nature genetics

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

Tomas Babak[1,6], 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] & Hunter B Fraser[1]

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 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 coexpressed pairs of maternally and 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 the discovery of imprinting in any species and sheds light on the causes and consequences of genomic imprinting in mammals.

Despite over 20 years of study[2–5], evolutionary explanations for genomic imprinting remain controversial. The conflict/kinship theory posits that imprinting evolved as a result of 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 the co-adaptation of mutually favorable traits in parent and offspring[6,7], among others[4], and it is not clear whether imprinting can be entirely explained by one model.

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 Figs. 1–3** and **Supplementary Data Set**). We detected imprinting as allele-specific expression (ASE) consistently biased toward either the maternal or paternal allele in reciprocally crossed F_1 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] (Online Methods). We sequenced the mRNA from 26 unique tissues and developmental stages (61 biological samples) and combined our data with those from 7 additional published tissues[9,12–15] (**Supplementary Table 1**). For the 207 imprinting measurements that have previously been reported[16], in the gene-tissue pairs assayed here, data for 95.6% of the loci agreed (**Supplementary Fig. 4a**); 5 of the 9 cases that disagreed were in tissues sampled at different developmental time points, and the remaining 4 cases included some equivocal evidence (for example, claims of imprinting without confirmation from a reciprocal cross[17]). We confirmed the reported non-canonical maternal expression of *Igf2* and the paternal expression of *Grb10* in adult brain[12], and we found this reciprocal pattern in all central nervous system (CNS) tissues (**Fig. 1** and **Supplementary Figs. 2** and **3**). We observed both paternal and maternal expression for *Copg2* and *Rtl1* as well (**Fig. 1** and **Supplementary Figs. 2** and **3**). Overall, our atlas increased the number of reported gene-tissue imprinting measurements[16,18] by nearly an order of magnitude (**Supplementary Fig. 4b** and **Supplementary Data Set**).

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

[1]Department of Biology, Stanford University, Stanford, California, USA. [2]University of California San Francisco School of Medicine, University of California, San Francisco, San Francisco, California, USA. [3]Department of Pathology, Stanford University, Stanford, California, USA. [4]Department of Genetics, Stanford University, Stanford, California, USA. [5]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [6]Present address: Department of Biology, Queen's University, Kingston, Ontario, Canada. Correspondence should be addressed to H.B.F. (hbfraser@stanford.edu).

Received 28 April 2014; accepted 13 March 2015; published online 13 April 2015; doi:10.1038/ng.3274

Figure 1 Atlas of genomic imprinting in mouse. All known and new validated imprinted genes with at least one expressed SNP in both reciprocal crosses are shown. (**a**) Tissue types analyzed. (**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 (IS; blue or pink) when allelic counts supported a parent-of-origin bias and by their level of gene expression (yellow; asinh(FPKM)) when parent-of-origin bias was absent. The *y* axis was clustered treating parent-of-origin expression (blue or 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 embryonic day (E) 9.5 placenta is not shown because we could not reliably exclude signal stemming from contaminating maternal tissue (**Supplementary Fig. 20**). Genomic clusters of at least two genes (within 1 Mb of each other) were each assigned a unique color, shown on the right, when these genes also clustered by imprinting pattern. The ASE data used to generate the plot are available in the **Supplementary Data Set**. 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] and uniparental disomy phenotypes (MouseBook; Harwell Phenotype Maps) (for example, maternal UPD (uniparental disomy) indicates that the gene is within a region that affects a phenotype when both copies of the region are maternal). Asterisks indicate promising new imprinted gene candidates only imprinted in one tissue (**Supplementary Table 3**).

(**Fig. 1**, **Supplementary Figs. 2** and **3**, and **Supplementary Data Set**). 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]. Of the genes that were imprinted in some tissues and biallelically expressed in others, 52 of 55 were imprinted in embryos but not in adults (the remaining 3 genes—*Phf17* (*Jade1*), *Gab1* and *Slc22a3*—were 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 because of shared *cis* regulatory elements[2]. We found a number of significant functional enrichments (false discovery rate (FDR) < 0.05) within clusters of co-imprinted genes, including growth (for example, decreased fetal weight), nutrient processing (for example, glucose transport and uptake), and CNS development and signaling (for example, nerve growth factor signaling) (Online 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 between this signaling and the 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 humans necessitates a more complex approach to identify parental-specific expression. Two major causes of autosomal ASE are imprinting and genetic variants affecting expression through regulation in *cis*. ASE caused by genetic variants typically leads to a consistent expression bias from the same allele in heterozygous individuals (in 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 the identification of imprinted genes.

We measured ASE in 1,687 RNA sequencing (RNA-seq) samples from 45 tissues in 178 individuals (Gene-Tissue Expression Project

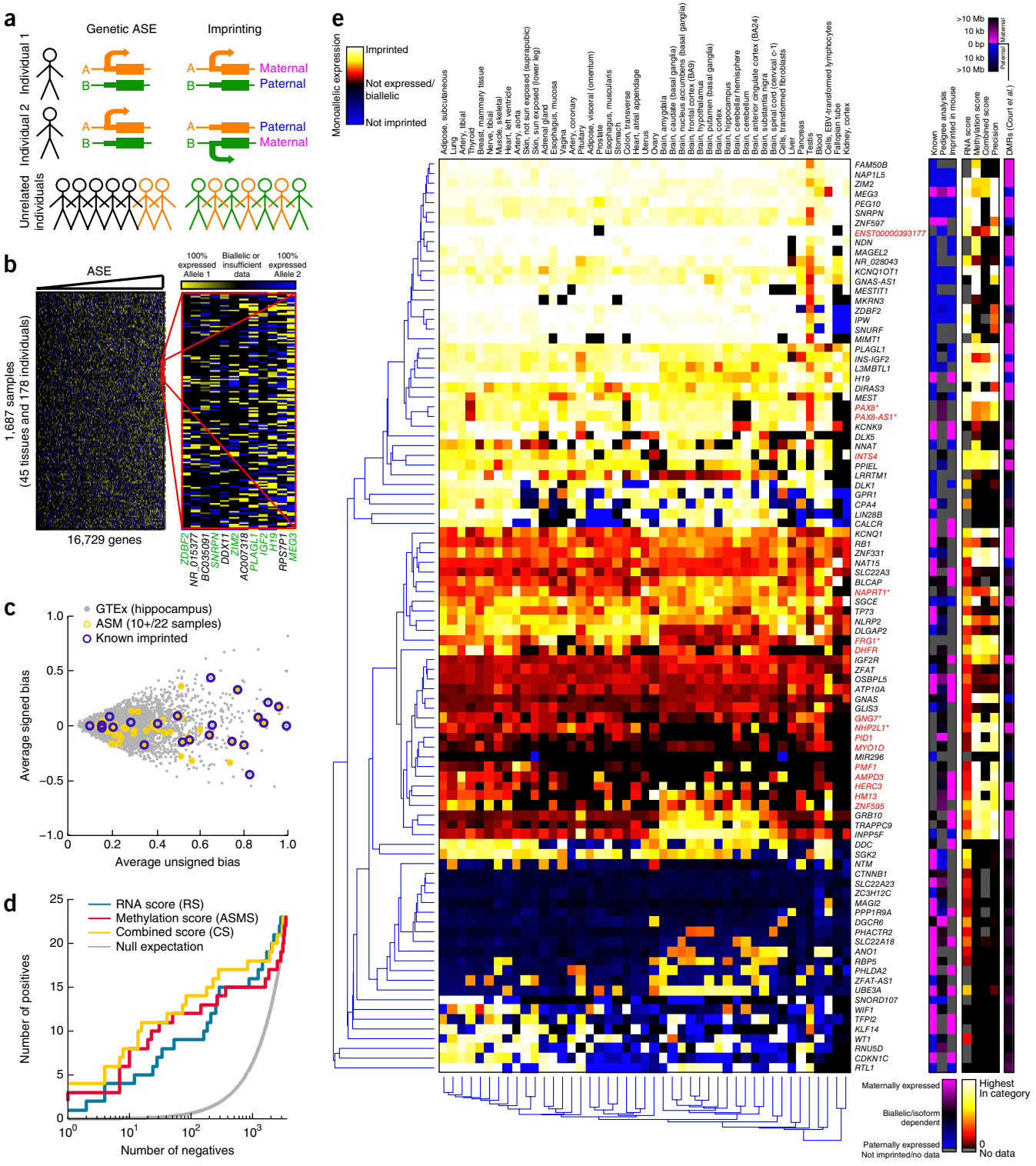

Figure 2 Atlas of genomic imprinting in human. (**a**) ASE caused by genetic polymorphisms will tend to be biased toward the same allele (orange) in heterozygotes, which are at most 50% of individuals. In contrast, ASE due to imprinting will be present in all individuals but will not favor either allele. (**b**) ASE for all genes powered (see Online Methods) in all available GTEx.v3 samples. ASE is scaled from −1 (100% expression of one allele) to +1 (100% expression of the other allele), and sorting genes on the basis of Σ|ASE| yielded 7 known imprinted genes (green text) among the top 12. (**c**) Resolving mean(|ASE|) further by plotting against mean(ASE) showed the tendency for imprinted genes to switch bias between alleles ($y \approx 0$) and for the strongest imprinted genes to have ASM[22] (shown are genes identified in 10 or more of 22 biological samples). (**d**) Number of known imprinted genes (positives), detected among all genes sorted using various scoring schemes. Combining the methylation score (ASMS) and RNA score (RS; into a combined score, CS) improves overall performance. The hippocampus is shown in **c** and **d**; other tissues behaved similarly. (**e**) Monoallelic expression of imprinted genes in 45 human tissues. Genes with CS < 0 (no evidence for imprinting) are colored blue. Clustering was performed using Manhattan distance. Right, pedigree analysis shown as the average parent-of-origin bias for 2 parents and 11 children. Precision is the proportion of positive calls that are known to correspond to imprinted genes given the 45 tissue-specific CS values used to establish a threshold (Online Methods). An asterisk indicates genes validated by mmPCR sequencing. Black text, known; red text, GeneImprint or new.

(GTEx.v3)[21]); **Supplementary Table 6**). We previously showed that concordance in ASE between two samples measured at the same SNP underestimates error as a result of systematic biases[11]. Therefore, we calibrated our parameters on the 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 for genotype imputation, phasing and quantification of ASE. We found that, similarly to in mouse, imprinted genes were highly over-represented among monoallelically expressed genes (with 7/76 known imprinted genes among the top 12 with the highest significance of ASE; **Fig. 2b**). The 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 for lymphocytes derived from 17 members of a 3-generation family[23]. This pedigree allowed us to identify imprinted genes, as their direction of ASE depended on each allele's parent of origin (Online Methods and **Supplementary Figs. 8** and **9**), and to estimate an FDR

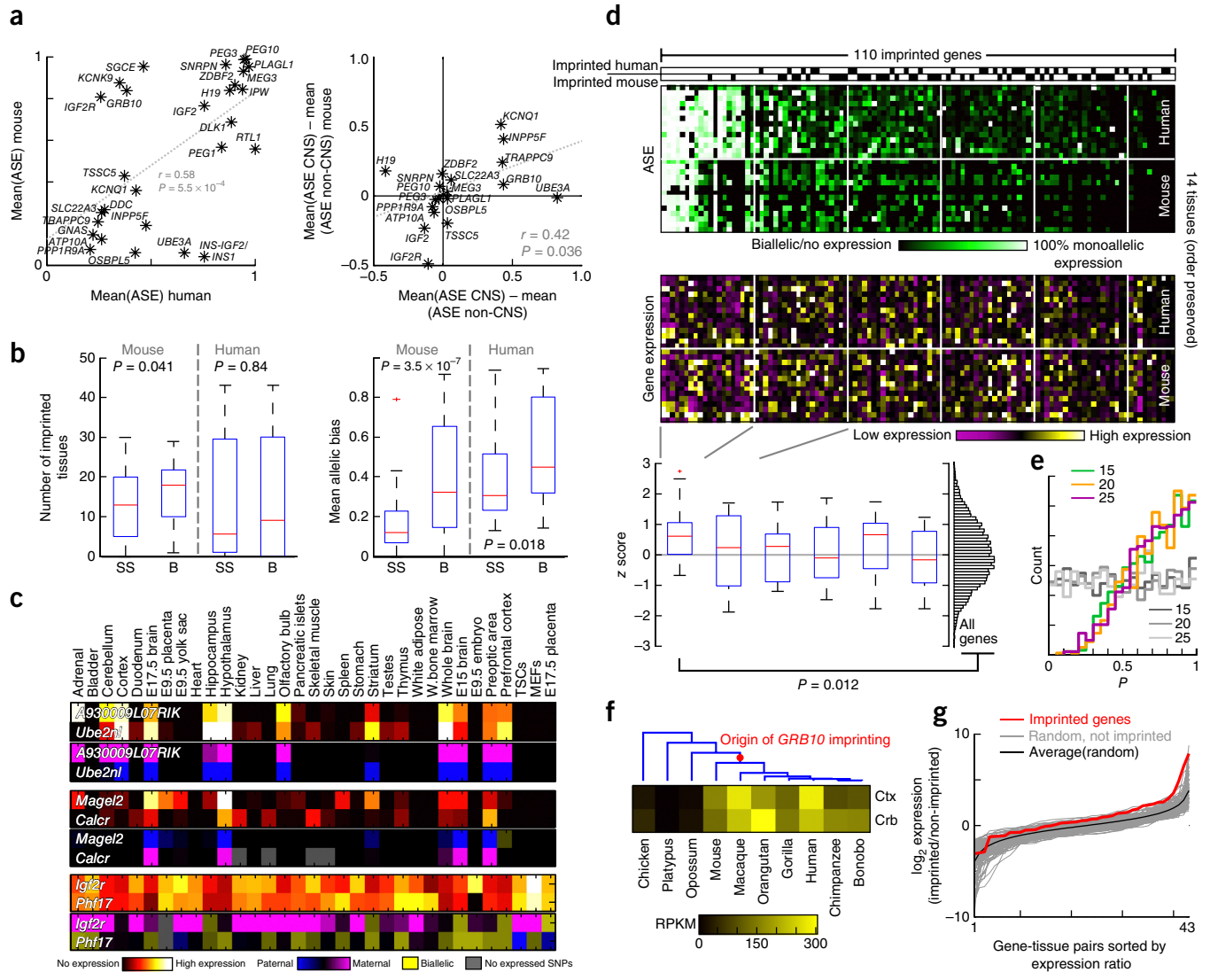

Figure 3 Species comparisons of imprinting. (**a**) Conservation of allelic bias between human and mouse. Left, average allelic bias is generally conserved between the two species. Right, tissue-specific imprinting was strongest between CNS and non-CNS samples and was conserved between the two species. (**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; each box delineates the 25th and 75th quartiles; whiskers show the range of non-outliers) and are imprinted in more tissues in mouse. (**c**) Examples of the strongest maternal-paternal pairings (see also **Supplementary Fig. 16**), showing expression (top) and imprinting (bottom) patterns for each pair in mouse. (**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 Figs. 18** and **19**; box plot metrics are the same as in **b**). *z* scores were computed against the gene expression divergence of randomly selected genes matched for breadth of expression (Online Methods). Expression is shown as asinh(FPKM), median subtracted for each gene within the species. (**e**) The significance of comparisons of within-human expression variation for the top 15, 20 and 25 most strongly imprinted genes (1,000 permutations). Colors correspond to the strongest imprinting bins from **d**; gray scale is used for genes randomized to demonstrate the null distribution. (**f**) Example of a gene with elevated expression in species and tissues where it is imprinted. Ctx, cortex; Cer, cerebellum; RPKM, reads per kilobase per million aligned reads. (**g**) Sorted log ratios of mean expression for mouse and human (imprinted) orthologs versus mean log ratios of expression for platypus and chicken (not imprinted) orthologs in gene-tissue combinations where there was imprinting in both human and mouse. Expression was elevated in mouse and human ($P = 0.0033$; see randomly selected, non-imprinted genes in the same tissues).

for our GTEx scoring scheme (**Fig. 2d**). This represents a general approach that can be applied to discover imprinted genes whenever multigenerational expression and genotype data are available.

As in mouse, we identified most genes known to be imprinted in human (63/76; **Fig. 2e** and **Supplementary Data Set**). The majority of the genes that we missed did not meet our stringent criteria (for example, four expressed heterozygous SNPs). Only a few new imprinted genes in human reached a level of significance comparable to that for 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 significance level corresponding to an FDR of 1% and achieved 100% ASE validation by microfluidics-based multiplex PCR (mmPCR) (**Supplementary Figs. 10–12**, **Supplementary Table 7** and **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 (RNA scores) that pushed them to significance in our study and distinguished them from many other genes with ASM but no evidence of ASE. Similarly to in 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 bias in allelic expression were enriched for functions including development via Hedgehog signaling, kidney development, skeletal system development, regulation of growth and synaptosome localization (**Supplementary Table 8**). Our new imprinted genes regulate glucose import in response to insulin (*PID1*), glucagon signaling and feeding behavior (*GNG7*) and growth (*PMF1*) and are associated with birth weight (*DHFR*) and type 2 diabetes (*MYO1D*) (**Supplementary Table 3**). As in mouse, genes with ubiquitous monoallelic expression (the first 19 genes in **Fig. 2e**) were highly enriched for oxytocin/vasopressin neuropeptide activity and genes governing eating behavior (**Supplementary Fig. 13**).

We identified several properties of genomic imprinting conserved between human and mouse. The dichotomy of imprinting between neural and non-neural adult tissues was shared (**Fig. 2e** and **Supplementary Fig. 3**), suggesting a conserved role for imprinting in neural function. We also found that genes that were imprinted in both species had stronger allelic bias but similar imprinting breadth when compared to species-specific imprinted genes (**Fig. 3**). 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 owing to genetic conflict over nutrient allocation[3]. For the 41 genes imprinted in mouse but not human, we observed an excess of maternally expressed genes (61%). This finding is consistent with theoretical predictions that the silencing of paternally derived alleles should be less evolutionarily stable, owing 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 among species-specific imprinted genes—which was 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 coevolve in evolutionary 'arms races', possibly leading to the coexpression of antagonistic gene pairs[26]. Consistent with this hypothesis, we observed an excess of maternal-paternal pairs among the most strongly coexpressed imprinted genes in mouse (excluding genes in close genomic proximity; Online Methods) (**Supplementary Fig. 15** and **Supplementary Table 9**). Many of the strongest maternal-paternal coexpressed pairs also had reciprocal imprinting patterns (**Fig. 3c** and **Supplementary Fig. 16**) and opposing functions. For instance, *Magel2* and *Calcr* are a coexpressed pair involved in neuropeptide hormonal signaling. Loss of *Magel2* function results in poor suckling and neonatal growth retardation[27], whereas *Calcr* affects appetite suppression through amylin regulation[28]. *Igf2r* and *Phf17* have also been linked to the regulation of growth: *Igf2r* suppresses growth[29], whereas *Phf17* promotes vasculogenesis of the placenta[30], where it is preferentially imprinted (**Fig. 3c**). The human orthologs of the maternal-paternal pairs were coexpressed 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 as a result of positive selection, in response to increases in the expression levels of their antagonistic counterparts. To investigate this possibility, 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 the imprinted genes with the strongest allelic bias (**Fig. 3d**, Online Methods and **Supplementary Figs. 18** and **19**). To test the possibility that strong imprinting itself causes variable expression, we searched for a similar pattern among human individuals but actually found less variation in expression for 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 the divergence.

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

In conclusion, our human and mouse imprinting atlases have shown 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: coexpression of maternally and paternally expressed genes, rapid divergence in expression levels and an overall pattern of upregulation associated with imprinting. We expect that these resources will be instrumental in refining the understanding of imprinting mechanisms at individual loci and that similar atlases in other species will improve understanding of the origins of imprinting.

URLs. GeneImprint, http://www.geneimprint.com/; Novoalign, http://www.novocraft.com/products/novoalign/; MouseBook (Harwell Phenotype Maps), http://www.mousebook.org/mousebook-catalogs/imprinting-resource.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The RNA-seq data have been deposited in the Sequence Read Archive (SRA) and are available under accession SRP020526.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of the Fraser laboratory for critical evaluation of the manuscript. This work was supported by US National Institutes of Health (NIH) grant 1R01GM097171-01A1. H.B.F. is an Alfred P. Sloan Fellow and a Pew Scholar in the Biomedical Sciences. The work of S.B.M. and J.B.L. was supported by US NIH grant U01HG007593. E.K.T. was supported by a Hewlett-Packard Stanford Graduate Fellowship. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by US National Science Foundation grant OCI-1053575. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the US NIH. Additional GTEx funds were provided by the National Cancer Institute, National Human Genome Research Institute, National Heart, Lung, and Blood Institute, National Institute on Drug Abuse, National Institute of Mental Health and National Institute Neurological Disorders and Stroke. Donors were enrolled at Biospecimen Source Sites funded by National Cancer Institute/SAIC-Frederick, Inc. subcontracts to the National Disease Research Interchange (10XS170), the Roswell Park Cancer Institute (10XS171) and Science Care, Inc. (X10S172). The Laboratory, Data Analysis and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to the Broad Institute. Biorepository operations were funded through an SAIC-Frederick subcontract to the Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-Frederick (HHSN261200800001E). The Brain Bank was supported by a supplement to University of Miami grant DA006227. Statistical Methods development grants were made to the University of Geneva (MH090941), the University of Chicago (MH090951 and MH090937), the University of North Carolina, Chapel Hill (MH090936) and Harvard University (MH090948).

AUTHOR CONTRIBUTIONS

B.D. and D.v.d.K. generated the mouse crosses and performed the dissections. T.B., Y.Z. and B.D. performed the mouse RNA-seq experiments. E.K.T., K.S.S., K.R.K., R.Z., J.B.L. and S.B.M. performed the human ASE validation by mmPCR-seq. X.L. and S.B.M. contributed the pedigree data. T.B. performed the analysis. T.B. and H.B.F. wrote the manuscript. All authors contributed to reading and editing the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Barlow, D.P. Gametic imprinting in mammals. *Science* **270**, 1610–1613 (1995).
2. Barlow, D.P. & Bartolomei, M.S. Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* **6**, a018382 (2014).
3. Moore, T. & Haig, D. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.* **7**, 45–49 (1991).
4. Spencer, H.G. & Clark, A.G. Non-conflict theories for the evolution of genomic imprinting. *Heredity* (*Edinb.*) **113**, 112–118 (2014).
5. Wilkins, J.F. & Haig, D. What good is genomic imprinting: the function of parent-specific gene expression. *Nat. Rev. Genet.* **4**, 359–368 (2003).
6. Curley, J.P., Barton, S., Surani, A. & Keverne, E.B. Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. *Proc. Biol. Sci.* **271**, 1303–1309 (2004).
7. Wolf, J.B. & Hager, R. A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol.* **4**, e380 (2006).
8. Babak, T. Identification of imprinted loci by transcriptome sequencing. *Methods Mol. Biol.* **925**, 79–88 (2012).
9. Babak, T. *et al.* Global survey of genomic imprinting by transcriptome sequencing. *Curr. Biol.* **18**, 1735–1741 (2008).
10. Wang, X. *et al.* Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS ONE* **3**, e3839 (2008).
11. DeVeale, B., van der Kooy, D. & Babak, T. Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.* **8**, e1002600 (2012).
12. Gregg, C. *et al.* High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* **329**, 643–648 (2010).
13. Tran, D.A., Bai, A.Y., Singh, P., Wu, X. & Szabo, P.E. Characterization of the imprinting signature of mouse embryo fibroblasts by RNA deep sequencing. *Nucleic Acids Res.* **42**, 1772–1783 (2014).
14. Calabrese, J.M. *et al.* Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell* **151**, 951–963 (2012).
15. Wang, X., Soloway, P.D. & Clark, A.G. A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics* **189**, 109–122 (2011).
16. Schulz, R. *et al.* WAMIDEX: a web atlas of murine genomic imprinting and differential expression. *Epigenetics* **3**, 89–96 (2008).
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* **77**, 91–98 (2001).
18. Morison, I.M., Ramsay, J.P. & Spencer, H.G. A census of mammalian imprinting. *Trends Genet.* **21**, 457–465 (2005).
19. Morison, I.M., Paton, C.J. & Cleverley, S.D. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res.* **29**, 275–276 (2001).
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.* **19**, 4895–4905 (2010).
21. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
22. Fang, F. *et al.* Genomic landscape of human allele-specific DNA methylation. *Proc. Natl. Acad. Sci. USA* **109**, 7332–7337 (2012).
23. Li, X. *et al.* Transcriptome sequencing of a large human family identifies the impact of rare noncoding variants. *Am. J. Hum. Genet.* **95**, 245–256 (2014).
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.* **24**, 554–569 (2014).
25. Wilkins, J.F. & Haig, D. Parental modifiers, antisense transcripts and loss of imprinting. *Proc. Biol. Sci.* **269**, 1841–1846 (2002).
26. Wilkins, J.F. & Haig, D. Genomic imprinting of two antagonistic loci. *Proc. Biol. Sci.* **268**, 1861–1867 (2001).
27. Bischof, J.M., Stewart, C.L. & Wevrick, R. Inactivation of the mouse *Magel2* gene results in growth abnormalities similar to Prader-Willi syndrome. *Hum. Mol. Genet.* **16**, 2713–2719 (2007).
28. Potes, C.S. & Lutz, T.A. Brainstem mechanisms of amylin-induced anorexia. *Physiol. Behav.* **100**, 511–518 (2010).
29. Wutz, A. *et al.* Non-imprinted *Igf2r* expression decreases growth and rescues the *Tme* mutation in mice. *Development* **128**, 1881–1887 (2001).
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.* **23**, 8553–8562 (2003).
31. Pereira, V., Waxman, D. & Eyre-Walker, A. A problem with the correlation coefficient as a measure of gene expression divergence. *Genetics* **183**, 1597–1600 (2009).
32. Renfree, M.B., Hore, T.A., Shaw, G., Graves, J.A. & Pask, A.J. Evolution of genomic imprinting: insights from marsupials and monotremes. *Annu. Rev. Genomics Hum. Genet.* **10**, 241–262 (2009).
33. Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816–831 (2012).

ONLINE METHODS

Identifying and quantifying imprinting in mouse. Tissues were dissected from 16 reciprocally crossed male C57BL/6J × CAST/EiJ F_1 mice (8 mice in each direction of the cross) aged 35–45 d and 2 embryonic stages (**Supplementary Table 1**). Mice were housed and euthanized in accordance with the current Animal Use Protocol approved by the Faculty of Medicine and Pharmacy Animal Care Committee at the University of Toronto. Tissues were rinsed in PBS and snap chilled in liquid nitrogen within 10 min of dissection. RNA was extracted with TRIzol (Invitrogen) according to the manufacturer's recommendations, and integrity was confirmed by BioAnalyzer (RNA integrity number (RIN) > 6). Sequencing libraries were prepared using TruSeq v2 RNA-Seq kits (RS-122-2001, Illumina) according to the manufacturer's recommendations. Fourteen libraries (seven tissues; **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 instrument (PE90) with an average of 7.2 Gb/sample at BGI. Genomic imprinting was quantified using previously established criteria[11]. In brief, reads were aligned with Novoalign (Novocraft) against gene models (**Supplementary Data Set**), 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. Before alignment, we masked 19.6 million C57BL/6 and CAST/EiJ SNPs, representing the union of the 2 collections[35,36], to minimize alignment biases (**Supplementary Fig. 21**). ASE was quantified for each gene by counting the number of uniquely mapped read pairs within the gene boundaries (including introns) that overlapped at least one SNP, such that allelic origin could be discerned. The probability of ASE was estimated from the cumulative binomial distribution with random expectation set to 50% (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 $\log_{10}$ value 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 values to background IS values computed from biological replicates with the same parental background[11] (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 the literature[11,18] and are available as a BED track along with new genes discovered in this study; genes with conflicting evidence in this study and the literature were not included (**Supplementary Data Set**). The seven samples sequenced in previous studies (**Fig. 1**) were analyzed in parallel from the fastq files. Putative new imprinted genes ($P < 1 \times 10^{-2}$; >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 s.d. of 182 biological replicate DNA measurements) and each ratio was reciprocally biased (in the opposite direction) 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 (SRA) under accession SRP020526. Functional enrichments were identified using hypergeometric tests against the JAX Phenome database[37], mouse Gene Ontology[38], human Reactome[39], GeneGO and Biobase pathways mapped to mouse genes via Ensembl orthologs[40]. All databases were downloaded between 10 May and 20 June 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 the significance of the gene set of interest.

Identifying and quantifying imprinting in human tissues. GTEx.v3 (ref. 21) imputed genotype and aligned RNA-seq (BAM) data were downloaded from the database of Genotypes and Phenotypes (dbGaP; phs000424.v3.p1) on 27 August 2013; 1,687 samples had both RNA-seq and genotypes available, comprising 45 unique tissues and organs from 178 individuals (**Supplementary Table 4**). No minimum number of individuals per tissue was required. Genotypes were phased using SHAPEIT.v2 (ref. 41) with genomic maps (human b37) from the authors' website downloaded on 27 August 2013 and all options set to default. Phased ASE read counts were compiled at all heterozygous sites using sra-pileup (v2.3.2 from the 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, and NSCAN and GENSCAN gene predictions, respectively, in a total of six iterations. Known imprinted genes in human were compiled from the literature[16,18] and are available in BED format (**Supplementary Data Set**). Gene-level ASE was quantified by aggregating counts across all phased heterozygous sites within gene boundaries such that each gene was assigned two integers for each sample: a_{ig} and b_{ig}, representing expression from allele *a* and allele *b* in sample *i* for gene *g*. Although assignment of *a* and *b* was arbitrary, phasing ensured that, when possible, assignment of alleles was consistent across individuals, thus enabling assessment of conserved allelic expression bias. Allelic bias (AB_{ig}) was quantified for each gene in each tissue as $2^*(a_{ig}/(a_{ig} + b_{ig}) - 0.5)$ and the RNA score (RS_{ig}) was computed as $\Sigma|AB_{ig}| - |\Sigma AB_{ig}|$ for $i = 1, 2, \ldots, n$ (where *n* was the number of subjects where that tissue was sequenced). More complicated combinations (for example, adding regression-derived weights to the two main terms) did not improve performance as gauged by receiver operating curve/area under the curve (ROC/AUC) analysis of known imprinted genes and assuming all negatives were true negatives (for example, see **Fig. 2d**). An ASM score (ASMS) was assigned to each gene by overlapping the ASM coordinates from 22 samples[22] with the above gene models and tallying the number of samples supporting ASM within the gene boundaries (including UTRs and introns). Cultured and uncultured differentiated cells did not have different distributions of ASM for known imprinted genes (**Supplementary Fig. 22**), and extending gene models to capture potential intergenic regulation did not improve performance (**Supplementary Fig. 23**). Multiple linear regression was used to compute a combined score (CS) by determining the ASMS and RS weights (mean ASMS weight = 0.29, mean RS weight = 0.11) that best distinguished 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 as we observed that it was affected by both the score and the number of tissues exceeding that score (an imprint was more likely to be real if supported by data from multiple tissues). Precision was estimated for each gene by choosing a CS threshold that optimized precision, given the number of tissues that exceeded 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 3-generation pedigree (NA1463; Gene Expression Omnibus (GEO), GSE56961). The two grandparent-parent trios and one parent-child pedigree were phased separately with SHAPEIT.v2 (ref. 41), and heterozygous SNPs within gene boundaries were used to track the inheritance of a child's alleles back to the grandparents (when possible; 67.9% and 68.5% of expressed genes in lymphocytes with at least 1 expressed heterozygote (ASE powered; $n = 17{,}855$) were traceable from each child back to the 2 sets of grandparents, and 94.1% of ASE-powered genes were traceable for at least 1 of 11 children for at least 2 alleles). Gene-level ASE was quantified as described above under "Identifying and quantifying imprinting in human tissues." Imprinting was considered to have been validated when all of the following criteria were met: (i) at least one allele switched from preferential expression to preferential silencing, consistent with its parent of origin (for example, the allele was expressed in the father when inherited from his mother but then silent in the father's children), (ii) parent-of-origin bias was consistent in all children and parents, and (iii) at least three children were powered for the detection of ASE. A minimum of ten ASE-informative reads were required, and bias was considered to be present when a binomial test (cumulative distribution function, as described previously[11]) yielded a significance of $P < 0.01$ when 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 of 14.2 tissues/subject; 745 total measurements). Tissue samples were contributed by the GTEx

doi:10.1038/ng.3274

Consortium[21] and processed as described previously[42] 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, to genotyped variants when exome SNPs were not detected and to imputed SNPs if no other evidence was available. Allelic ratios were quantified as described previously[42] (see **Supplementary Table 2** for design details).

Human-mouse comparisons. Mean allelic bias was quantified for each gene as the sum of the 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). Each tissue contributed a value between 0 (100% biallelic expression) and 1 (100% monoallelic expression).

Enrichment for maternal-paternal coexpressed 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 the gene) and was quantile normalized in each species separately. These data were used to compute the distances between all unique pairwise combinations of genes. The distance similarity metric (dist.sim) used for all comparisons was standard Euclidean distance. In comparisons of the number 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 (for example, 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). The significance of deviation from the null expectation 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 and mouse orthologs[40] 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 because the mouse and human counterparts did not cluster next to each other, possibly as a result of the human dissection including muscular tissue (**Supplementary Fig. 24**), leaving 14 pairs for analysis. Data were median subtracted (for each gene; the median across all tissues was 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 could contribute a value ranging from 0 (no ASE) to 1 (100% monoallelic expression). Divergence was measured using the Euclidean distance between each set of orthologs and was quantified as a *z* score (number of standard deviations from the mean) relative to a background set matched for the degree of expression. A gene was included in the background if it was expressed in the same number of tissues (nonzero expression after median subtraction) with the highest FPKM value being within 10% of the highest FPKM value of the imprinted gene; the background was further trimmed such that the same number of genes had expression above and below the imprinted maximal expression. There was no association between background scores and imprinting breadth. Imprinting breadth was the sum of allelic bias across all tissues. The sum of the Euclidean distances between matched tissue pairs was also used to quantify divergence, where the inputs were groups of genes and matched background sets that comprised randomly substituted genes from all orthologs. Human-human comparisons (**Fig. 3e**) were run on mean expression across 2 random subsets of individuals for the top 15, 20 and 25 genes sorted on strength of ASE. For comparison, random sets of genes were also analyzed in the same way.

Comparison of expression for imprinted genes to that of non-imprinted orthologs. Normalized gene expression data from six tissues for ten species[43] were 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 $\log_2$ 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 analysis was repeated 10,000 times on randomly selected non-imprinted genes using expression data from the same tissues.

34. Parkhomchuk, D. *et al.* Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res.* **37**, e123 (2009).
35. Frazer, K.A. *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050–1053 (2007).
36. Keane, T.M. *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289–294 (2011).
37. Grubb, S.C., Bult, C.J. & Bogue, M.A. Mouse phenome database. *Nucleic Acids Res.* **42**, D825–D834 (2014).
38. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
39. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
40. Guberman, J.M. *et al.* BioMart Central Portal: an open database network for the biological community. *Database* (*Oxford*) **2011**, bar041 (2011).
41. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2013).
42. Zhang, R. *et al.* Quantifying RNA allelic ratios by microfluidic multiplex PCR and sequencing. *Nat. Methods* **11**, 51–54 (2014).
43. Brawand, D. *et al.* The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).