted from the maternal chromosome imprinting is lost from the entire region and Dio3 shows 50:50 expression from both alleles[33].

We therefore propose to define "canonical imprinting" as genes that are expressed predominantly (>70%) from one allele in a parent-of-origin specific manner in at least one tissue. We have taken a 70% cut-off as this was used previously by others[20]. Canonical imprinted genes are also under the control of a differentially methylated ICR that is established during gametogenesis with expression returning to biallelic levels or completely lost on perturbation of the ICR. Genes with a weak bias at the periphery of known imprinted domains that lose the expression bias on loss of the ICR we term "weak canonical imprinting". The term "non-canonical imprinting" is often used to refer to those genes for which imprinting is established in the germline via regions of differential of H3K27me3 (DMKs) leading to tissue-specific imprinting in the extra-embryonic compartment, such as *Sfmbt2, Smoc1 and Gab1*[9,22,38]. Finally, genes with a weak bias of less than 70% in tissues and with no known mechanism for the establishment of the bias we call "parental-origin-specific biased genes".

Although many parental-origin-specific biased genes were identified in the brain in the original studies, only 2 of the Class-1 or 2 genes tested here validated in more than two tissues (*Nhlrc1* and *Pcdhb12*). Our analysis also revealed little overlap of novel weakly biased genes between the four datasets detailed above, even when compared between the same tissues. This lack of reproducibility could be due to several factors. Firstly, expression levels can influence ASE calling: the lack of read depth in lowly expressed genes may erroneously lead to genes being called as biased, because a small difference in read numbers produces larger bias in weakly expressed transcripts.  Of the twenty-three Class-1 or 2 genes tested by pyrosequencing, seven could not be confirmed due to low expression levels. Secondly, the number of biological replicates can greatly influence the number of biased genes called. We found that analysis of datasets with only one or two replicates leads to a higher number of false positive novel biased genes being called compared with datasets with >6 replicates (Figure 3). Thirdly, the influence of genetic background on gene expression needs to be considered. We found a greater overlap between the original studies with genes called as strain biased by ISoLDE than those called with a parental bias. Moreover, 21 of the genes we experimentally tested had a strain bias in two or more tissues compared with 16 genes with parental origin bias in one or more tissues. Finally, the method for calling ASE greatly affects the results: minimal overlap was observed between the same data analysed by two different methods (Figure 3A-C). It is noteworthy that the two Class-1 or 2 genes that experimentally validated in most tissues were both called by two or more of the original studies suggesting a more accurate picture of ASE may be achieved by combining two calling methods. Taken together, these observations highlight the need for careful planning before embarking on projects to call parental-bias, stringent filtering for low expression and strain effects, and validation via an independent method to confirm findings.

Most of the unique novel genes identified in the original studies are in regions where imprinting has not previously been identified. Although only six of the 23 we tested validated experimentally (three in placenta) there may be others that display a true parent of origin bias. None of the Class-1 and 2 genes that validated show a bias greater than 60:40. Such weak biases can be explained in different ways (Figure 7). Firstly, a weak bias is seen in each cell in the population (Figure 7A). Secondly, bias may be due to true imprinting in a subset of cells within the population which could be random, clonal or cell type specific (Figure 7B and C). If this is due to canonical imprinting, we would expect to see a DMR. Indeed, at the *Nhlrc1* locus we found that a previously reported Zfp57 bound, germline DMR[35,36] is partially retained in postnatal neural tissues indicating that its expression bias could be regulated by canonical imprinting methods. We assessed the genomic intervals with other validated novel genes for germline DMRs (gDMRs) or Zfp57 binding using previously published data[36,39]. No gDMRs or Zfp57 binding sites were reported in the *Pcdhb* cluster. In the novel cluster between *Prdx5* and *Pla2g16* there are 3 oocyte-methylated gDMRs[39]. However, none of them overlap with the biased genes or a Zfp57 bound region[36]. Interestingly, there are 2 oocyte-methylated gDMRs in the interval between *Wnk4* and *Rdm1*, both of which overlap with *Wnk4*: one over the promoter and one over exons 5 and 6 (Supplementary Figure S3)[39]. As *Wnk4* is maternally biased it is hard to reconcile with methylation at the maternal promoter since most maternally methylated promoters lead to repression of the maternal allele. However, the second DMR overlaps with a lncRNA (*Gm11615*): further investigation is needed to establish if this transcript is reciprocally biased and offers a possible mechanism for the observed *Wnk4* bias. Finally, bias could be due to a skew towards activating one allele in random monoallelic expression (Figure 8D). The 22 *Pcdhb* isoforms exhibit allelic exclusion and are mono-allelically expressed in a stochastic and combinatorial fashion in individual neurons so each cell expresses two genes from the maternally inherited chromosome and two from the paternal copy [40]. However, *Pcdhb12* showed a maternal bias in all postnatal tissues tested. This may reflect interindividual random bias: but, of the 60 tissue samples we tested 54 showed a maternal bias in *Pcdhb12* and only 6 showed a paternal bias, indicating there are more neurons expressing the maternally inherited copy of the gene than randomly expected. This is not imprinting *per se* as does not reflect a bias in single cells but rather a population bias within the brain of individuals.

Most novel genes called in more than one study reside in or close to known imprinted regions. These genes were also more likely to validate (11 out of 20) but again, it is not known on a population level if these biases occur in every cell or are the result of complete imprinting is a subset of cells. We find that, parental bias is stronger for known and novel genes *within* imprinted clusters than those at the periphery of the cluster. At the periphery, biased genes show temporal and tissue specific changes: *Tpx2* only validates in P60 samples whereas *Smim17*'s bias decreases at this later stage showing biases are dynamic. We have found that weak Class-3 genes are more likely to be repressed on the chromosome with the unmethylated ICR where repression of known imprinted genes relies on repressive histone marks. Perez et al. previously showed that biases decrease towards the edge of imprinted clusters. Together, these observations suggest the influence of ICRs diminishes with distance and that differences in local environment on each chromosome may play a role in biased gene expression. One possibility is that differential 3D architecture between the two chromosome is contributing to the bias. It has long been known that the *Igf2*/*H19* cluster shows different conformation on each parental chromosome[41] and a more recent single cell 3D analysis found eight known imprinted regions have parental-origin-specific TADs in postnatal cortex and hippocampus[42]. TADs are flanked by convergent CTCF sites and were originally believed to be stable domains that demarcate regions of shared regulation[43,44]. However, recent findings show CTCF binding is dynamic[45] and the boundaries of TADs can vary between cells[46,47]. It is also known that expression of an imprinted ncRNA from one chromosome can recruit polycomb repressive complex to repress other genes in cis[48]. It is therefore possible that in imprinted regions, parental chromosome specific conformations are affecting the expression of peripheral genes either by preventing the spread of repressive marks to the periphery or by reducing the enhancer/silencer interactions on the weaker allele (Figure 7 E-H). Such mechanisms could bring about weak bias in all cells or complete imprinting in a subset and are likely locus dependent; higher resolution analysis at the end of clusters and in single cells is needed to address this.

We have shown that the number of novel imprinted genes in the genome has been over-estimated by RNA-seq. However, some weakly biased genes do exist. We confirmed the weakly biased imprinting of the novel singleton *Nhlrc1* and many Class-3 genes. What is the functional relevance of such weak biases? As strain bias has a greater effect on expression than parent-of-origin bias then imprinting per se, may not be functionally relevant, especially if the bias is occurring within each cell. If on the other hand the bias is due to canonical imprinting in a subset of cells, then it may be important. One known weak canonically imprinted gene is *Th*[34], which was picked up by two of the studies[17,18] and shows 50-60 % maternal expression bias. Importantly, this gene was found to be monoallelic in a subset of cells and be associated with a behavioural phenotype depending on maternal or paternal inheritance[17]. It is therefore possible that other weakly biased genes are robustly imprinted in certain cell types and where their monoallelic expression is functionally important. Indeed, the *Bcl2l1* is a weak paternally biased gene in the *H13/Mcts2* imprinted domain, deletion of which leads to loss of certain neuron types and reduction in brain mass upon paternal but not maternal transmission[19]. Thus, genomic imprinting is not as widespread in the genome as recent studies claimed. However, at the periphery of known imprinted clusters weak parent-of-origin specific bias appears to be conferred to some genes. Whether all these weakly bias genes are functionally relevant or if some are simply innocent bystanders of local allele- and tissue-specific environments remains to be established.

## Methods

**Meta-analysis of previous data.** Co-ordinates from the Datasets A and B were converted to mm10 using the LiftOver function at UCSC (Kent et al). Any overlapping genes with different names were individually assessed and merged if found to represent the same transcript. Overlaps between studies were then assessed.

The maximum bias from each gene was assigned into one of 5 bins (50-60, 60-70, 70-80, 80-90, 90-100). For Dataset C these were taken straight from elife-07860-supp1-v2.xlsx (Table G). For the Dataset A read counts for the reciprocal crosses were taken from the original source data. The bias for each gene in every tissue was calculated by:

$$Paternal\ reads\ (BxC + CxB)/Total\ reads\ (BxC + CxB)$$

or

$$Maternal\ reads\ (BxC + CxB)/Total\ reads\ (BxC + CxB)$$

For Dataset B the maternal bias for each gene in every tissue was calculated as:

$$(Fold\ Change\left(\frac{Maternal}{Paternal}\right)/1 +\ Fold\ Change\ \left(\frac{Maternal}{Paternal}\right))\ x\ 100$$

For Dataset D the maximum bias was calculated by taking the mean paternal and maternal bias from each reciprocal cross to eliminate strain biases.

For promoter analysis the 500 bp upstream of the transcriptional start site was taken for each Class-3 and known imprinted gene. Methylation levels for fetal and male 6 week frontal cortex[30,49] and Histone H3 lysine 27 trimethylation (E16 and P0 forebrain)[31,32,50] were extracted using UCSC Table Browser[51]. The mean level across the 500bp interval was calculated then heatmap produced using ggplot2[52].

**Allele Specific Expression analysis**. Raw sequencing read FASTQ files were downloaded from EMBL-EBI European Nucleotide Archive for each of the RNA-seq data sets [16–20]. Low quality bases and adapters were removed with trim_galore (v0.4.1) (Babraham Bioinformatics - Trim Galore!). SNPSplit (v0.3.4) [53] was used to separate reads by parent of origin, which first required the preparation of allele-specific reference genomes for C57BL6/CAST_Eij and CAST_Eij/FVB (based on C57BL6) with the following commands SNPsplit_genome_preparation --vcf_file mgp.v5.merged.snps_all.dbSNP142.vcf.gz --reference_genome GRCm38_fasta/ --strain CAST_EiJ and SNPsplit_genome_preparation --vcf_file mgp.v5.merged.snps_all.dbSNP142.vcf.gz --reference_genome GRCm38_fasta/ --strain CAST_EiJ --strain2 FVB_NJ --dual_hybrid. VCF files for strain specific SNPs were obtained from www.sanger.ac.uk/data/mouse-genomes-project.

The Clusterflow pipeline tool was used to enable running multiple jobs in parallel across multiple processors on an HPC, however all scripts are also compatible with running on a single processor [54]. Trimmed reads were aligned to either the C57BL6/CAST_Eij or CAST_Eij/FVB reference genomes using HiSat2 (v2.1.0) [55], run via the hisat2 ClusterFlow module. Aligned reads were then name sorted to be compatible with SNPSplit, run via the samtools (v1.9) [56] Clusterflow module. Aligned files were run through SNPSplit to produce separated parent-specific alignment files using the SNP files produced by the genome preparation: all_SNPs_CAST_EiJ_GRCm38.txt.gz or all_FVB_NJ_SNPs_CAST_EiJ_reference.based_on_GRCm38.txt.gz. A custom Clusterflow module was created for SNPsplit [SNPSplit.cfmod]. Gene counts from each parent-specific alignment BAM file produced by SNPSplit were calculated using featureCounts (v1.5.0-p2) [57] via Clusterflow. A custom Rscript DESeq2_featureCounts_2_CountsTables.R is used to make a single counts table, including normalised reads, from the individual featureCount files. All scripts to reproduce the analysis are freely available at github.com/darogan/ASE_Meta_Analysis.

The counts tables were then manipulated into the configuration needed for the ISoLDE R package [26]. ISoLDE was used to test both allele specific parental and strain biases. The default resampling method was used with nboot = 3000 for datasets with more than 2 replicates. For datasets with two replicates or fewer the threshold method was used [26]. Overlaps were then identified between the different datasets.

**Mice.** All animal procedures were subject to local institutional ethical approval and performed under a UK Government Home Office license (project license number: PC213320E). Reciprocal crosses of the mouse lines C57BL/6J and castaneus (CAST/EiJ) were generated. For loss of IG-DMR studies mutant mice were maintained on a C57BL6/J background by crossing males heterozygous for the deletion with wild-type females. For expression analysis female or male heterozygotes were mated to castaneus (CAST/EiJ) mice to generate IG-DMR deletion heterozygotic conceptuses and wildtype littermates.

**Tissue collection**. Samples were harvested from three to four mice from each cross at three developmental stages: e16.5 (16.5 days after conception), P7 (7 days after birth) and P60 (60 days after birth). At e16.5, whole brain, liver and placenta were harvested. At P7 and P60, brain stem, cerebellum, cortex, hippocampus and hypothalamus were harvested. Samples were snap-frozen in liquid nitrogen and stored at -80 degrees. For loss of IG-DMR studies samples were collected at E15.5.

**qRT–PCR.** DNA and RNA were extracted from samples using AllPrep DNA/RNA Mini kit (Qiagen) according to the manufacturer's instructions.  5 µg RNA was treated with DNaseI (Thermo Scientific) as per the manufacturer's instructions. 1 µg RNA was reverse transcribed with Revert Aid RT (Thermo Scientific). Assays were designed using PyroMark Assay Design SW 2.0 (Qiagen) and provided in Table S9. The annealing temperature for each primer set was optimized by gradient PCR. The qPCR reactions were run on a LightCycler 480 (Roche) with Brilliant III Ultra-Fast SYBR Green qPCR Master Mix and the following conditions: 95$^o$C for 5 min followed by 45 cycles of 95$^o$C - 10 s, specific annealing temperature - 10 s, 72$^o$C - 10 s. All reactions were run in duplicate, and the relative expression of each gene calculated using the ΔCt method and normalised to the housekeeping gene *Tbp*. Genes with expression lower than 0.05 times that of the housekeeping gene *Tbp* after qRT-PCR were not analysed further as weak expression leads to inconsistent results between technical replicates in pyrosequencing.

**Allelic expression analysis.** Parental allelic expression quantification was performed by pyrosequencing. Streptavidin Sepharose High Performance beads (GE healthcare) dissolved in binding buffer (10 mM Tris-HCL pH7.6, 2 M NaCl, 1 mM EDTA, 0.1% Tween-20) were shaken with the qPCR product at 1,400 rpm for 20 min. The biotinylated strand was purified using a PyroMark Q96 Vacuum Workstation (QIAGEN) then sequencing primers annealed in annealing buffer (20mMTris-acetate pH7.6, 2 M magnesium acetate) at 85$^o$C for 3 min. Sequencing was performed on a PyroMark Q96 MD pyrosequencer (Qiagen) using PyroMark Gold Q96 Reagents (Qiagen). The mean expression bias from 3 or 4 biological replicates of tissue was then calculated. Genomic DNA was also assessed to identify any amplification bias from the primers. If the amplification bias was greater than 53:47, the cDNA values for each sample was corrected to the mean gDNA bias of the stage.

**Clonal Methylation Analysis**. DNA (0.5-1µg) was bisulfite treated using the two-step protocol of the Imprint DNA Modification Kit (Sigma). converted DNA was amplified using primers Fwd 5' TTGATGGAGTAAAAGGAATTGTTTTAGG and Rev 5' CCAATTCAAAAATTTAAAAAAAACAAAACC with HotStarTaq DNA Polymerase (QIAGEN). The PCR conditions were: 1) 95°C – 5 min; 2) 94°C – 30 s, 55°C – 30 s, 72°C – 55 s, 40 cycles; 3) 72°C – 5 min. PCR Products were run on an 1.5% agarose gel, bands were then cut out and DNA was extracted using MinElute Gel Extraction Kit (QIAGEN). Purified DNA was ligated into the pGEM®-T Easy Vector and transformed into Stellar™ Competent Cells (Cat# 636766). Selected colonies were Sanger sequenced by GENEWIZ.

## Figures and Tables

Figure 1

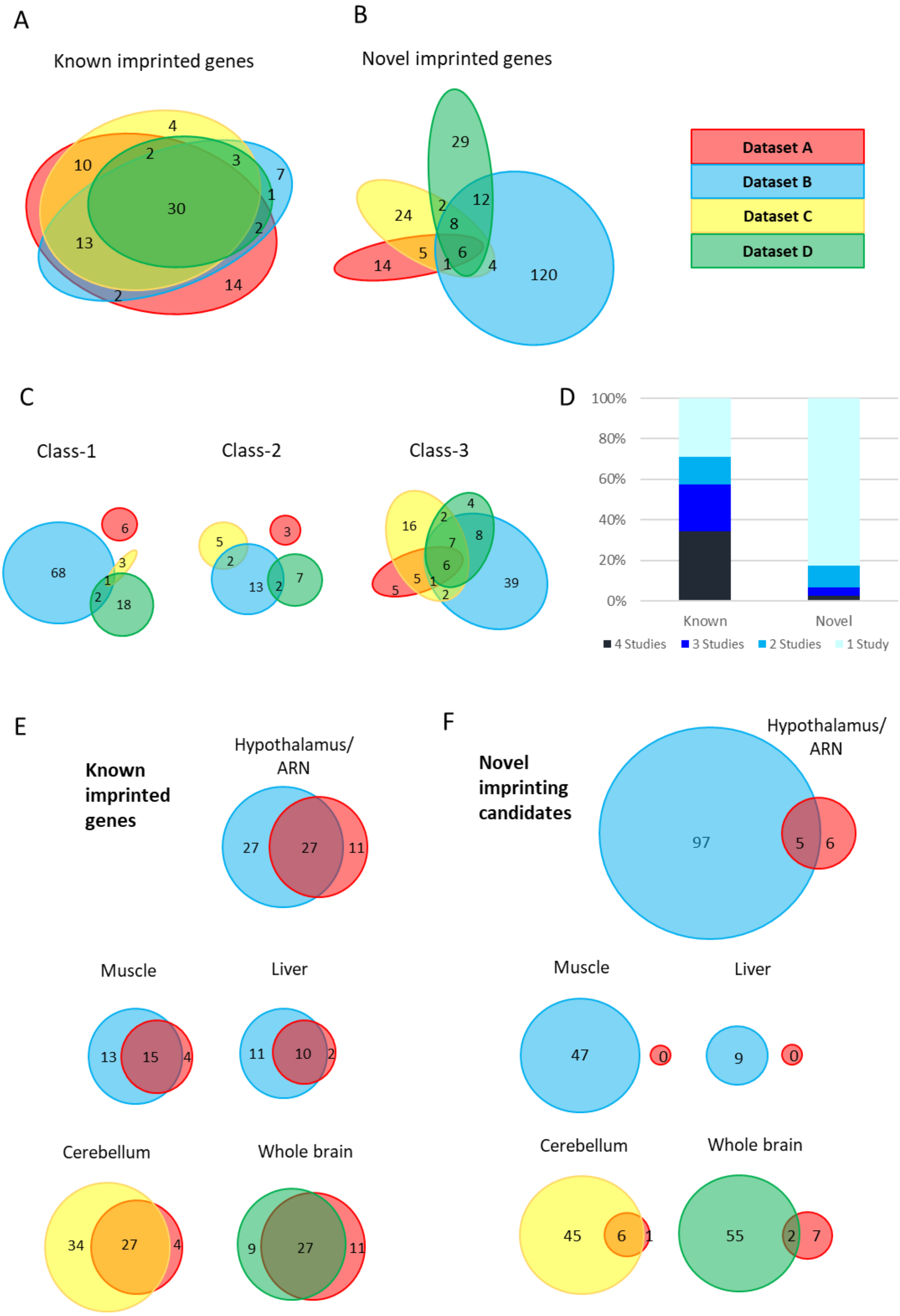


Figure 1 - limited overlap between novel genes identified in the four studies

(A-B) Euler diagrams showing overlap of A - known imprinted genes and B – novel biased genes between the four studies. (C) – overlaps between different classes of novel genes. Class-1 = novel singletons >1Mb from another gene identified in any of the studies. Class-2 = novel clusters where two or more novel genes within 1 Mb of each other but over 1Mb from a known imprinted gene. Class-3 = novel genes are within 1 Mb of a known imprinted gene. (D) - proportion of known and novel genes by number of studies they were identified in. (E-F) Venn diagrams showing tissue specific overlaps in E - known imprinted genes and F – novel biased genes.

Figure 2

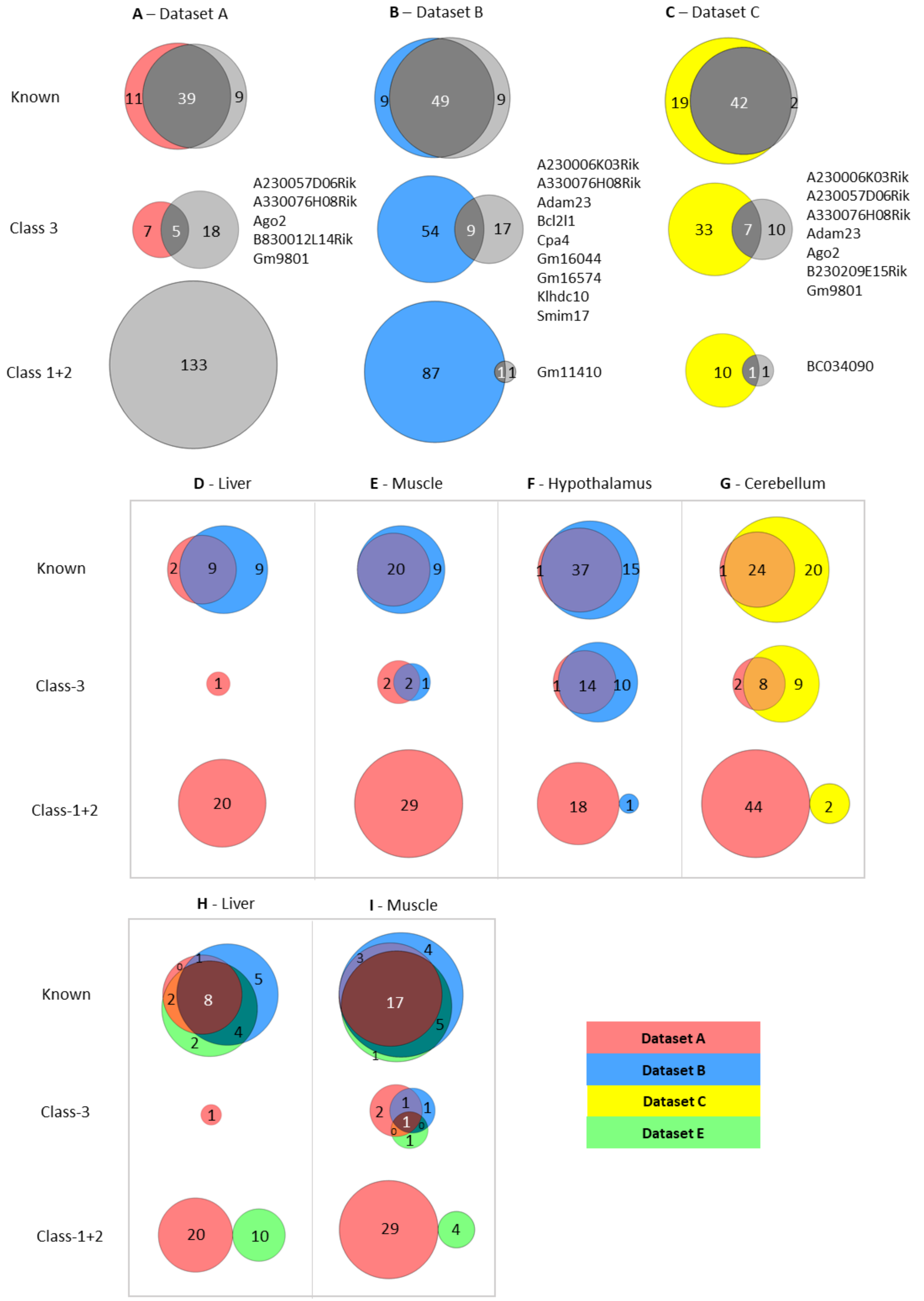

A – Dataset A
B – Dataset B
C – Dataset C
Known
11
39
9
9
49
9
19
42
2
Class 3
7
5
18
A230057D06Rik
A330076H08Rik
Ago2
B830012L14Rik
Gm9801
54
9
17
A230006K03Rik
A330076H08Rik
Adam23
Bcl2l1
Cpa4
Gm16044
Gm16574
Klhdc10
Smim17
33
7
10
A230006K03Rik
A230057D06Rik
A330076H08Rik
Adam23
Ago2
B230209E15Rik
Gm9801
Class 1+2
133
87
1
1
Gm11410
10
1
1
BC034090
D - Liver
E - Muscle
F - Hypothalamus
G - Cerebellum
Known
2
9
9
20
9
1
37
15
1
24
20
Class-3
1
2
2
1
1
14
10
2
8
9
Class-1+2
20
29
18
1
44
2
H - Liver
I - Muscle
Known
0
1
5
2
8
4
2
3
4
17
5
1
Class-3
1
2
1
1
1
0
0
1
Class-1+2
20
10
29
4
Dataset A
Dataset B
Dataset C
Dataset E

Figure 2 - Limited overlap between novel genes called by different analysis pipelines.

(A-C) Venn diagrams showing overlap between allelic biased genes called by our pipeline with the ISoLDE package (grey circles) versus the original studies. A – Dataset A, B - Dataset B, C - Dataset C. Overlapping novel genes are listed to the right. Dataset A only genes called in hypothalamus, cerebellum, liver, muscle and whole adult brain were analysed. (D-G) Venn diagrams showing the overlap between allelic biased genes called by our pipeline from sequence data generated from the same tissue by different studies: liver (D), muscle (E), hypothalamus (F) from Dataset A and Dataset B and cerebellum (G) from Dataset A and Dataset C. (H-I) Venn diagrams of 3-way overlap between allelic biased genes called by our pipeline from sequence data from liver (H) and muscle (I) generated by Dataset A, Dataset B and Dataset E.

Figure 3

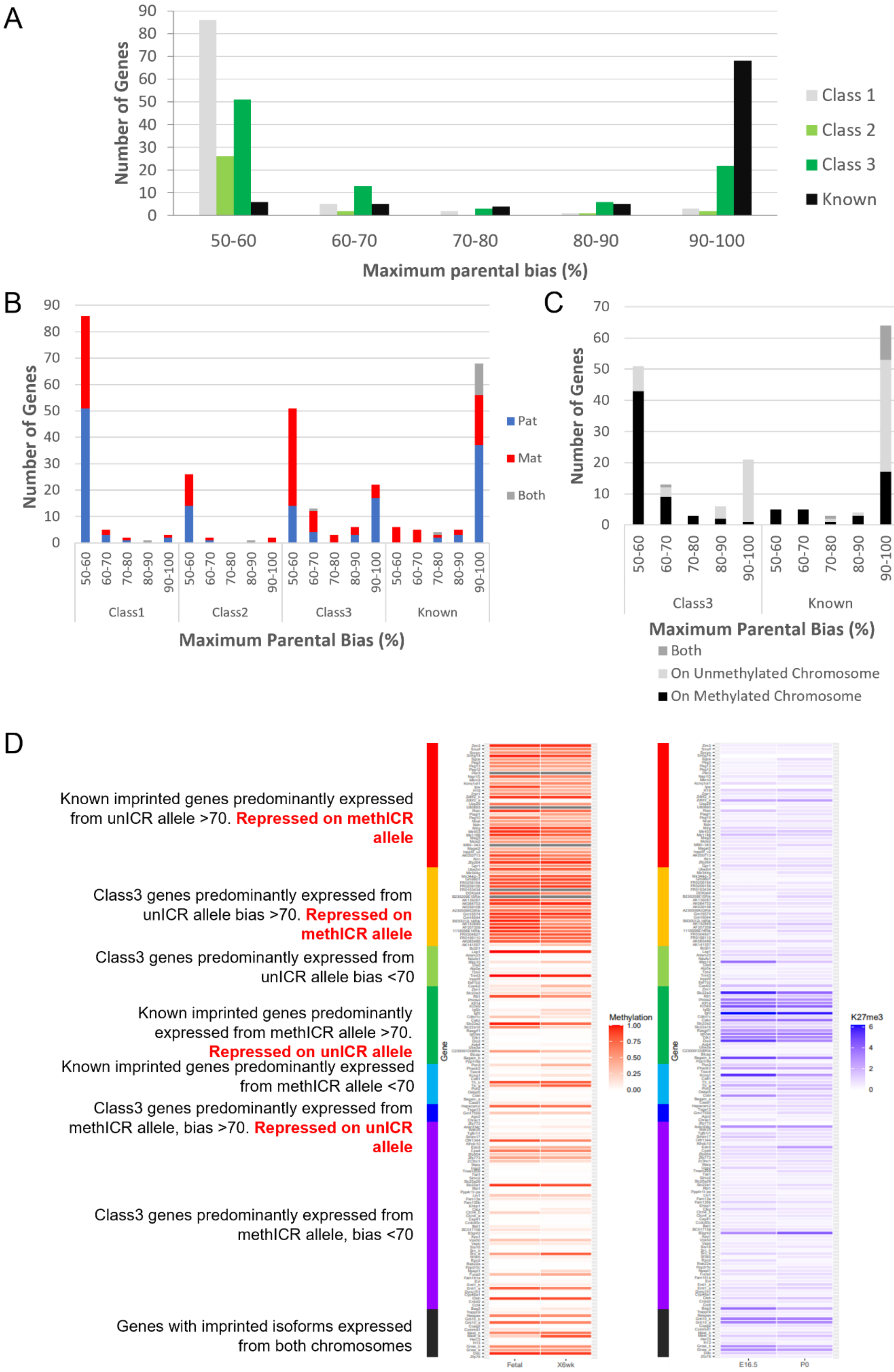

A
Number of Genes
90
80
70
60
50
40
30
20
10
0
50-60
60-70
70-80
80-90
90-100
Maximum parental bias (%)
Class 1
Class 2
Class 3
Known
B
Number of Genes
Pat
Mat
Both
50-60
60-70
70-80
80-90
90-100
Class1
Class2
Class3
Known
Maximum Parental Bias (%)
C
Number of Genes
Class3
Known
Maximum Parental Bias (%)
Both
On Unmethylated Chromosome
On Methylated Chromosome
D
Known imprinted genes predominantly expressed from unICR allele >70. Repressed on methICR allele
Class3 genes predominantly expressed from unICR allele bias >70. Repressed on methICR allele
Class3 genes predominantly expressed from unICR allele bias <70
Known imprinted genes predominantly expressed from methICR allele >70. Repressed on unICR allele
Known imprinted genes predominantly expressed from methICR allele <70
Class3 genes predominantly expressed from methICR allele, bias >70. Repressed on unICR allele
Class3 genes predominantly expressed from methICR allele, bias <70
Genes with imprinted isoforms expressed from both chromosomes
Gene
Methylation
1.00
0.75
0.50
0.25
0.00
Fetal
X6wk
Gene
K27me3
6
4
2
0
E16.5
P0

Figure 3

A- Distribution of biased genes by maximum reported bias. Class-1 = novel singletons >1 Mb from another gene identified in any of the studies. Class-2 = novel clusters where two or more novel genes within 1 Mb of each other but over 1Mb from a known imprinted gene. Class-3 = novel genes are within 1 Mb of a known imprinted gene. B – Distribution of biased genes by preferential parental chromosome and maximum reported bias. Genes preferentially expressed from the maternal chromosome are shown in red, paternal chromosome in blue and preferentially expressed from both chromosomes in a tissue specific manner in grey. C - Distribution of known and Class-3 genes by methylation status of ICR on preferentially expressed allele. Genes preferentially expressed from the Meth-ICR chromosome are shown in black, Un-ICR chromosome in pale-grey and preferentially expressed from both chromosomes in a tissue specific manner in dark-grey. D – Heap maps of DNA methylation (fetal and 6-week male frontal cortex[32,50]) and Histone H3 lysine 27 trimethylation (E16 and P0 forebrain[31,32]) over the promoters of known and Class-3 novel genes. Promoters are defined as 500bp upstream of the transcription factor binding site. Genes are sorted by maximum reported bias and methylation status of the ICR on the preferentially expressed allele.

Figure 4

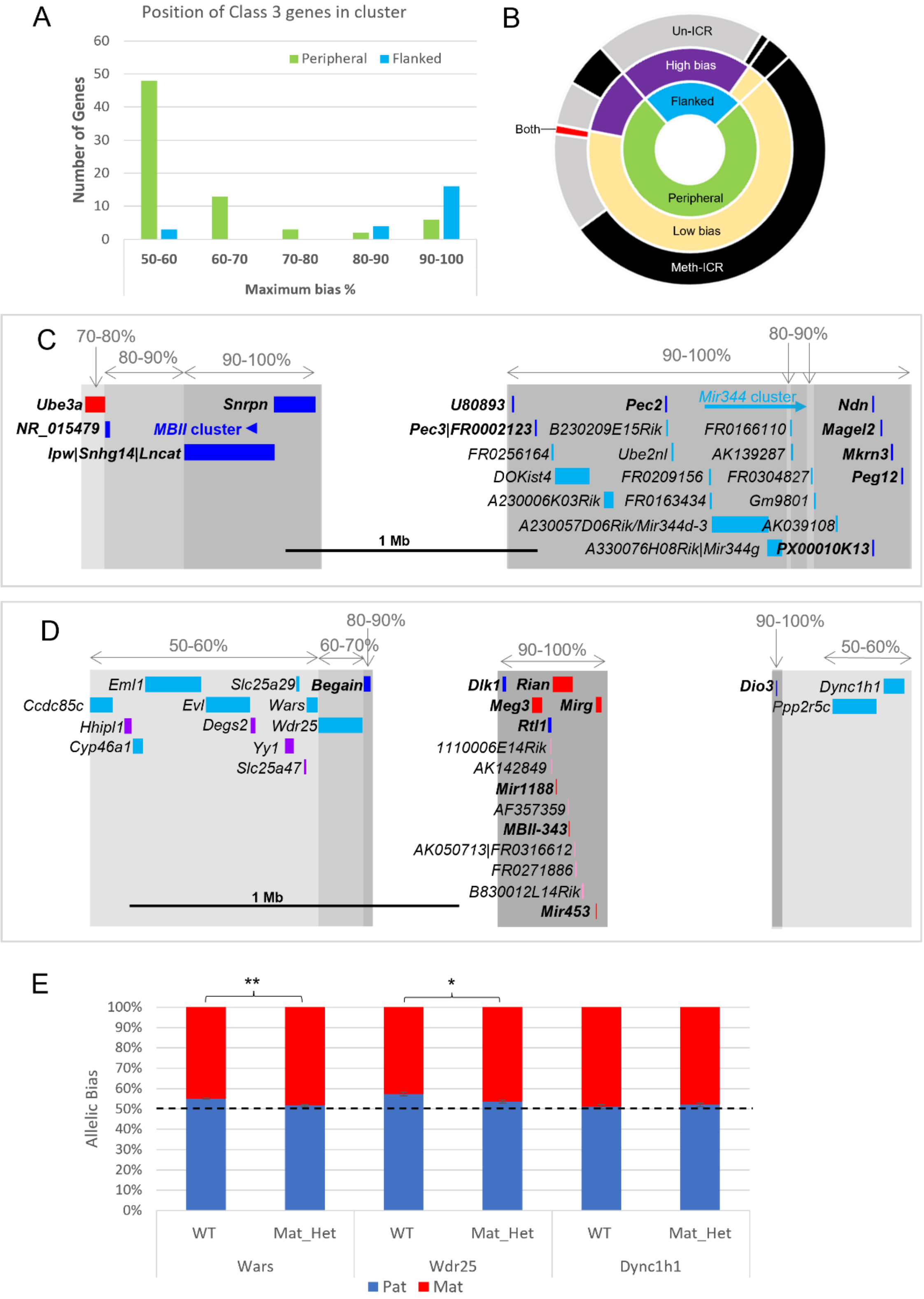

A
Position of Class 3 genes in cluster
🟩 Peripheral 🟦 Flanked
Number of Genes
0
10
20
30
40
50
60
50-60
60-70
70-80
80-90
90-100
Maximum bias %
B
Un-ICR
High bias
Flanked
Both
Peripheral
Low bias
Meth-ICR
C
70-80%
80-90%
90-100%
Ube3a
NR_015479
Ipw|Snhg14|Lncat
Snrpn
MBII cluster
1 Mb
80-90%
90-100%
U80893
Pec3|FR0002123
FR0256164
DOKist4
A230006K03Rik
A230057D06Rik/Mir344d-3
A330076H08Rik|Mir344g
Pec2
B230209E15Rik
Ube2nl
FR0209156
FR0163434
Mir344 cluster
FR0166110
AK139287
FR0304827
Gm9801
AK039108
PX00010K13
Ndn
Magel2
Mkrn3
Peg12
D
50-60%
80-90%
60-70%
Eml1
Ccdc85c
Hhipl1
Cyp46a1
Evl
Degs2
Yy1
Slc25a47
Slc25a29
Wars
Wdr25
Begain
1 Mb
90-100%
Dlk1
Rian
Meg3
Mirg
Rtl1
1110006E14Rik
AK142849
Mir1188
AF357359
MBII-343
AK050713|FR0316612
FR0271886
B830012L14Rik
Mir453
90-100%
50-60%
Dio3
Dync1h1
Ppp2r5c
E
**
*
Allelic Bias
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
WT
Mat_Het
WT
Mat_Het
WT
Mat_Het
Wars
Wdr25
Dync1h1
🟦 Pat 🟥 Mat

Figure 4

A – Distribution of Class-3 biased genes in relation to known imprinted genes. Genes flanked by known imprinted genes are shown in turquoise and those peripheral to known imprinted genes are shown in green. B – Sunburst graph of relationship between position in cluster, extent of bias and ICR methylation in Class-3 genes. Low bias = <70% expression from preferential chromosome, high bias = >70% expression from preferential chromosome. Preferential expression from the methylated ICR chromosome is shown in black, preferential expression from the unmethylated ICR chromosome is shown in grey, genes reported as being biased on both chromosomes depending on tissue or study are shown in red.  C and D – Schematics of the *Snrpn* (C) and *Dlk1* (D) regions. Highly biased novel genes (80-100%) are located between known imprinted transcripts whereas low biased genes (50-70%) are located at the periphery. Red boxes = known maternally expressed genes. Blue boxes = known paternally expressed genes. Pink boxes = novel maternally biased genes called by original studies. Turquoise boxes = novel paternally biased genes called by original studies. Blue arrow = cluster of imprinted MBII snoRNAs. Turquoise arrow Mir344 cluster. E – Allele specific expression analysis in of peripheral genes in the *Dlk1* region in IG-DMR knockout mice. Female mice, heterozygous for IG-DMR knockout were crossed with male CastEiJ mice and expression was assessed by pyrosequencing. Wildtype (n=5) and maternal heterozygote (n=6) expression biases were compared using an unpaired T-test.

Figure 5

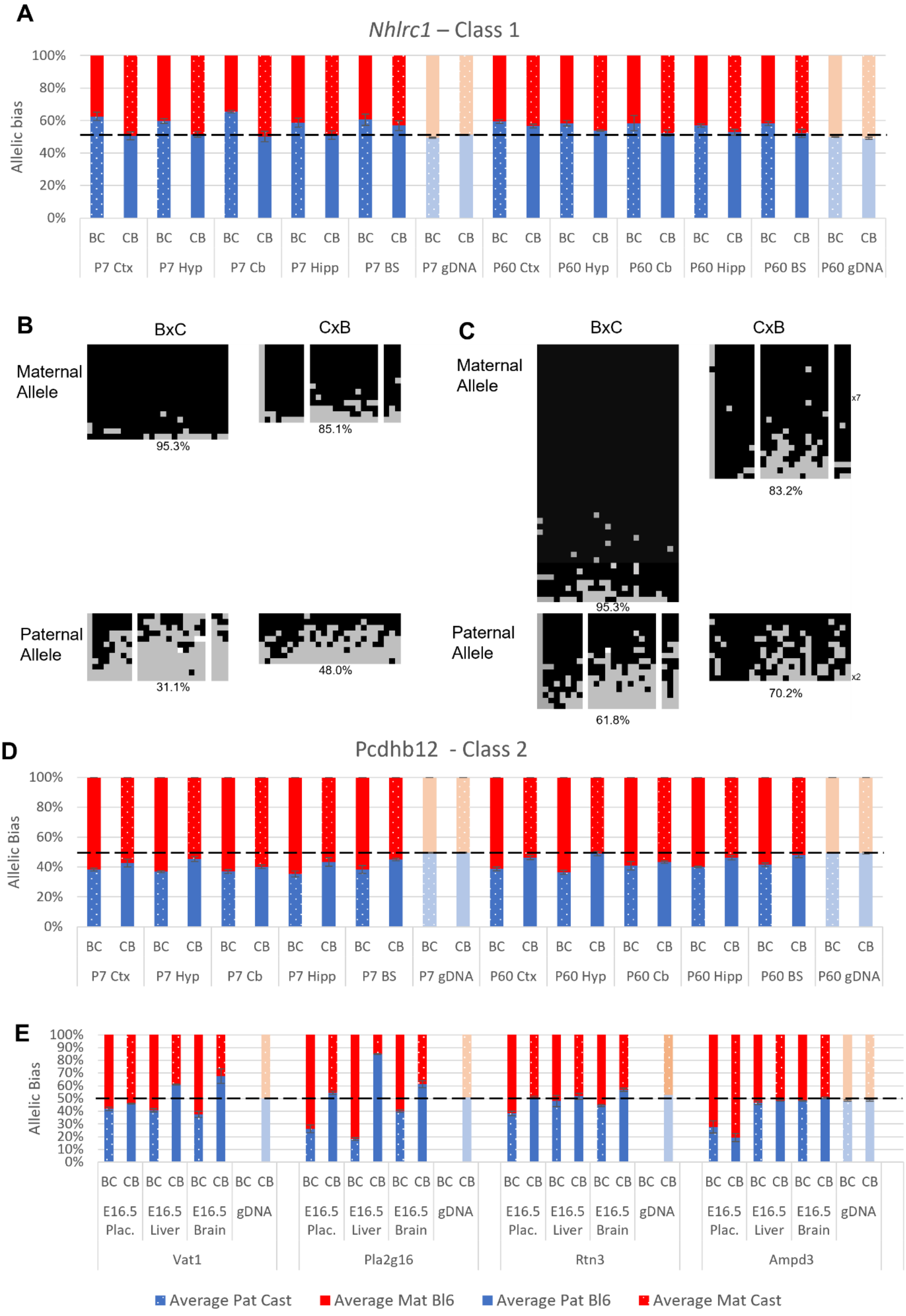

A
*Nhlrc1* – Class 1
Allelic bias
100%
80%
60%
40%
20%
0%
BC CB
P7 Ctx
P7 Hyp
P7 Cb
P7 Hipp
P7 BS
P7 gDNA
P60 Ctx
P60 Hyp
P60 Cb
P60 Hipp
P60 BS
P60 gDNA
B
BxC
CxB
Maternal Allele
95.3%
85.1%
Paternal Allele
31.1%
48.0%
C
BxC
CxB
Maternal Allele
x7
95.3%
83.2%
Paternal Allele
x2
61.8%
70.2%
D
Pcdhb12 - Class 2
Allelic Bias
100%
80%
60%
40%
20%
0%
BC CB
P7 Ctx
P7 Hyp
P7 Cb
P7 Hipp
P7 BS
P7 gDNA
P60 Ctx
P60 Hyp
P60 Cb
P60 Hipp
P60 BS
P60 gDNA
E
Allelic Bias
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
BC CB
E16.5 Plac.
E16.5 Liver
E16.5 Brain
gDNA
Vat1
Pla2g16
Rtn3
Ampd3
Average Pat Cast
Average Mat Bl6
Average Pat Bl6
Average Mat Cast

Figure 5

**Experimentally validated Class-1 and Class-2 genes.** A – *Nhlrc1* (Class-1) is paternally biased in all postnatal neuronal tissues tested. B-C Bisulfite sequencing analysis in P7 tissues. B – Cerebellum, C-Liver. Each line represents a different clone of bisulfite sequencing derived from 2 BxC animals and 2 CxB animals. Numbers of identical clones sequenced are indicated to the right. Black = methylated CpG and Grey = unmethylated CpG, white = CpG absent from clone. Percentage of methylated CpGs from all clones is indicated underneath. D – *Pcdhb12* (Class-2) is maternally biased in all postnatal tissues tested. E – Three Class-2 genes show a maternal bias in e16.5 placenta: *Vat1*, *Pla2g16* and *Rtn3*. These biases are weaker than seen in *Ampd3* which is imprinted in the placenta [34]. Allele specific expression graphs (A, D and E) show mean expression (%) from the paternal allele (deep blue) and maternal allele (red) in C57BL/6 x CastEiJ (BC) and 4 CastEiJ x C57BL/6 (CB) crosses. Castaneus allele is denoted by spotted pattern. Standard error of the mean is shown N = 3 or 4. Amplification bias was assessed in genomic DNA paternal allele (pale blue) and maternal allele (salmon). *Pcdhb12*, *Vat1* and *Ampd2* assays had an amplification bias and the data in these graphs are corrected accordingly.

Figure 6

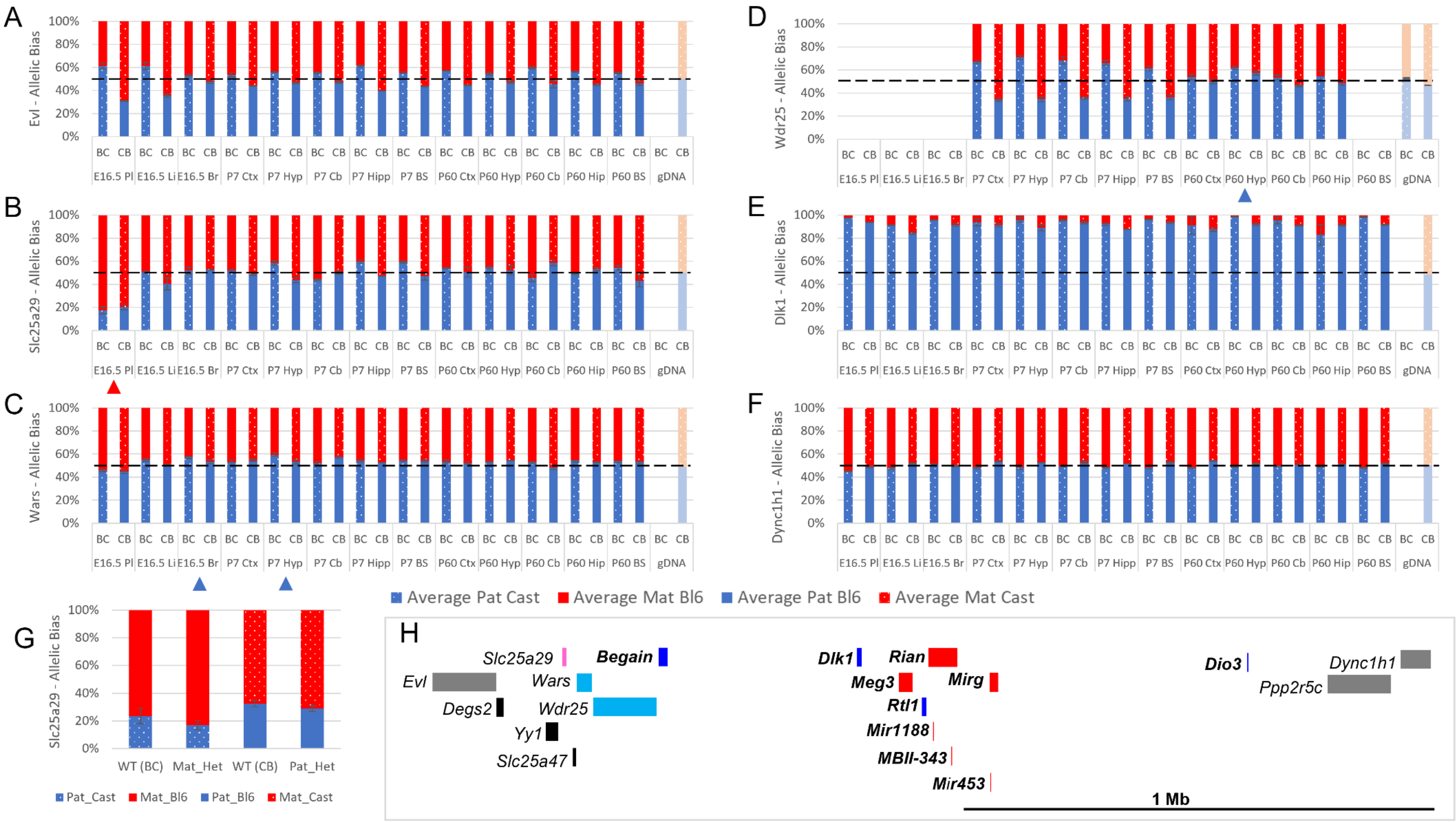
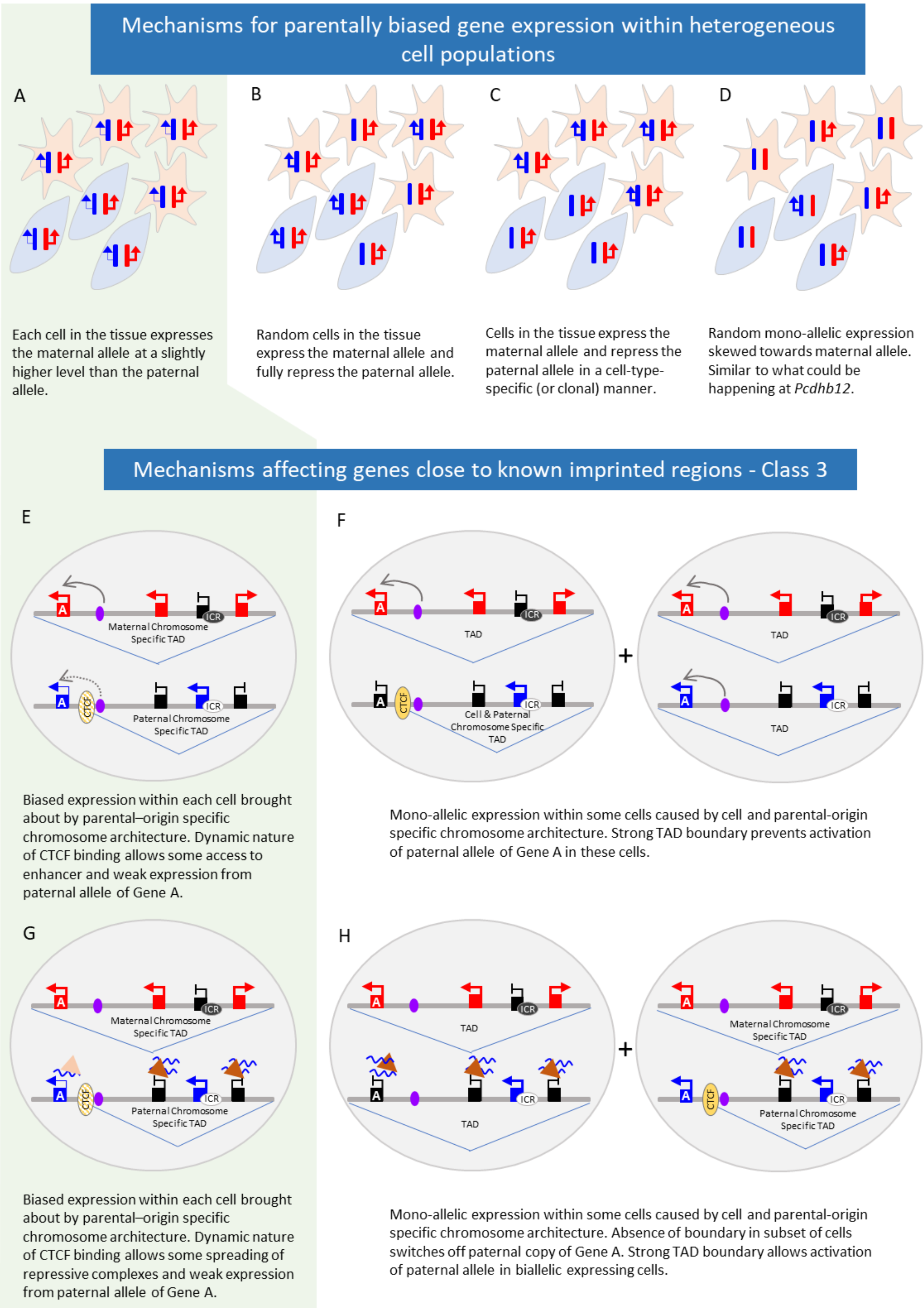

A
Evl - Allelic Bias
B
Slc25a29 - Allelic Bias
C
Wars - Allelic Bias
D
Wdr25 - Allelic Bias
E
Dlk1 - Allelic Bias
F
Dync1h1 - Allelic Bias
100%
80%
60%
40%
20%
0%
BC CB BC CB BC CB BC CB BC CB BC CB BC CB BC CB BC CB BC CB BC CB BC CB BC CB BC CB
E16.5 Pl
E16.5 Li
E16.5 Br
P7 Ctx
P7 Hyp
P7 Cb
P7 Hipp
P7 BS
P60 Ctx
P60 Hyp
P60 Cb
P60 Hip
P60 BS
gDNA
■ Average Pat Cast ■ Average Mat Bl6 ■ Average Pat Bl6 ■ Average Mat Cast
G
Slc25a29 - Allelic Bias
WT (BC)
Mat_Het
WT (CB)
Pat_Het
■ Pat_Cast ■ Mat_Bl6 ■ Pat_Bl6 ■ Mat_Cast
H
Slc25a29
Begain
Evl
Wars
Degs2
Wdr25
Yy1
Slc25a47
Dlk1
Rian
Meg3
Mirg
Rtl1
Mir1188
MBII-343
Mir453
Dio3
Dync1h1
Ppp2r5c
1 Mb

Figure 6

**Allele specific expression analysis of the *Dlk1* domain** - *Evl* (A), *Slc25a29* (B), *Wars* (C) *Wdr24* (D) *Dlk1* (E) and *Dync1h1* (F) Graphs show mean expression (%) from the paternal allele (deep blue) and maternal allele (red) C57BL/6 x CastEiJ (BC) and 4 CastEiJ x C57BL/6 (CB) crosses. Castaneus allele is denoted by spotted pattern. Standard error of the mean is shown, N = 3 or 4. Tissues with a bias greater than 45:55 are indicated by arrow heads. Amplification bias was assessed in genomic DNA paternal allele (pale blue) and maternal allele (salmon). The *Evl* assay had an amplification bias and the data are corrected. (G) – Imprinting of Slc25a29 in e15.5 placenta is not under the control of the IG-DMR. WT (BC) = maternal allele is wildtype for the IG-DMR paternal allele is CastEiJ (N = 5). Mat_Het = maternal allele has IG-DMR deletion and paternal allele is CastEiJ (N = 6). WT (CB) = paternal allele is wildtype for the IG-DMR maternal allele is CastEiJ (N = 5). Pat_Het = Paternal allele has IG-DMR deletion and maternal allele is CastEiJ (N = 7) . (H) Schematic of the validated expression data in the Dlk1 region. Red boxes = known maternally expressed genes. Blue boxes = known paternally expressed genes. Pink boxes = novel validated maternally biased genes. Turquoise boxes = novel validated paternally biased genes. Grey boxes = biallelically expressed genes.

Figure 7



Mechanisms for parentally biased gene expression within heterogeneous cell populations

A: Each cell in the tissue expresses the maternal allele at a slightly higher level than the paternal allele.

B: Random cells in the tissue express the maternal allele and fully repress the paternal allele.

C: Cells in the tissue express the maternal allele and repress the paternal allele in a cell-type-specific (or clonal) manner.

D: Random mono-allelic expression skewed towards maternal allele. Similar to what could be happening at *Pcdhb12*.

Mechanisms affecting genes close to known imprinted regions - Class 3

E
A
ICR
Maternal Chromosome Specific TAD
CTCF
Paternal Chromosome Specific TAD
F
TAD
Cell & Paternal Chromosome Specific TAD
+
G
H

E: Biased expression within each cell brought about by parental–origin specific chromosome architecture. Dynamic nature of CTCF binding allows some access to enhancer and weak expression from paternal allele of Gene A.

F: Mono-allelic expression within some cells caused by cell and parental-origin specific chromosome architecture. Strong TAD boundary prevents activation of paternal allele of Gene A in these cells.

G: Biased expression within each cell brought about by parental–origin specific chromosome architecture. Dynamic nature of CTCF binding allows some spreading of repressive complexes and weak expression from paternal allele of Gene A.

H: Mono-allelic expression within some cells caused by cell and parental-origin specific chromosome architecture. Absence of boundary in subset of cells switches off paternal copy of Gene A. Strong TAD boundary allows activation of paternal allele in biallelic expressing cells.

Figure 7 – Mechanisms behind parent-of-origin expression biases in tissues. A-D - Scenarios causing biased expression in heterogenous cell populations: bias occurs in every cell in the tissue (A), random imprinting in subset of cells (B), cell-type specific imprinting (C) or random monoallelic expression that is skewed towards one allele (D). E-H – Possible mechanisms behind parent-of-origin biases at the periphery of imprinted domains.

Table 1

| *Genes* | Chr. | Dataset | Direction previously reported | Class | e16.5 Plac. | e16.5 Liver | e16.5 Brain | P7 Cortex | P7 Hyp. | P7 Cb. | P7 Hipp. | P7 B.S | P60 Cortex | P60 Hyp. | P60 Cb. | P60 Hipp. | P60 B.S | Strain Bias | Validation status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Class 1 (Novel Singletons)* | | | | | | | | | | | | | | | | | | | |
| *L3mbtl1* | 2 | B | pat | 1 | - | - | - | 48.0 | 53.5 | 53.4 | 46.3 | 50.1 | 49.4 | 49.7 | 46.7 | 51.0 | 49.2 | | Biallelic |
| *Ahi1* | 10 | B,D | pat | 1 | 52.7 | 46.5 | n/a | 47.1 | 51.6 | 50.9 | 52.8 | n/a | 48.6 | 49.4 | 50.0 | 52.4 | 51.3 | ✓ | Biallelic |
| *Platr20* | 11 | A | pat | 1 | 51.7 | 49.9 | 50.0 | 49.9 | 50.4 | 50.1 | 50.4 | 50.2 | 49.7 | 49.8 | 50.1 | 50.2 | 49.7 | | Biallelic |
| *Calm1* | 12 | B,D | pat | 1 | - | - | - | 54.5 | 49.6 | 51.4 | 46.0 | 43.4 | n/a | n/a | n/a | n/a | n/a | ✓ | Biallelic |
| *Nhlrc1* | 13 | B, C, D | pat | 1 | - | - | - | **56.6** | **55.6** | **57.8** | 54.8 | **58.8** | **58.1** | **56.1** | **55.4** | **55.0** | **55.2** | ✓ | **Paternal** |
| *Tnk1* | 11 | A | mat | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Mlana* | 18 | B | mat | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Gm16299* | 19 | C | pat | 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Class 2 (Novel Clusters)* | | | | | | | | | | | | | | | | | | | |
| *Stx6* | 1 | C | mat | 2 | - | - | - | 50.7 | 49.2 | 50.3 | 49.8 | 50.1 | n/a | n/a | n/a | n/a | n/a | ✓ | Biallelic |
| *Gabra5* | 7 | B,D | pat | 2 | - | - | - | 51.5 | 53.1 | 52.3 | 54.8 | 52.6 | 52.3 | 50.7 | 48.6 | 51.5 | 51.5 | | Biallelic |
| *Wnk4* | 11 | C | mat | 2 | - | - | - | 50.1 | 52.0 | 45.3 | 50.6 | 44.0 | 48.1 | 44.6 | 50.9 | 45.4 | 48.1 | ✓ | Maternal |
| *Vat1* | 11 | B | mat | 2 | 43.9 | 50.9 | 52.6 | 49.6 | 48.8 | 50.4 | 51.0 | 49.9 | 46.4 | 51.3 | 50.8 | 49.8 | 50.4 | ✓ | Placental |
| *Rdm1* | 11 | A | mat | 2 | 49.4 | - | - | 48.8 | 48.8 | 50.1 | 50.2 | 50.7 | 46.9 | 49.5 | 50.5 | 50.1 | 48.7 | ✓ | Biallelic |
| *Gaa* | 11 | B,D | pat | 2 | 48.9 | 52.4 | 53.6 | 51.3 | 47.1 | 51.7 | 53.1 | 51.9 | 51.8 | 51.6 | 47.9 | 51.9 | 50.8 | ✓ | Biallelic |
| *Pcdhb10* | 18 | D | mat | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Pcdhb12* | 18 | B,C | mat | 2 | - | - | - | **40.5** | **41.2** | **41.0** | **39.4** | **41.8** | **42.6** | **42.7** | **42.1** | **43.2** | **44.9** | ✓ | **Maternal** |
| *Pcdhb20* | 18 | B,C | pat | 2 | - | - | 52.7 | 51.4 | 51.6 | 52.0 | 52.6 | 53.1 | 51.5 | 52.3 | 51.6 | 51.6 | 50.1 | ✓ | Biallelic |
| *Prdx5* | 19 | B | pat | 2 | 46.9 | 51.1 | 49.4 | 49.0 | 50.4 | 48.8 | 49.4 | 49.9 | 49.2 | 50.2 | 50.0 | 48.3 | 49.2 | ✓ | Biallelic |
| *Rtn3* | 19 | D | pat | 2 | 44.6 | 49.9 | 50.7 | 49.4 | 47.4 | 49.5 | 49.7 | 49.3 | 49.1 | 50.9 | 49.7 | 50.4 | 50.4 | ✓ | Placental |
| *Pla2g16* | 19 | B | mat | 2 | 40.4 | 51.6 | 50.6 | 51.7 | 49.7 | 48.9 | 48.5 | 50.7 | 51.5 | 51.2 | 49.2 | 46.9 | 48.3 | ✓ | Placental |
| *Mr1* | 1 | C | mat | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *BC034090* | 1 | C | mat | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Tmem106a* | 11 | A | mat | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Class 3 (Close to known imprinted genes)* | | | | | | | | | | | | | | | | | | | |
| *Adam23* | 1 | A,B,C,D | pat | 3 | 48.0 | 52.0 | **59.1** | **56.5** | **57.8** | **56.5** | **58.0** | 53.7 | **55.6** | **58.7** | 53.5 | **56.5** | 54.1 | | **Paternal** |
| *Mcts2* | 2 | A,B,C | pat | K | **77.2** | **82.8** | **64.6** | **71.8** | **80.2** | **73.0** | **77.2** | **80.4** | **85.7** | **78.3** | **70.6** | **87.7** | **84.1** | | **Paternal** |
| *Cox4i2* | 2 | C | pat | 3 | 49.3 | - | - | **69.8** | 51.0 | **56.6** | **60.0** | **56.5** | **55.3** | **55.0** | 52.8 | **56.8** | 54.9 | | **Paternal** |
| *Bcl2l1* | 2 | A,B,C,D | pat | 3 | 49.5 | 50.2 | **61.6** | **60.9** | **61.8** | **58.3** | **59.5** | **58.8** | **60.0** | **61.4** | **57.4** | **59.2** | **59.0** | | **Paternal** |
| *Tpx2* | 2 | C | pat | 3 | - | 49.7 | 49.1 | 53.4 | 52.6 | 50.7 | 53.6 | 51.4 | **56.4** | **55.4** | **64.5** | **62.6** | **61.0** | ✓ | **Paternal** |
| *Herc3* | 6 | A,B,C,D | mat | K | 47.0 | 45.8 | **43.1** | 45.7 | **40.7** | **40.5** | 49.9 | **32.4** | **44.0** | **29.5** | **39.1** | **42.3** | **24.5** | ✓ | **Maternal** |
| *Fam13a* | 6 | B,D | mat | 3 | - | 46.1 | 53.2 | 47.9 | 47.1 | 51.3 | 49.4 | 48.4 | n/a | n/a | n/a | n/a | n/a | ✓ | Biallelic |
| *Zfp78* | 7 | B | both | 3 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Smim17* | 7 | B,D | mat | 3 | **-** | **-** | **38.3** | **48.8** | **35.3** | **50.8** | **44.0** | **44.6** | **45.2** | **43.3** | **52.1** | **50.8** | **49.3** | | **Maternal** |
| *Peg3* | 7 | A,B,C,D | pat | K | **96.2** | **99.2** | **99.6** | **92.8** | **93.2** | **94.5** | **94.6** | **95.3** | **97.5** | **98.0** | **98.0** | **97.3** | **97.7** | | **Paternal** |
| *Zfp954* | 7 | B | mat | 3 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Zfp773* | 7 | B | mat | 3 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Zfp772* | 7 | B | mat | 3 | - | - | - | - | - | - | - | - | - | - | - | - | - | | low expression |
| *Clcn4-2* | 7 | B | mat | 3 | 44.6 | 54.5 | 48.5 | 48.9 | 48.8 | 49.9 | 50.6 | 48.0 | 49.5 | 50.7 | 51.0 | 49.6 | 50.7 | ✓ | Placental |
| *Ifitm10* | 7 | C,D | mat | 3 | 46.9 | 52.3 | 48.7 | 47.5 | **41.5** | 53.0 | 49.4 | **45.0** | 49.0 | **43.2** | **43.6** | 50.7 | 47.0 | ✓ | **Maternal** |
| *Ctsd* | 7 | B,D | mat | 3 | 46.8 | 47.8 | 50.1 | 50.0 | 49.0 | 48.4 | 48.9 | 49.8 | 49.3 | 50.2 | 49.0 | 50.5 | 49.5 | | Biallelic |
| *Evl* | 12 | B | pat | 3 | 46.2 | 48.3 | 50.2 | 48.7 | 51.3 | 51.1 | 50.6 | 49.5 | 50.9 | 50.9 | 52.6 | 50.8 | 50.6 | ✓ | Biallelic |
| *Slc25a29* | 12 | B | pat | 3 | **18.9** | 45.5 | 52.3 | 50.4 | 51.1 | 47.4 | 53.5 | 53.2 | 52.3 | 53.3 | 52.1 | 51.3 | 48.7 | ✓ | **Placental** |
| *Wars* | 12 | C | pat | 3 | 45.2 | 52.5 | **55.1** | 53.9 | **56.2** | 54.5 | 53.3 | 54.2 | 52.8 | 53.9 | 50.6 | 53.6 | 53.2 | | **Paternal** |
| *Wdr25* | 12 | B,D | pat | 3 | - | - | - | 50.6 | 52.8 | 51.9 | 50.6 | 48.8 | 51.8 | **59.7** | 49.9 | 51.6 | 51.3 | ✓ | **Paternal** |
| *Dlk1* | 12 | A,B,C,D | pat | K | **95.6** | **87.8** | **93.6** | **92.4** | **92.5** | **94.5** | **90.1** | **94.9** | **89.4** | **95.4** | **93.1** | **87.1** | **95.0** | | **Paternal** |
| *Ppp2r5c* | 12 | B,C | pat | 3 | 52.4 | 51.3 | 48.5 | 48.6 | 49.3 | 54.0 | 50.6 | 48.2 | 50.6 | 48.7 | 45.2 | 51.0 | 49.3 | ✓ | Biallelic |
| *Dync1h1* | 12 | B,C | pat | 3 | 46.9 | 49.8 | 50.4 | 50.7 | 50.7 | 51.1 | 49.9 | 50.4 | 51.0 | 50.4 | 50.1 | 50.1 | 49.8 | | Biallelic |
| *Ago2* | 15 | A,B,C,D | mat | 3 | 47.7 | 50.7 | **30.2** | **24.4** | **26.2** | **28.6** | **28.7** | **19.3** | **25.9** | **28.7** | **38.1** | **33.4** | **24.6** | | **Maternal** |
| *Ampd3* | 7 | B | mat | K | **23.4** | 47.5 | - | 48.6 | 48.5 | 48.3 | 51.8 | 50.9 | 50.1 | 46.7 | 50.9 | 49.4 | 51.5 | ✓ | **Placental** |
| *Gab1* | 8 | A | pat | K | **74.9** | 50.4 | 49.8 | 52.7 | 49.6 | 48.6 | 49.4 | 51.0 | 47.1 | 49.6 | 48.8 | 50.1 | 50.2 | | **Placental** |

Table 1 – Table showing summary of all the allele-specific pyrosequencing performed to validate putative biased genes. Values show the mean expression (%) from the paternal allele of both reciprocal crosses to eliminate strain bias. Values above 55% are called as paternally biased (blue) and values below 45% are called as maternally biased (red). Assays with a strain bias of greater than 45:55 in more than 1 tissue are indicated in the 19th column. Genes that only validate in the placenta are called as Placental in the 20th column (Red = maternal, Blue = Paternal).

## References

1. Tucci, V. *et al.* Genomic Imprinting and Physiological Processes in Mammals. *Cell* **176**, 952–965 (2019).

2. Plasschaert, R. N. & Bartolomei, M. S. Genomic imprinting in development, growth, behavior and stem cells. *Dev.* **141**, 1805–1813 (2014).

3. Cleaton, M. A. M., Edwards, C. A. & Ferguson-Smith, A. C. Phenotypic Outcomes of Imprinted Gene Models in Mice: Elucidation of Pre- and Postnatal Functions of Imprinted Genes. *Annu. Rev. Genomics Hum. Genet.* **15**, 93–126 (2014).

4. Ishida, M. & Moore, G. E. The role of imprinted genes in humans. *Mol Asp. Med* (2012). doi:S0098-2997(12)00081-7 [pii]10.1016/j.mam.2012.06.009

5. Uribe-Lewis, S., Woodfine, K., Stojic, L. & Murrell, A. Molecular mechanisms of genomic imprinting and clinical implications for cancer. *Expert Rev. Mol. Med.* **13**, 1–22 (2011).

6. Li, X. *et al.* A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev Cell* **15**, 547–557 (2008).

7. Takahashi, N. *et al.* ZNF445 is a primary regulator of genomic imprinting. *Genes Dev.* (2019). doi:10.1101/gad.320069.118

8. Edwards, C. A. & Ferguson-Smith, A. C. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* **19**, 281–289 (2007).

9. Inoue, A., Jiang, L., Lu, F., Suzuki, T. & Zhang, Y. Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547**, 419–424 (2017).

10. Babak, T. *et al.* Global Survey of Genomic Imprinting by Transcriptome Sequencing. *Curr. Biol.* (2008). doi:10.1016/j.cub.2008.09.044

11. Wang, X. *et al.* Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One* (2008). doi:10.1371/journal.pone.0003839

12. Gregg, C. *et al.* High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science (80-. ).* (2010). doi:10.1126/science.1190830

13. Gregg, C., Zhang, J., Butler, J. E., Haig, D. & Dulac, C. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science (80-. ).* (2010). doi:10.1126/science.1190831

14. Wang, X., Soloway, P. D. & Clark, A. G. A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics* (2011). doi:10.1534/genetics.111.130088

15. DeVeale, B., van der Kooy, D. & Babak, T. Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. *PLoS Genet.* (2012). doi:10.1371/journal.pgen.1002600

16. Babak, T. *et al.* Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat. Genet.* **47**, 544–549 (2015).

17. Bonthuis, P. J. *et al.* Noncanonical genomic imprinting effects in offspring. *Cell Rep.* **12**, 979–991 (2015).

18. Crowley, J. J. *et al.* Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat. Genet.* **47**, 353–360 (2015).

19. Perez, J. D. *et al.* Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife* **4**, 41 (2015).

20. Andergassen, D. *et al.* Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *Elife* **6**, (2017).

21. Zwart, R., Sleutels, F., Wutz, A., Schinkel, A. H. & Barlow, D. P. Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. *Genes Dev* **15**, 2361–2366 (2001).

22. Okae, H. *et al.* Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Hum. Mol. Genet.* (2012). doi:10.1093/hmg/ddr488

23. Paulsen, M. *et al.* Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum Mol Genet* **9**, 1829–1841 (2000).

24. Kuzmin, A. *et al.* The PcG gene Sfmbt2 is paternally expressed in extraembryonic tissues. *Gene Expr. Patterns* (2008). doi:10.1016/j.modgep.2007.09.005

25. Wang, Q. *et al.* Recent acquisition of imprinting at the rodent Sfmbt2 locus correlates with insertion of a large block of miRNAs. *BMC Genomics* (2011). doi:10.1186/1471-2164-12-204

26. Reynès, C. *et al.* ISoLDE: A data-driven statistical method for the inference of allelic imbalance in datasets with reciprocal crosses. *Bioinformatics* **36**, 504–513 (2020).

27. Xu, H. *et al.* Landscape of genomic imprinting and its functions in the mouse mammary gland. doi:10.1093/jmcb/mjaa020

28. Fitzpatrick, G. V, Soloway, P. D. & Higgins, M. J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet* **32**, 426–431 (2002).

29. Pandey, R. R. *et al.* Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Mol. Cell* (2008). doi:10.1016/j.molcel.2008.08.022

30. Lister, R. *et al.* Global Epigenomic Reconfi guration During Mammalian Brain Development. doi:10.1126/science.1237905

31. Gorkin, D. U. *et al.* Systematic mapping of chromatin state landscapes during mouse development. *bioRxiv* 166652 (2017). doi:10.1101/166652

32. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* (2012). doi:10.1038/nature11243

33. Lin, S. P. *et al.* Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet* **35**, 97–102 (2003).

34. Schulz, R. *et al.* Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. *Nucleic Acids Res.* **34**, (2006).

35. Proudhon, C. *et al.* Protection against De Novo Methylation Is Instrumental in Maintaining Parent-of-Origin Methylation Inherited from the Gametes. *Mol. Cell* (2012). doi:10.1016/j.molcel.2012.07.010

36. Strogantsev, R. *et al.* Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol.* **16**, (2015).

37. Tsai, C. E. *et al.* Genomic imprinting contributes to thyroid hormone metabolism in the mouse embryo. *Curr Biol* **12**, 1221–1226 (2002).

38. Hanna, C. W. *et al.* Endogenous retroviral insertions drive non-canonical imprinting in extra-embryonic tissues. *Genome Biol.* **20**, 225 (2019).

39. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish Oocyte-specific heritable marks. *PLoS Genet.* **8**, (2012).

40. Hirayama, T. & Yagi, T. Regulation of clustered protocadherin genes in individual neurons. *Seminars in Cell and Developmental Biology* **69**, 122–130 (2017).

41. Murrell, A., Heeson, S. & Reik, W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet* **36**, 889–893 (2004).

42. Tan, L. *et al.* Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* (2021). doi:10.1016/j.cell.2020.12.032

43. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

44. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* (2014). doi:10.1016/j.cell.2014.11.021

45. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* (2017). doi:10.7554/elife.25776

46. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* (2013). doi:10.1038/nature12593

47. Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* (2020). doi:10.1038/s41588-020-00716-8

48. Terranova, R. *et al.* Polycomb Group Proteins Ezh2 and Rnf2 Direct Genomic Contraction and Imprinted Repression in Early Mouse Embryos. *Dev. Cell* (2008). doi:10.1016/j.devcel.2008.08.015

49. Song, Q. *et al.* A Reference Methylome Database and Analysis Pipeline to Facilitate Integrative and Comparative Epigenomics. *PLoS One* **8**, e81148 (2013).

50. Sloan, C. A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, (2016).

51. Karolchik, D. *et al.* The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* **32**, D493-6 (2004).

52. Wickham, H. *Ggplot2: Elegant graphics for data analysis*. (Springer, Cham, 2016).

53. Krueger, F. & Andrews, S. R. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes [version 2; referees: 3 approved]. *F1000Research* **5**, 1479 (2016).

54. Ewels, P., Krueger, F., Käller, M. & Andrews, S. Cluster Flow: A user-friendly bioinformatics workflow tool. *F1000Research* **5**, 2824 (2017).

55. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

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

57. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for

assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).