Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse

Tomas Babak[1,*], **Brian DeVeale**[2], **Emily K. Tsang**[3], **Yiqi Zhou**[1], **Xin Li**[3], **Kevin S. Smith**[3], **Kim R. Kukurba**[3,4], **Rui Zhang**[4], **Jin Billy Li**[4], **Derek van der Kooy**[5], **Stephen B. Montgomery**[3,4], and **Hunter B. Fraser**[1,§]

[1]Department of Biology, Stanford University, Stanford, CA, 94305, USA

[2]UCSF School of Medicine, UCSF, San Francisco, CA, 94143, USA

[3]Department of Pathology, Stanford University, Stanford, CA, 94305, USA

[4]Department of Genetics, Stanford University, Stanford, CA, 94305, USA

[5]Department of Molecular Genetics, University of Toronto, Toronto, ON, M5S 3E1, Canada

Abstract

Genomic imprinting is an epigenetic process that restricts gene expression to either the maternally or paternally inherited allele[1,2]. Many theories have been proposed to explain its evolutionary origin[3,4], but our understanding has been limited by a paucity of data mapping the breadth and dynamics of imprinting within any organism. We generated an atlas of imprinting spanning 33 mouse and 45 human developmental stages and tissues. Nearly all imprinted genes were imprinted in early development and either retained their parent-of-origin expression in adults, or lost it completely. Consistent with an evolutionary signature of parental conflict, imprinted genes were enriched for co-expressed pairs of maternally/paternally expressed genes, showed accelerated expression divergence between human and mouse, and were more highly expressed than their non-imprinted orthologs in other species. Our approach demonstrates a general framework for imprinting discovery in any species, and sheds light on the causes and consequences of genomic imprinting in mammals.

§Correspondence should be addressed to: hbfraser@stanford.edu.
*Current affiliation: Department of Biology, Queen's University, Kingston, ON, K7L 3N6, Canada

Despite over 20 years of study[2-5], evolutionary explanations for genomic imprinting remain controversial. The conflict/kinship theory posits that imprinting evolved due to different selection pressures on maternally and paternally-derived alleles[3,5,6]. For example, in species where litters of multiple paternities are common, increased expression of genes that promote fetal growth at the expense of the mother and littermates can be advantageous for paternally inherited alleles. In contrast, the inclusive fitness of maternally inherited alleles is maximized by more controlled nutrient exchange to enable equal allocation to all littermates. Other prominent theories include co-adaptation of mutually favorable traits in parent and offspring[6,7] and others[4], and it is not clear whether imprinting can be entirely explained by one model.

In order to systematically identify imprinted genes and measure the breadth of tissues and developmental stages in which they are imprinted, we constructed an atlas of genomic imprinting in mouse (Fig. 1; Supplementary File 1; Supplementary Fig. 1, Supplementary Fig. 2, Supplementary Fig. 3). We detected imprinting as allele-specific expression (ASE) consistently biased towards either the maternal or paternal allele in reciprocally crossed F1 hybrids of diverged inbred mouse strains[8-10] (C57Bl6/J and CAST/EiJ), using methods that reliably discriminate imprinting from technical and biological variation[11] (see Methods). We sequenced mRNA from 26 unique tissues/developmental stages (61 biological samples), and combined our data with seven additional published tissues[9,12-15] (**Supplementary Table 1**). Among 207 imprinting measurements that have been previously reported[16] in the gene-tissue pairs assayed here, 95.6% agree (**Supplementary Fig. 4a**); five of the nine cases that disagreed were in tissues sampled at different developmental time-points, and the remaining four include some equivocal evidence (e.g. claims of imprinting without confirmation from a reciprocal cross[17]). We confirmed the reported non-canonical maternal expression of *Igf2* and paternal expression of *Grb10* in adult brain[12], and we found this reciprocal pattern in all central nervous system (CNS) tissues (**Fig 1; Supplementary Fig. 2, Supplementary Fig. 3**). We observed both paternal and maternal expression for *Copg2* and *Rtl1* as well (**Fig. 1; Supplementary Fig. 2, Supplementary Fig. 3**). Overall, our atlas increased the number of reported gene-tissue imprinting measurements[16,18] by nearly an order of magnitude (**Supplementary Fig. 4b; Supplementary File 2**).

We identified 74 candidates for novel imprinted genes that we tested by pyrosequencing, finding evidence of imprinted expression for 12 (Supplementary Table 2; Supplementary Table 3; Supplementary Fig. 5). We designated 7/12 as "high confidence", on account of evidence from imprinting in multiple tissues and/or biological replicates, high allelic bias, and concordance at multiple SNPs. 7/12 were also located in regions previously shown to lead to parent-of-origin phenotypes (2.34 expected by chance; *p*=0.006; **Supplementary File 3**). We also found biallelic expression of eight genes previously reported as imprinted (*Htr2a, Pde4d, Tbc1d12, Gatm, Dlx5, Gabrb3, Nap1l4*, and *Pon2*), which together with other conflicting evidence[19] indicates that these are likely not imprinted (**Supplementary Table 4**). **Supplementary File 4** lists all high-confidence imprinted genes.

One of the most striking features of the mouse atlas was the robust conservation of imprinting across tissues; the majority of imprinted genes were imprinted in nearly all tissues where they were expressed (Fig. 1, Supplementary Fig. 2, Supplementary Fig. 3;

Supplementary File 1). Embryonic, extra-embryonic, and CNS tissues had the highest proportion of genes imprinted (**Fig. 1a-b**), consistent with the known role of imprinting in development and social cognition[2]. Among genes that are imprinted in some tissues and biallelic in others, 52/55 were imprinted in embryos, but not in adults (the remaining three genes [*Phf17, Gab1, Slc22a3*] are imprinted in placenta and yolk-sac). These observations support a model where imprinted expression manifests during embryogenesis and then either persists through adulthood, or is lost during development.

Genes with the most similar imprinting patterns ("co-imprinting") were often clustered in the genome (**Fig. 1c**), as expected due to shared *cis*-regulatory elements[2]. We found a number of significant functional enrichments within clusters of co-imprinted genes, including growth (e.g. decreased fetal weight), nutrient processing (e.g. glucose transport and uptake), and CNS development and signaling (e.g. nerve growth factor signaling) (see Methods and **Supplementary Table 5**). We also found a strong enrichment for neuropeptide hormone activity mediated by oxytocin/vasopressin signaling (**Supplementary Fig. 6**), consistent with a recent link to regulation of feeding behavior[20] and widespread imprinting in the hypothalamus (**Fig. 1**).

To enable comparisons of imprinting patterns between species we also generated an atlas of human imprinting. The lack of engineered crosses in human necessitates a more complex approach to identify parental-specific expression. Two major causes of autosomal ASE are imprinting and genetic variants affecting expression through *cis*-regulation. ASE caused by genetic variants typically leads to a consistent expression bias from the same allele in heterozygous individuals (at most ~50% of individuals). In contrast, imprinted genes have ASE in all individuals, but without bias toward any particular allele (**Fig. 2a**). With ASE data from many individuals, these differences could potentially allow identification of imprinted genes.

We measured ASE in 1,687 RNA-Seq samples from 45 tissues in 178 individuals (GTEx.v3[21]; **Supplementary Table 6**). We previously showed that concordance in ASE between two samples measured at the same SNP underestimates error due to systematic biases[11]. Therefore, we calibrated our parameters on concordance of ASE between different SNPs within the same gene. We found excellent agreement between genotyped and imputed SNPs (Pearson r^2=0.94; **Supplementary Fig. 7**), demonstrating high accuracy of genotype imputation, phasing, and quantification of ASE. We found that, similar to mouse, imprinted genes are highly over-represented among monoallelically expressed genes (**Fig. 2b**; 7/76 known imprinted genes among the top 12). Detection of imprinted genes was further improved by eliminating genes with consistent ASE directionality across individuals (likely due to *cis*-regulatory variants) (**Fig. 2c**) and incorporating allele-specific methylation[22] (ASM) data (**Fig. 2d**).

To assess the accuracy of our predictions, we analyzed RNA-Seq data from lymphocytes derived from 17 members of a three-generation family[23]. This pedigree allowed us to identify imprinted genes, since their direction of ASE depends on each allele's parent-of-origin (**Supplementary Fig. 8, Supplementary Fig. 9**; see Methods), and to estimate a false discovery rate (FDR) for our GTEx scoring scheme (**Fig. 2d**). This represents a general

approach that could be applied to discover imprinted genes whenever multi-generation expression and genotype data are available.

As in mouse, we identified most genes known to be imprinted in human (63/76; **Fig. 2e**; **Supplementary File 5**). The majority of the genes that we missed did not meet our stringent criteria (e.g. four expressed heterozygous SNPs). Only a few novel human imprinted genes reached a level of significance comparable to well-established imprinted genes, supporting the expectation that the majority of genes imprinted in adult tissues have already been discovered. We identified 17 strong candidates (**Fig. 2e**) at a level corresponding to a 1% FDR, and 100% ASE validation by microfluidics-based multiplex PCR (Supplementary Fig. 10, Supplementary Fig. 11, Supplementary Fig. 12, Supplementary Table 7, Supplementary Note). *NHP2L1, PMF1, PPIEL*, and *ZNF595* were previously predicted on the basis of strong ASM[22,24], but it was their patterns of ASE (i.e. RNA score) that pushed them to significance in our study, and distinguished them from many other genes with ASM but no evidence of ASE. Similar to mouse, when a gene was imprinted in adult tissues, it tended to have a consistent strong allelic bias in all tissues sampled. Nonetheless, distinct patterns of allelic expression bias were enriched for functions including development via hedgehog signaling, kidney development, skeletal system development, regulation of growth, and synaptosome localization (**Supplementary Table 8**). Our novel imprinted genes regulate glucose import in response to insulin (*PID1*), glucagon signaling and feeding behavior (*GNG7*), growth (*PMF1*), and are associated with birth weight (*DHFR*) and type II diabetes (*MYO1D*) (**Supplementary Table 3**). As in mouse, ubiquitously monoallelic genes (the first 19 genes in **Fig. 2e**) are highly enriched for oxytocin/vasopressin neuropeptide activity and genes governing eating behavior (**Supplementary Fig. 13**).

We identified several conserved properties of genomic imprinting between human and mouse. The dichotomy of imprinting between neural and non-neural adult tissues is shared (**Fig. 3a; Fig. 2e; Supplementary Fig. 3**), suggesting a conserved role for imprinting in neural function. We also found that genes that are imprinted in both species have stronger allelic bias but similar imprinting breadth compared to species-specific imprints (**Fig. 3b**). Interestingly, among the most strongly conserved imprinted genes, "response to growth factor stimulus" was the most highly enriched function (**Supplementary Fig. 14**), consistent with the theory that imprinting evolved due to genetic conflict over nutrient allocation[3]. Among 41 genes imprinted in mouse but not human, we observed an excess of maternally expressed genes (61%). This is consistent with theoretical predictions that silencing of paternally derived alleles should be less evolutionarily stable, due to maternal alleles having greater control over the *in utero* environment[25]. If indeed paternal silencing is less stable, it should be less common overall—despite being enriched in species-specific imprints—which is indeed the case (average 35% maternal expression/paternal silencing across all mouse tissues).

If imprinted genes are indeed often involved in genetic conflict, pairs of maternally and paternally expressed genes with opposing roles may co-evolve in evolutionary "arms races", possibly leading to co-expression of antagonistic gene pairs[26]. Consistent with this hypothesis, we observed an excess of maternal/paternal pairs among the most strongly co-expressed imprinted genes in mouse (excluding genes in close genomic proximity; see

Methods) (**Supplementary Fig. 15; Supplementary Table 9**). Many of the strongest maternal/paternal co-expressed pairs also have reciprocal imprinting patterns (**Fig. 3c; Supplementary Fig. 16**) and opposing functions. For instance, *Magel2* and *Calcr* are a co-expressed pair involved in neuropeptide hormonal signaling. *Magel2* loss of function results in poor suckling and neonatal growth retardation[27] while *Calcr* affects appetite suppression through amylin regulation[28]. *Igf2r* and *Phf17* have also been linked to regulation of growth: *Igf2r* suppresses growth[29], while *Phf17* promotes vasculogenesis of the placenta[30], where it is preferentially imprinted (**Fig. 3c**). The human orthologs of the maternal/paternal pairs were co-expressed as well (**Supplementary Fig. 17**), suggesting conservation of these antagonistic interactions.

An additional prediction of the conflict/arms race model is that the expression levels of imprinted genes may increase due to positive selection, in response to increases in their antagonistic counterparts. To investigate this, we computed the expression divergence (Euclidean distance[31]) between all mouse-human orthologs in 14 tissues profiled in both species. We found a higher rate of divergence among imprinted genes with the strongest allelic bias (Fig. 3d; Supplementary Fig. 18, Supplementary Fig. 19; see Methods). To test the possibility that strong imprinting itself causes variable expression, we searched for a similar pattern among human individuals, but actually found less expression variation among imprinted genes (**Fig. 3e**). This pattern of high interspecies divergence, coupled with low intraspecies variation, is consistent with the idea that positive selection contributed to their divergence.

To test whether this rapid expression divergence reflects up-regulation, as predicted by the conflict/arms race model, we compared imprinted gene expression levels in human/mouse with the levels of their orthologs in platypus and chicken, where they are not likely to be imprinted[32]. Of ten imprinted genes with 10-way 1:1 orthologs across amniotes, nine had higher expression in human/mouse vs. platypus/chicken (**Fig. 3f:** *GRB10* shown as an example; **Fig. 3g:** all data; *p*=0.0033, see Methods). This difference is in spite of only being expressed from one allele in human/mouse (and thus having 50% lower expected expression, all else being equal), consistent with up-regulation due to antagonistic co-evolution.

In conclusion, our human and mouse imprinting atlases have revealed the patterns of imprinting – across development, tissues, and species – in unprecedented detail. Tissue-specific imprinting is surprisingly rare, with most genes either imprinted in all adult tissues where they are expressed, or in none. In addition, genetic conflict between imprinted loci can explain several key observations: co-expression of maternally/paternally expressed genes, rapid expression divergence, and an overall pattern of up-regulation associated with imprinting. We expect that these resources will be instrumental in refining our understanding of imprinting mechanisms at individual loci and that similar atlases in other species will improve our understanding of the origins of imprinting.

METHODS

Identifying and quantifying imprinting in mouse

Tissues were dissected from 16 reciprocally crossed male C57Bl/6J × CAST/EiJ F1 mice (8 mice in each direction of cross) aged 35-45d and two embryonic stages (**Supplementary Table 1**). Animals were housed and euthanized in accordance with the current Animal Use Protocol approved by the Faculty of Medicine and Pharmacy Animal Care Committee at University of Toronto. Tissues were rinsed in PBS and snap-chilled in liquid nitrogen within 10 minutes of dissection. RNA was extracted with Trizol (Invitrogen) according to the manufacturer's recommendations and integrity confirmed by Bionalyzer (RIN>6). Sequencing libraries were prepared using TruSeq v2 RNA-Seq kits (RS-122-2001; Illumina) according to the manufacturer's recommendations. 14 libraries (7 tissues; see **Supplementary Table 1**) were treated with UNG nuclease to retain strand specificity in the sequenced libraries[34]. Libraries were indexed and sequenced on a HiSeq 2000 (PE90) to an average 7.2 Gb/sample at BGI (Hong Kong, China). Genomic imprinting was quantified using previously established criteria[11]. In brief, reads were aligned with Novoalign (Novocraft) against gene models (**Supplementary File 6**), all possible splice junctions representing up to two exon skipping events, and the mouse genome (mm9). Reads mapping to opposite strands were analyzed independently for libraries with strand-specificity. 19.6 million C57Bl/CAST SNPs, representing the union of two collections[35-36], were masked prior to alignment in order to minimize alignment biases (**Supplementary Fig. 21**). Allele-specific expression was quantified for each gene by counting uniquely mapped read-pairs within the gene boundaries (incl. introns) that overlapped at least one SNP, such that allelic origin could be discerned. Probability of ASE was estimated from the cumulative binomial distribution with random expectation set to 50% (i.e. 50% expression from both alleles), consistent with observed global ASE distributions (**Supplementary Fig. 20**). An Imprinting Score (*IS*), which was previously shown to reliably distinguish genuine imprinting events[11], was computed as the $\log_{10}$ of the less significant binomial *p*-value of the two reciprocally crossed tissues and paternal bias was arbitrarily set to be negative. Significance was established by comparing *IS*s to background *IS*s computed from biological replicates with the same parental background[11] (i.e. not from a reciprocal cross). We note that incomplete imprinting may be due to mixtures of imprinted and non-imprinted cell types within individual tissue samples. Known mouse imprinted genes were compiled from literature[11,18] and are available as a BED track along with novel genes discovered in this study and genes with conflicting evidence from this study and literature removed (**Supplementary File 4**). The seven samples sequenced in previous studies (**Fig. 1**) were analyzed in parallel from the fastq files. Putative novel imprinted genes (*p*<1e-2, >50% ASE in both crosses) were validated by pyrosequencing as previously described[11]. In brief, imprinting was considered validated if the difference in allelic bias between the reciprocal samples was >5% (2 standard deviations of 182 biological replicate DNA measurements) and each ratio was reciprocally biased (i.e. in opposite directions) relative to the ratio obtained with DNA. Primer sequences and assay details are available in **Supplementary Table 2**. Raw data are available in the Sequence Read Archive under accession SRP020526. Functional enrichments were identified using hypergeometric tests against the JAX Phenome database[37], mouse Gene Ontology[38], human Reactome[39], GeneGO[40], and Biobase[41]

pathways mapped onto mouse genes via ENSEMBL orthologs[42]. All databases were downloaded between May 10 and Jun 20, 2013. FDR was established as the proportion of randomly sampled background sets (selected from genes powered to detect ASE) of the same size that exceeded significance of the gene set of interest.

Identifying and quantifying imprinting in human tissues

GTEx.v3[21] imputed genotype and aligned RNA-Seq (BAM) data were downloaded from dbGaP (phs000424.GTEx.v3.p1.c1) on August 27, 2013. 1,687 samples had both RNA-Seq and genotypes available, comprising 45 unique tissues/organs from 178 individuals (**Supplementary Table 4**). No minimum number of individuals per tissue was required. Genotypes were phased using SHAPEIT.v2[43] with genomic maps (human b37) from the author's website downloaded on August 27, 2013 and all options set to default. Phased ASE read counts were compiled at all heterozygous sites using sra-pileup (v2.3.2 from SRA Toolkit) and custom scripts. Gene models were constructed from RefSeq transcripts, ENSEMBL transcripts, UCSC gene models, GenBank mRNAs, and computational gene predictions by collapsing overlapping transcripts on the same strand and iteratively adding new models; starting with collapsed RefSeq transcripts, then collapsed ENSEMBL transcripts completely outside RefSeq annotated boundaries, and similarly UCSC genes, mRNAs, NSCAN and GENSCAN gene predictions, respectively, in a total of six iterations. Known imprinted genes in human were compiled from literature[16,18], and are available in BED format (**Supplementary File 7**). Gene-level ASE was quantified by aggregating counts across all phased heterozygous sites within gene boundaries such that each gene is assigned two integers for each sample: a_{ig} and b_{ig}, representing expression from allele a and allele b, in sample i and gene g. Although assignment of a and b is arbitrary, phasing ensures that, when possible, assignment of alleles is consistent across individuals, thus enabling assessment of conserved allelic expression bias. Allelic Bias (AB_{ig}) was quantified for each gene in each tissue as $a_{ig}/(a_{ig}+b_{ig})$ and the RNA-Score (RS_{ig}) computed as $\Sigma|AB_{ig}|-|\Sigma AB_{ig}|$, for $i=1..n$ (n=number of subjects where that tissue was sequenced). More complicated combinations (e.g. adding regression-derived weights to the two main terms) did not improve performance as gauged by ROC/AUC analysis of known imprinted genes and assuming all negatives are true-negatives (e.g. **Fig. 2d**). An allele-specific methylation score ($ASMS$) was assigned to each gene by overlapping allele-specific methylation coordinates from 22 samples[22] with gene models above and tallying the number of samples supporting ASM within the gene boundaries (incl. UTRs and introns). Cultured and uncultured differentiated cells did not have different distributions of ASM among known imprinted genes (**Supplementary Fig. 22**) and extending gene models to capture potentially intergenic regulation did not improve performance (**Supplementary Fig. 23**). Multiple linear regression was used to compute a combined score (CS) by determining $ASMS$ and RS weights (mean $ASMS$ weight=0.29, mean RS weight=0.11) that best distinguish known imprinted genes from all other genes within each tissue. Negative CS values were regarded as a lack of evidence for imprinting. Precision, typically defined as the proportion of all positive calls at some score threshold that are true, was more broadly defined since we observed that it was impacted by both the score and the number of tissues exceeding that score (i.e. an imprint was more likely to be real if supported by multiple tissues). Precision

was estimated for each gene by choosing a *CS* threshold that optimized precision, given the number of tissues that exceed that threshold (**Supplementary Fig. 10**).

Pedigree imprinting analysis

Genotype data (generated from whole-genome sequencing by Complete Genomics) and BAM files of mapped RNA-Seq data[23] (generated from lymphocytes) were downloaded for 17 individuals from a CEPH/Utah three-generation pedigree (NA1463; GSE56961). The two grandparent/parent trios and one parent/children pedigree were phased separately with SHAPEIT.v2[43] and heterozygous SNPs within gene boundaries were used to track inheritance of child alleles back to the grandparents [when possible; 67.9% and 68.5% of expressed genes in lymphocytes with at least one expressed heterozygote (i.e. ASE-powered; n=17,855) were traceable from each child back to the two sets of grandparents, and 94.1% of ASE-powered genes were traceable in at least one of 11 children for at least one allele]. ASE was quantified as described above under "Identifying and quantifying imprinting in human tissues". Imprinting was considered validated when all of the following criteria were met: 1) at least one allele switched between preferential expression to preferential silencing consistent with its parent-of-origin (e.g. expressed in the father when inherited from his mother, but then silent in father's children), 2) parent-of-origin bias was consistent in all children and parents, 3) at least three children were powered for detection of ASE. A minimum of ten ASE-informative reads were required and bias was considered present when a binomial test (cumulative distribution function, as described previously[11]) yielded a significance of $p<0.01$ comparing the ASE read counts of the two alleles.

ASE validation in human samples

ASE validation was performed on 9 genes in 9 subjects across 42 unique tissues (average 14.2 tissues/subject; 745 total measurements). Tissue samples were contributed by the GTEx consortium[21] (see Acknowledgements) and processed as described previously[44] with some modifications (**Supplementary Note**). One heterozygous SNP was identified for each gene and subject; preference was given to heterozygous SNPs identified by exome-sequencing, genotyped variants when exome SNPs were not detected, and imputed SNPs if no other evidence was available. Allelic ratios were quantified as described previously[44] (see **Supplementary Table 2** for design details).

Human-Mouse comparisons

Mean allelic bias was quantified for each gene as the sum of allelic biases across all tissues where imprinting was detected (max(p)<0.01 from the two reciprocal crosses in mouse with consistent parent-of-origin bias and *CS*>0 in humans; see sections above for more details). Each tissue contributes a value between 0 (100% biallelic) and 1 (100% monoallelic).

Enrichment for maternal/paternal expressed genes among similar imprinting patterns

Gene expression was quantified as arcsinh (similar to the natural log but allowing for values of zero) of FPKM (Fragments Per Kilobase per Million uniquely aligned reads that overlap gene) and was quantile-normalized in each species separately. These data were used to compute distances between all unique pairwise combinations of genes. The distance

similarity metric (*dist.sim*) used for all comparisons was standard Euclidean distance. In the comparisons of numbers of maternal/paternal interactions, an interaction was defined between two genes if they did not exceed the maximum *dist.sim* threshold. Interactions between genes less than 1 MB apart were not considered to avoid effects from shared *cis*-regulation (e.g. two genes affected by the same DMR). When multiple interactions existed between genes within a cluster (within 1 MB of each other) and one other gene, only the lowest *dist.sim* value was considered. Two interaction types were considered: *mm/pp* (maternal-maternal or paternal-paternal), and *mp/pm* (maternal-paternal or paternal-maternal). Significance of deviation from the null was tested with a binomial test (cumulative), where the null expectation was set to the ratio of all possible *mm/pp* pairs to *mp/pm* pairs (after removing interactions within the same genomic regions).

Divergence of gene expression among imprinted genes

Gene expression was quantified for all human/mouse orthologs[42] as described above in 15 pairs of matching tissues. Mean FPKM was used when multiple individuals were sequenced for the same tissue. Stomach was excluded since the mouse-human counterparts did not cluster next to each other, possibly due to the human dissection including muscular tissue (**Supplementary Fig. 24**), leaving 14 pairs for analysis. Data were median-subtracted (for each gene: median across all tissues subtracted from its value) and quantile-normalized within each species separately, then merged and quantile-normalized again to minimize species-specific biases. Imprinting strength was measured as the sum of ASE across all tissues, such that each tissue can contribute a value of 0 (no ASE) to 1 (100% monoallelic). Divergence was measured using Euclidean distance between each set of orthologs and was quantified as a *z*-score (number of standard deviations from mean) relative to a background set matched for degree of expression. A gene was included in the background if it was expressed in the same number of tissues (non-zero expression post-median subtraction) with the highest FPKM value being within 10% of the highest FPKM of the imprinted gene; the background was further trimmed such that the same number of genes was above and below the imprinted maximal expression. There was no association between background scores and imprinting breadth. Imprinting breadth is the sum of allelic bias across all tissues. The sum of Euclidean distances between matched tissue pairs was also used to quantify divergence, where the inputs were groups of genes and matched background sets comprised randomly substituted genes from all orthologs as described above. Human-human comparisons (**Fig. 3e**) were run on mean expression across two random subsets of individuals for the top 15, 20, and 25 genes sorted on strength of ASE (see above). For comparison, random sets of genes were also analyzed in the same way.

Comparison of imprinted gene expression to non-imprinted orthologs

Normalized gene expression data from six tissues from 10 species[45] was averaged across biological replicates (disregarding sex). Among 41 genes imprinted in both human and mouse, 10 were represented among 10-way 1:1:...:1 orthologs. For each gene, an average log2 ratio of mean(human,mouse)/mean(chicken,platypus) RPKM was computed across all tissues where the gene was imprinted in both human and mouse (43 total gene-tissue pairs). To quantify significance, this was repeated 10,000 times on randomly selected non-imprinted genes using expression from the same tissues.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Barlow DP. Gametic imprinting in mammals. Science. 1995; 270:1610–3. [PubMed: 7502071]

2. Barlow DP, Bartolomei MS. Genomic imprinting in mammals. Cold Spring Harb Perspect Biol. 2014; 6

3. Moore T, Haig D. Genomic imprinting in mammalian development: a parental tug-of war. Trends Genet. 1991; 7:45–9. [PubMed: 2035190]

4. Spencer HG, Clark AG. Non-conflict theories for the evolution of genomic imprinting. Heredity (Edinb). 2014

5. Wilkins JF, Haig D. What good is genomic imprinting: the function of parent- specific gene expression. Nat Rev Genet. 2003; 4:359–68. [PubMed: 12728278]

6. Curley JP, Barton S, Surani A, Keverne EB. Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. Proc Biol Sci. 2004; 271:1303–9. [PubMed: 15306355]

7. Wolf JB, Hager R. A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol. 2006; 4:e380. [PubMed: 17105351]

8. Babak T. Identification of imprinted loci by transcriptome sequencing. Methods Mol Biol. 2012; 925:79–88. [PubMed: 22907492]

9. Babak T, et al. Global survey of genomic imprinting by transcriptome sequencing. Curr Biol. 2008; 18:1735–41. [PubMed: 19026546]

10. Wang X, et al. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. PLoS One. 2008; 3:e3839. [PubMed: 19052635]

11. DeVeale B, van der Kooy D, Babak T. Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet. 2012; 8:e1002600. [PubMed: 22479196]

12. Gregg C, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science. 2010; 329:643–8. [PubMed: 20616232]

13. Tran DA, Bai AY, Singh P, Wu X, Szabo PE. Characterization of the imprinting signature of mouse embryo fibroblasts by RNA deep sequencing. Nucleic Acids Res. 2014; 42:1772–83. [PubMed: 24217910]

14. Calabrese JM, et al. Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell. 2012; 151:951–63. [PubMed: 23178118]

15. Wang X, Soloway PD, Clark AG. A survey for novel imprinted genes in the mouse placenta by mRNA-seq. Genetics. 2011; 189:109–22. [PubMed: 21705755]

16. Schulz R, et al. WAMIDEX: a web atlas of murine genomic imprinting and differential expression. Epigenetics. 2008; 3:89–96. [PubMed: 18398312]

17. Kim J, Bergmann A, Wehri E, Lu X, Stubbs L. Imprinting and evolution of two Kruppel-type zinc-finger genes, ZIM3 and ZNF264, located in the PEG3/USP29 imprinted domain. Genomics. 2001; 77:91–8. [PubMed: 11543637]

18. Morison IM, Ramsay JP, Spencer HG. A census of mammalian imprinting. Trends Genet. 2005; 21:457–65. [PubMed: 15990197]

19. Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. Nucleic Acids Res. 2001; 29:275–6. [PubMed: 11125110]

20. Schaller F, et al. A single postnatal injection of oxytocin rescues the lethal feeding behaviour in mouse newborns deficient for the imprinted Magel2 gene. Hum Mol Genet. 2010; 19:4895–905. [PubMed: 20876615]

21. GTEx-Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013; 45:580–5. [PubMed: 23715323]

22. Fang F, et al. Genomic landscape of human allele-specific DNA methylation. Proc Natl Acad Sci U S A. 2012; 109:7332–7. [PubMed: 22523239]

23. Li X, et al. Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. Am J Hum Genet. 2014; 95:245–56. [PubMed: 25192044]

24. Court F, et al. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of the human imprintome and suggests a germline methylation independent establishment of imprinting. Genome Res. 2014

25. Wilkins JF, Haig D. Parental modifiers, antisense transcripts and loss of imprinting. Proc Biol Sci. 2002; 269:1841–6. [PubMed: 12350273]

26. Wilkins JF, Haig D. Genomic imprinting of two antagonistic loci. Proc Biol Sci. 2001; 268:1861–7. [PubMed: 11564340]

27. Bischof JM, Stewart CL, Wevrick R. Inactivation of the mouse Magel2 gene results in growth abnormalities similar to Prader-Willi syndrome. Hum Mol Genet. 2007; 16:2713–9. [PubMed: 17728320]

28. Potes CS, Lutz TA. Brainstem mechanisms of amylin-induced anorexia. Physiol Behav. 2010; 100:511–8. [PubMed: 20226802]

29. Wutz A, et al. Non-imprinted Igf2r expression decreases growth and rescues the Tme mutation in mice. Development. 2001; 128:1881–7. [PubMed: 11311167]

30. Tzouanacou E, Tweedie S, Wilson V. Identification of Jade1, a gene encoding a PHD zinc finger protein, in a gene trap mutagenesis screen for genes involved in anteroposterior axis development. Mol Cell Biol. 2003; 23:8553–2. [PubMed: 14612400]

31. Pereira V, Waxman D, Eyre-Walker A. A problem with the correlation coefficient as a measure of gene expression divergence. Genetics. 2009; 183:1597–600. [PubMed: 19822726]

32. Renfree MB, Hore TA, Shaw G, Graves JA, Pask AJ. Evolution of genomic imprinting: insights from marsupials and monotremes. Annu Rev Genomics Hum Genet. 2009; 10:241–62. [PubMed: 19630559]

33. Xie W, et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell. 2012; 148:816–31. [PubMed: 22341451]

34. Parkhomchuk D, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 2009; 37:e123. [PubMed: 19620212]

35. Frazer KA, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature. 2007; 448:1050–3. [PubMed: 17660834]

36. Keane TM, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011; 477:289–94. [PubMed: 21921910]

37. Grubb SC, Bult CJ, Bogue MA. Mouse phenome database. Nucleic Acids Res. 2014; 42:D825–34. [PubMed: 24243846]

38. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25:25–9. [PubMed: 10802651]

39. Croft D, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014; 42:D472–7. [PubMed: 24243840]

40. Gene GO. Thompson Reuters. 2009

41. BioBase. BioBase Corp.; 2009.

42. Guberman JM, et al. BioMart Central Portal: an open database network for the biological community. Database (Oxford). 2011 2011, bar041.

43. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2012; 10:5–6. [PubMed: 23269371]

44. Zhang R, et al. Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. Nat Methods. 2014; 11:51–4. [PubMed: 24270603]

45. Brawand D, et al. The evolution of gene expression levels in mammalian organs. Nature. 2011; 478:343–8. [PubMed: 22012392]
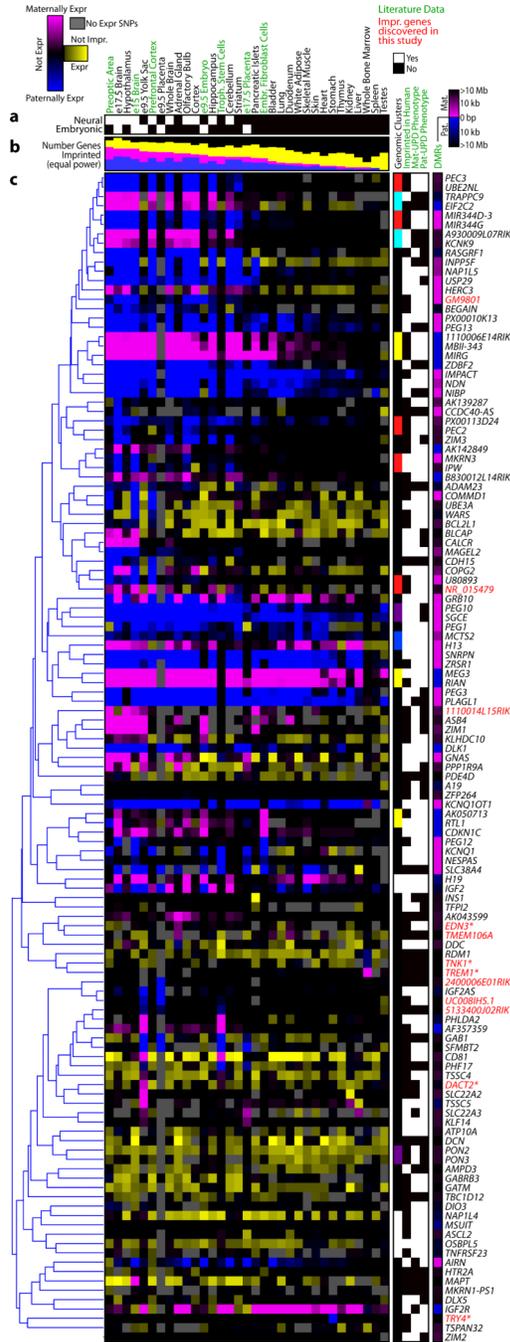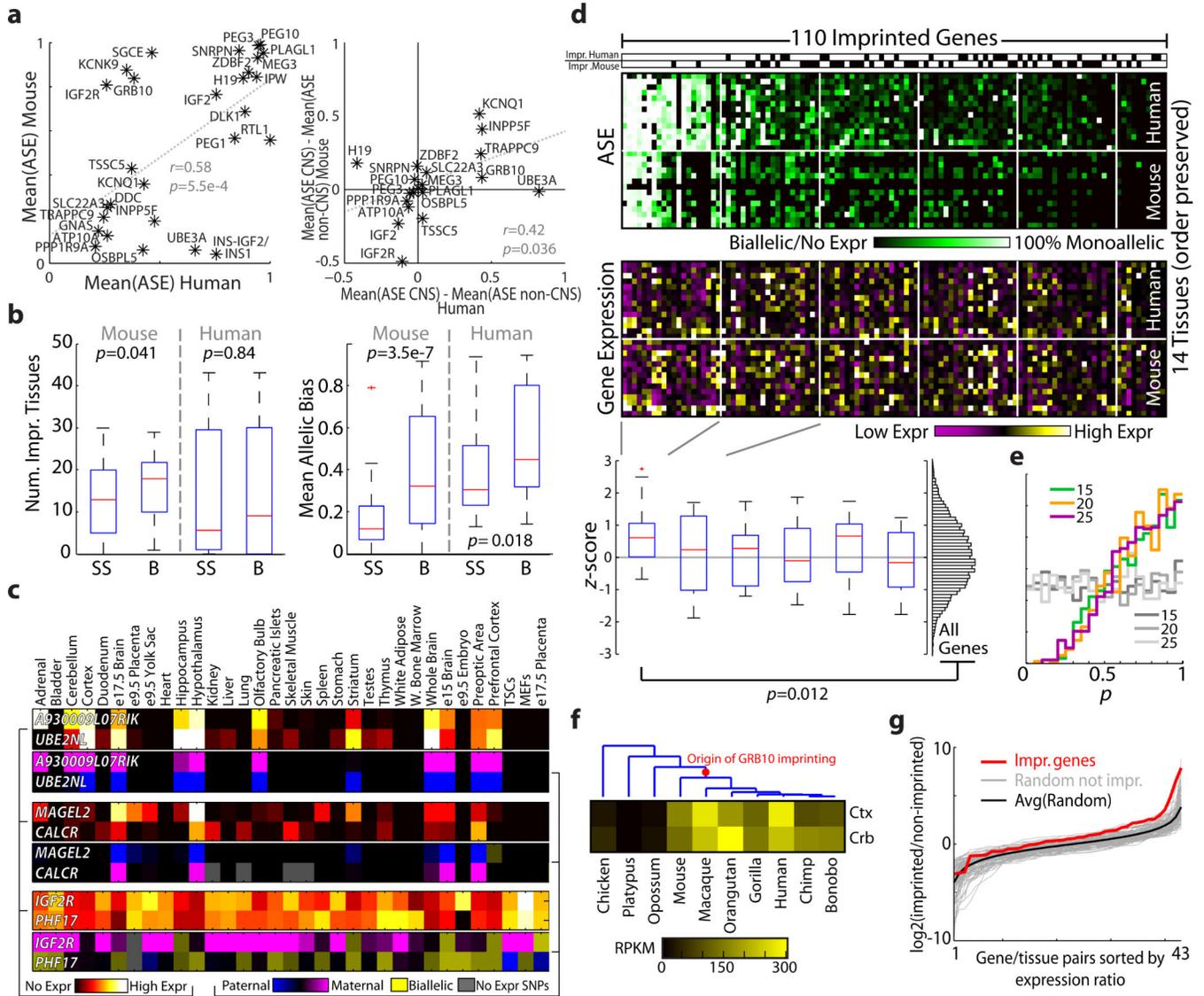
Figure 1.

Atlas of genomic imprinting in mouse. All known and novel validated imprinted genes with at least one expressed SNP in both crosses are shown. (**a**), Tissue type. (**b**), Proportion of genes imprinted, detected using the same number of allele-specific sequencing reads in all samples. (**c**), Atlas generated using all sequencing reads. Genes are colored by their Imprinting Score (blue/pink) when allelic counts supported a parent-of-origin bias, and on their levels of gene expression in yellow (asinh(FPKM)) when parent-of-origin bias was absent. The *y*-axis was clustered treating parent-of-origin expression (blue/pink)

equivalently by setting imprinting scores to positive values and non-imprinted expression (yellow) to negative values, thereby grouping similarly expressed and imprinted genes together. The *x*-axis was sorted on the proportion of genes imprinted. Maternal expression in e9.5 placenta is not shown since we could not reliably exclude signal stemming from contaminating maternal tissue (**Supplementary Fig. 20**). Genomic clusters of at least two genes (within 1 MB) were each assigned a unique color, shown on the right when these genes also cluster by imprinting pattern. ASE data used to generate the plot are available in **Supplementary File 1**. Previously published samples: Pre-optic area[12], e15 Brain[12], Prefrontal cortex[12], e9.5 Embryo[9], Trophoblast stem cells[14], e17.5 Placenta[15], Embryonic fibroblasts[13], differentially methylated regions (DMRs)[33], uniparental disomy phenotypes (Mousebook/Harwell Phenotype Maps) (e.g. Mat-UPD indicates that gene is within a region that affects a phenotype when both copies of the region are maternal). *Promising novel imprinted gene candidates only imprinted in one tissue (**Supplementary Table 3**).
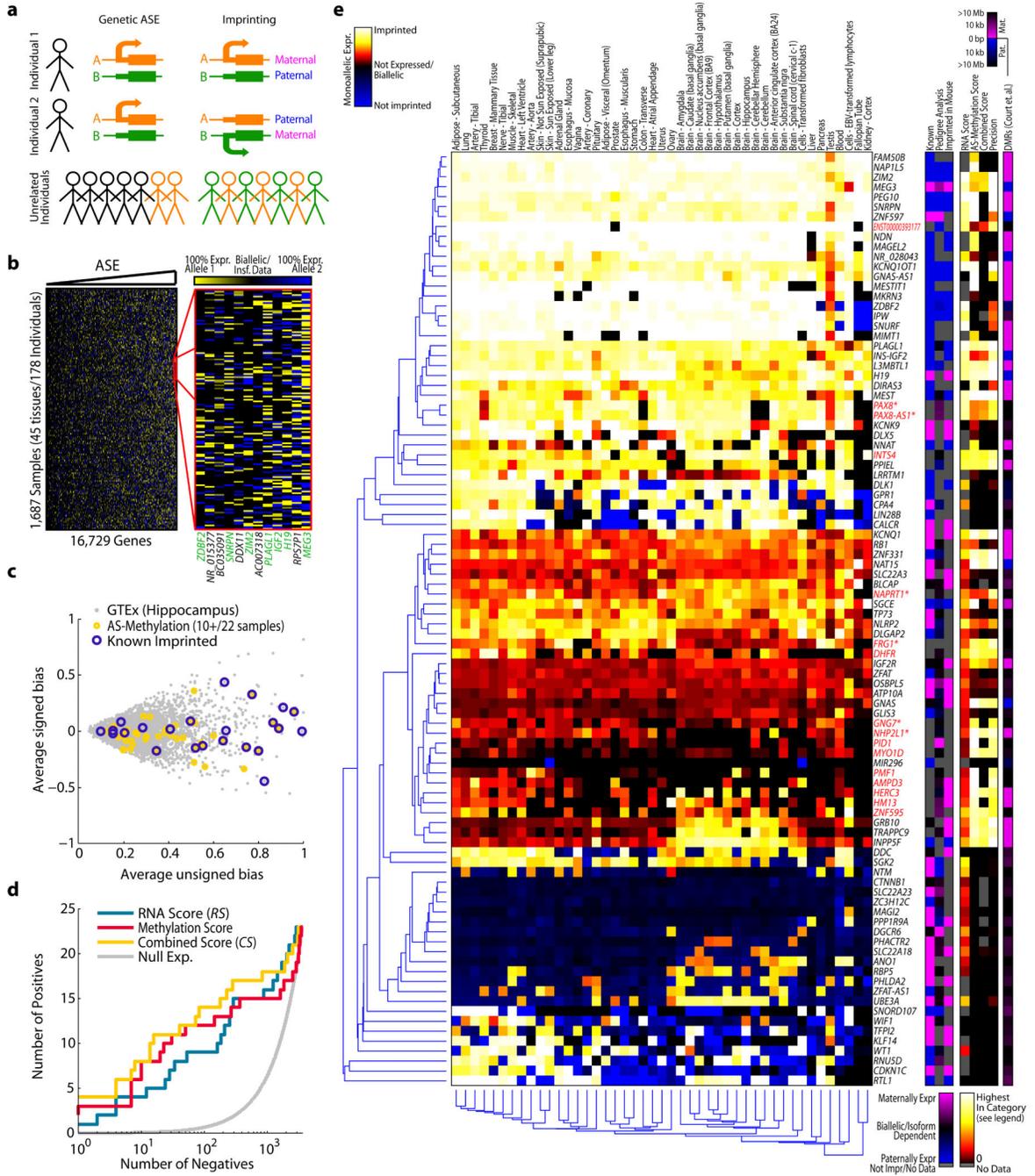
Figure 2.

Atlas of genomic imprinting in human. (**a**), ASE caused by genetic polymorphisms will tend to be biased towards the same allele (orange) in heterozygotes, which are at most 50% of individuals. In contrast, ASE due to imprinting will be present in all individuals, but will not favor either allele. (**b**), ASE for all genes powered (see Methods) in all available GTEx.v3 samples. ASE is scaled from −1 (100% expression of one allele) to +1 (100% expression from the other allele) and sorting genes on Σ yielded 7 known imprinted genes (green) among the top 12. (**c**), Resolving mean(|ASE|) further by plotting against mean(ASE)

reveals the tendency for imprinted genes to switch bias between alleles (i.e. $y \approx 0$) and for the strongest imprinted genes to have allele-specific methylation[22] (ASM; shown are genes identified in 10/22 or more biological samples) marks. (**d**), Number of known imprinted genes, detected among all genes sorted using various scoring schemes. Combining ASM and RNA scores (into *CS*) improves overall performance. Hippocampus shown in (**c-d**), other tissues behaved similarly. (**e**),Monoallelic expression of imprinted genes in 45 human tissues. Genes with a *CS*<0 (no evidence for imprinting) are colored blue. Clustering was done using Manhattan distance. Right panels: Pedigree analysis shown as average parent-of-origin bias of 2 parents and 11 children. Precision is the proportion of positive calls that are known imprinted genes given the 45 tissue-specific *CS* values to establish a threshold (see Methods). *Validated by mmPCR-Seq. Known (black): Geneimprint, novel (colored in red).

Figure 3.

Species comparisons of imprinting. (**a**), Conservation of allelic bias between human and mouse. Average allelic bias is generally conserved between the two species. Tissue-specific imprinting was strongest between CNS and non-CNS samples and is conserved between species. (**b**), Genes that are imprinted in both species (B; n_{human}=47, n_{mouse}=54) have stronger allelic bias relative to species-specific (SS; n_{human}=47, n_{mouse}=71) imprinted genes (median in red, box delineates 25th/75th quartiles, whiskers show range of non-outliers) (**c**), Examples of top maternal/paternal pairings (also see **Supplementary Fig. 16**). (**d**), Strongly imprinted genes (high allelic bias in many tissues) have more divergent gene expression between human and mouse relative to all genes (Wilcoxon p=0.012; **Supplementary Fig. 18, Supplementary Fig. 19**; box plot metrics same as **b**). z-scores were computed against gene expression divergence of randomly selected genes matched for breadth of expression (see Methods). $Expr$ = asinh(FPKM), median subtracted for each gene within species. (**e**), Significance of within-human expression variation comparisons for top 15, 20, and 25 most

strongly imprinted genes (1,000 permutations); colors=strongest imprinting bins from **d**, grayscale=genes randomized to demonstrate null. (**f**) Example of a gene with elevated expression in species and tissues where it is imprinted. Ctx=cortex, Cer=cerebellum. (**g**) Sorted log-ratios of mean mouse/human (imprinted) expression vs. mean platypus/chicken (not imprinted) orthologs in gene-tissue combinations where imprinted in both human and mouse. Expression was elevated in mouse/human (*p*=0.0033; c.f. randomly selected non-imprinted genes in the same tissues).