Genetic Differentiation of Hypothalamus Parentally Biased Transcripts in Populations of the House Mouse Implicate the Prader–Willi Syndrome Imprinted Region as a Possible Source of Behavioral Divergence

Anna Lorenc,[†,1] Miriam Linenbrink,[1] Inka Montero,[‡,1] Markus B. Schilhabel,[2] and Diethard Tautz*[1]

[1]Max-Planck Institute for Evolutionary Biology, Department Evolutionary Genetics, Plön, Germany

[2]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany

[†]Present address: National Institute for Health Research Biomedical Research Centre at Guy's and St Thomas' Hospital Foundation Trust and King's College, London, United Kingdom

[‡]Present address: Medical Faculty, Eberhard Karls Universität Tübingen, Tübingen, Germany

*Corresponding author: E-mail: tautz@evolbio.mpg.de.

Associate editor: Michael Nachman

Abstract

Parentally biased expression of transcripts (genomic imprinting) in adult tissues, including the brain, can influence and possibly drive the evolution of behavioral traits. We have previously found that paternally determined cues are involved in population-specific mate choice decisions between two populations of the Western house mouse (*Mus musculus domesticus*). Here, we ask whether this could be mediated by genomically imprinted transcripts that are subject to fast differentiation between these populations. We focus on three organs that are of special relevance for mate choice and behavior: The vomeronasal organ (VNO), the hypothalamus, and the liver. To first identify candidate transcripts at a genome-wide scale, we used reciprocal crosses between *M. m. domesticus* and *M. m. musculus* inbred strains and RNA sequencing of the respective tissues. Using a false discovery cutoff derived from mock reciprocal cross comparisons, we find a total of 66 imprinted transcripts, 13 of which have previously not been described as imprinted. The largest number of imprinted transcripts were found in the hypothalamus; fewer were found in the VNO, and the least were found in the liver. To assess molecular differentiation and imprinting in the wild-derived *M. m. domesticus* populations, we sequenced the RNA of the hypothalamus from individuals of these populations. This confirmed the presence of the above identified transcripts also in wild populations and allowed us to search for those that show a high genetic differentiation between these populations. Our results identify the *Ube3a–Snrpn* imprinted region on chromosome 7 as a region that encompasses the largest number of previously not described transcripts with paternal expression bias, several of which are at the same time highly differentiated. For four of these, we confirmed their imprinting status via single nucleotide polymorphism-specific pyrosequencing assays with RNA from reciprocal crosses. In addition, we find the paternally expressed *Peg13* transcript within the *Trappc9* gene region on chromosome 15 to be highly differentiated. Interestingly, both regions have been implicated in Prader–Willi nervous system disorder phenotypes in humans. We suggest that these genomically imprinted regions are candidates for influencing the population-specific mate-choice in mice.

Key words: imprinting, genetic differentiation, RNASeq, *Mus musculus*, *Mus domesticus*, hypothalamus, vomeronasal organ, liver.

Introduction

Parent-of-origin effects on gene function and expression (genomic imprinting) have been implicated in many developmental processes and genetic diseases (Wood and Oakey 2006). Imprinted genes are autosomal loci expressed from only one parental allele due to epigenetically inherited differential methylation marks. However, exclusion of the alternative parental allele does not need to be complete, that is, the expression of the two parental alleles to different degrees, depending on their parent-of-origin, is also known as genomic imprinting (Wolf et al. 2008; Gregg, Zhang, Weissbourd, et al. 2010; DeVeale et al. 2012). Most studies on genomic imprinting have focused on effects in early development, but independent of the early embryonic effects genomic imprinting can also influence behavioral patterns controlled by the adult brain, such as maternal care or social dominance (Wilkins and Haig 2003; Wolf and Hager 2009; Curley 2011; Garfield et al. 2011).

We have recently found that mate choice decisions in mice appear to be influenced by paternally inherited cues; these cues appear to be subject to fast evolution and thus contribute to within-species population divergence (Montero et al. 2013). In the respective study, we conducted an experiment that allowed the free choice of mates in a seminatural

Open Access

environment. We used two wild-caught populations of *Mus musculus domesticus* from France and Germany, which diverged approximately 3,000 years ago. Mating success was assessed through molecular paternity analysis. F1 offspring showed a general assortative pattern according to their population of origin. Matings between a hybrid and an animal of pure origin showed a strong preference for matching the paternal side of the hybrid. The study therefore suggests that a combination of learned and inherited cues must exist to allow such differential decisions. Moreover, the inherited cues must have diverged between the two populations to allow the population-specific decisions, that is, they are expected to belong to the molecularly highly differentiated fraction of their genomes. Given that the F1 hybrids in the study shared the same autosomal genome combinations yet displayed differential choice according to their paternal population origin, it seems possible that genomically imprinted loci mediated the respective genetic component. Here, we search for such candidate loci by identifying imprinted transcripts that are highly diverged and thus could be involved in the observed behavioral pattern.

Imprinted loci can be detected in reciprocal crosses between animals that differ in diagnostic single nucleotide polymorphisms (SNPs) using RNA sequencing (RNAseq) (Wang and Clark 2014). However, this approach has yielded conflicting results, because the statistical evaluation of the data can be problematic. Based on RNAseq data and binomial sampling statistics, Gregg, Zhang, Weissbourd, et al. (2010) suggested that a large number of brain-expressed genes is affected by imprinting (Gregg, Zhang, Butler, et al. 2010). However, DeVeale et al. (2012) showed that this approach leads to an overestimate of imprinted genes and they suggest a statistical procedure that reduces biases in RNASeq data by estimating a false discovery rate (FDR) from the data through mock comparisons between samples of the same cross direction. This leads to much fewer identified loci, but independent confirmation assays showed that this is the more reliable way to find imprinted transcripts.

In our study, we focus on genes expressed in the hypothalamus (HYP), the vomeronasal organ (VNO), and the liver (LIV), because these are of particular interest for behavior and mate recognition. The HYP controls sexual behavior and maternal care and is responsive to olfactory stimuli, such as pheromones. The VNO is the olfactory sense organ in mice that detects pheromones. It has been implicated in detecting major histocompatibility complex (MHC) peptides that are thought to be relevant for optimal mate choice (Leinders-Zufall et al. 2009). The LIV produces the major urinary proteins (MUPs) that are excreted with the urine and play a general role in chemical communication among mice, including signaling of reproductive status (Roberts et al. 2010; Janotova and Stopka 2011; Nelson et al. 2013; Stockley et al. 2013).

We first used RNAseq data from reciprocal crosses of inbred strains from two *Mus* subspecies: *M. m. domesticus* and *M. m. musculus*. We apply the recommendations for deriving a false discovery rate (FDR) from the data proposed by DeVeale et al. (2012) and confirm their findings that a stringent statistical treatment is advantageous to obtain reliable data. We find many previously studied imprinted transcripts but also a number of new ones, most of which may be noncoding long RNAs and are of yet unknown function. A large number of novel imprinted transcripts is found in a region of chromosome 7 that is orthologous to the region involved in the Prader–Willi/Angelman syndromes in humans. These are complex disorders with parent-of-origin effects, including behavioral and cognitive deficiencies, that are driven by deletions or disruptions of imprinted gene clusters on human chromosome 15 (Horsthemke and Wagstaff 2008; Cassidy et al. 2012).

After having identified the imprinted transcripts in the cross of the inbred *Mus* subspecies, we also sequenced HYP RNA from eight individuals from each of the two *M. m. domesticus* populations of our original mate choice study (see above). Interestingly, we find that the transcripts in the Prader–Willi syndrome (PWS) region show highly differentiated alleles between the populations. Further, we find a second highly differentiated transcript, *Peg13* on chromosome 15, that has also been implicated in Prader–Willi-like phenotypes in humans. We propose that these candidate genes could contribute to behavioral divergence patterns of natural mouse populations.

Results

Tissues for RNAseq analysis were derived from reciprocal crosses of animals from two inbred strains, WSB and PWD, that represent the two subspecies *M. m. domesticus* and *M. m. musculus*, respectively. RNA was extracted from the LIV, HYP, and the VNO. Between 16 and 53 million uniquely mapping paired reads were obtained for each cross direction in each tissue from several individuals (supplementary table S1, Supplementary Material online).

SNPs were initially considered to be imprinted when they showed the same direction of change in the reciprocal crosses (i.e., same parent of origin effect) with an initial *P* value cutoff of 0.05 (χ^2 test). For example, overrepresentation of the WSB allele in the PWDxWSB cross and overrepresentation of the PWD allele in the WSBxPWD cross would be called paternally biased expression. However, for most of these "significant" SNPs the actual differences between maternal and paternal read counts were rather small. Because DeVeale et al. (2012) show that the application of such a simple cutoff leads to many false positives, we have followed their approach of a mock comparison to estimate a *P* value that would reduce the empirical FDR to less than 0.05 (fig. 1). The proposed mock comparison is a simple negative control that accounts for systematic error, technical variation, and biological variation. Basically, it asks how many SNPs exceed significance in a mock reciprocal cross (i.e., comparing samples with the same parental background as though they were from reciprocal crosses). In such a comparison, any reciprocally biased expression cannot be caused by genomic imprinting and is a measure of the technical and biological variation of the experimental approach (DeVeale et al. 2012).

Based on this analysis, we use cutoff *P* values of $P < 10^{-2}$ for HYP, 10^{-6} for VNO, and 10^{-5} for LIV. In addition, for those
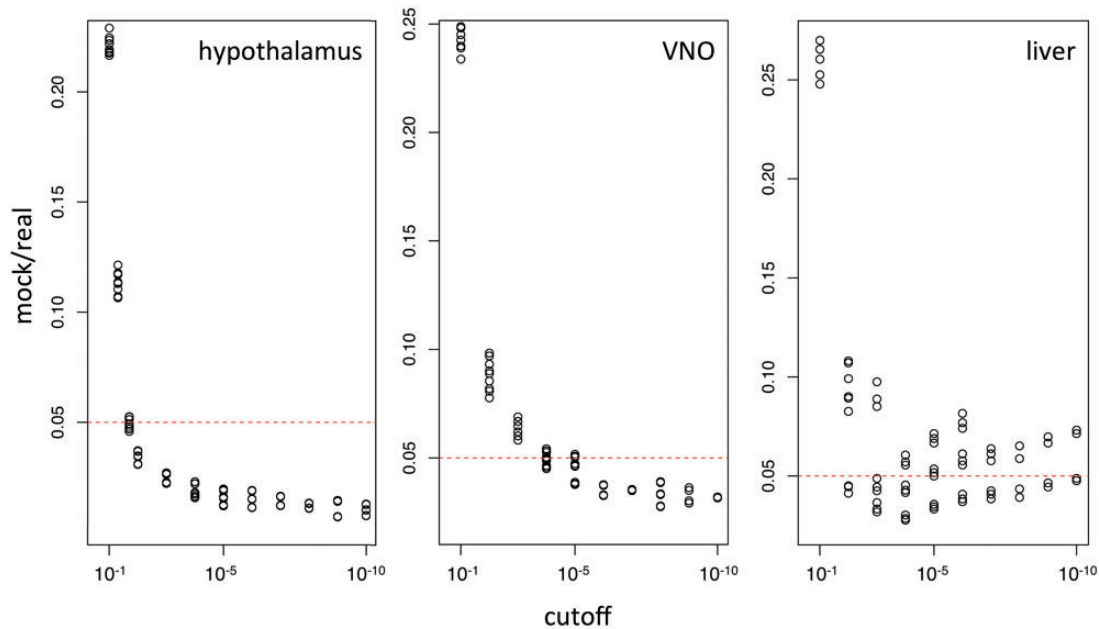
FIG. 1. Determination of FDR cutoff for the three tissues. To find a chi-square test cutoff corresponding to an FDR of 5%, reads from the same cross direction were split randomly into two groups and analyzed as if they originated from reciprocal crosses. Then the ratio of SNPs from such mock comparisons to the number of SNPs identified in reciprocal crosses was calculated. Values are plotted as a function of different P value cutoffs. The acceptable level of false positives was set to 5% (dashed lines), and corresponding cutoff values were applied for the data analysis. The relative variability in the LIV data is due to the small number of actually imprinted transcripts.

transcripts that were not already known from other studies we required that at least two SNPs within the annotated transcript show the same bias. These criteria were used to obtain the list of significantly imprinted transcripts for the three tissues (table 1). The largest number was found in HYP, and the smallest was found in LIV; these results support the notion that the brain is subject to more imprinting than other tissues.

We also found a large number of apparently imprinted transcripts on the X-chromosome, but these are not included in table 1 because the expression bias on the X-chromosome could also be due to unequal ratios of X-chromosome inactivation (Wang et al. 2010; Wang and Clark 2014). The respective SNPs are included in the overall results table (supplementary table S2, Supplementary Material online), but we do not discuss them further.

We find 64 (26 maternal/38 paternal) transcripts preferentially expressed from one parental allele in the HYP, 20 (6/14) in the VNO, and 8 (4/4) in the LIV. They fall into 13 chromosomal regions with two or more imprinted transcripts and seven regions containing only a single imprinted transcript (table 1). Many transcripts are expressed in two or all three tissues but may be imprinted in only one of them (table 1). No transcripts were found that are exclusively imprinted in the LIV, and only two were found that are exclusively imprinted in the VNO. A total of 45 transcripts with parentally biased expression (20 maternal/25 paternal) were found only in the HYP. Approximately two-thirds of the transcripts detected as significantly imprinted in our study are well-known imprinted genes (table 1). In addition, 9 of the 24 novel transcripts were also found in the DeVeale et al. (2012) study, and

two of them were confirmed by pyrosequencing by these authors (table 1).

Three genes showed opposite imprinting in different tissues. *Grb10* (chr11) is paternally expressed in HYP and maternally in VNO. This asymmetry was noted before where *Grb10* is expressed from the paternal allele in the whole brain and from the maternal one in other tissues (Arnaud et al. 2003). It is interesting to note that the sensory epithelia of the VNO behave like the other tissues in this respect. The second locus that shows opposite imprinting is *Igf2* (chr7); it shows a slight but significant maternal bias in HYP and a stronger paternal bias in VNO. *Igf2* has so far mostly been known to be only paternally expressed during development and in various tissues (Chao and D'Amore 2008), but Gregg, Zhang, Weissbourd, et al. (2010) found this paternal and maternal expression in their data as well. Finally, the minor histocompatibility gene *H13* (chr2) has two transcript variants with opposite patterns of imprinting in our data. This confirms the results by Wood et al. (2008) that differential utilization of polyadenylation sites can be epigenetically regulated.

Most of the detected genes occurred in clusters (table 1) in agreement with what is generally known for imprinted genes (Wood and Oakey 2006). In fact, most of the imprinted transcripts found in our survey that were not previously described are part of known clusters. The two exceptions are ENSMUST00000061241 (chr13) and ENSMUST00000168236 (chr17) that showed slight maternal biases (table 1). The largest number of new imprinted transcripts was found in the region between *Ube3a* and *Magel2* on chromosome 7, which represents the region that is involved in the

Table 1. List of Imprinted Transcripts Identified in the Present Study.

Region in mm10	Gene Name	LIV	HYP	VNO	Status	SNPs $F_{st} > 0.8$
chr1:63,273,269-63,314,575	Zdbf2	—	pat > 0.95	—	Known	—
chr1:63,445,904-63,596,515	Adam23	+	pat > 0.55	+	Known[a,b]	—
chr2:152,669,461-152,708,668	H13 (short)	—	pat > 0.8	pat > 0.95	Known	—
chr2:152,669,461-152,708,668	H13 (long)	mat > 0.7	mat > 0.8	mat > 0.6	Known	—
chr2:152,780,668-152,831,682	Bcl2l1	+	pat > 0.55	+	Known[a]	—
chr2:157,556,362-157,566,361	Blcap	+	mat > 0.62	+	Known	—
chr2:157,560,110-157,562,519	Nnat	—	pat > 0.99	pat > 0.75	Known	—
chr6:4,674,350-4,747,204	Sgce	—	pat > 0.95	pat > 0.95	Known	—
chr6:4,747,306-4,760,516	Peg10	—	pat > 0.95	—	Known	—
chr6:4,903,320-5,165,661	Ppp1r9a	—	mat > 0.55	+	Known	—
chr6:5,383,386-5,433,021	Asb4	—	mat > 0.8	—	Known	—
chr6:30,401,909-30,455,174	Klhdc10	+	mat > 0.55	+	Known[a]	—
chr6:30,738,050-30,748,466	Mest	—	pat > 0.95	pat > 0.95	Known	—
chr6:30,804,784-30,807,552	Copg2os2	—	pat > 0.95	—	Known	—
chr6:30,809,559-30,896,760	Copg2	+	mat > 0.7	+	Known	—
chr6:58,833,700-58,920,396	Herc3	—	mat > 0.65	+	Known	—
chr6:58,905,233-58,907,126	Nap1l5	—	pat > 0.99	pat > 0.95	Known	—
chr7:6,671,269-6,672,888	AK003710	—	mat > 0.9	—	New	—
chr7:6,675,443-6,696,432	Zim1	—	mat > 0.9	—	Known	—
chr7:6,703,901-6,730,554	Peg3	pat > 0.95	pat > 0.99	pat > 0.99	Known	—
chr7:6,730,741-6,967,220	Usp29	—	pat > 0.99	—	Known	—
chr7:59,228,750-59,306,727	Ube3a	+	mat > 0.8	+	Known	—
chr7:59,262,923-59,263,934	AK038761	—	pat > 0.99	—	New[c]	—
chr7:59,281,852-59,290,247	A230073K19Rik	—	pat > 0.99	—	New[c]	1
chr7:59,307,924-59,324,149	C230091D08Rik	+	mat > 0.8	+	Known[a]	—
chr7:59,327,318-59,328,016	AK020709	—	pat > 0.95	—	New	—
chr7:59,937,467-59,975,759	D7Ertd715e	—	pat > 0.99	—	New[c]	4
chr7:59,976,740-59,980,676	AK139082	—	pat > 0.99	—	New[c]	1
chr7:59,982,501-60,140,219	Snrpn / Snurf	pat > 0.99	pat > 0.99	pat > 0.99	Known	1
chr7:61,010,256-61,012,230	AK046019	—	pat > 0.99	—	Known	—
chr7:61,072,752-61,089,737	AK038418	—	pat > 0.95	—	New	—
chr7:61,089,568-61,221,965	DOKist4	—	pat > 0.99	—	Known[a]	—
chr7:61,529,410-61,615,327	B230209E15Rik	—	pat > 0.95	—	New	—
chr7:61,705,850-61,927,574	A230057D06Rik	—	pat > 0.95	—	New	—
chr7:61,751,446-61,753,692	AK031915/AK046509	—	pat > 0.99	+	New	—
chr7:61,930,789-61,982,715	ENSMUST00000181804	—	pat > 0.99	—	New	—
chr7:61,930,944-61,934,821	AK048029	—	pat > 0.99	—	New	—
chr7:62,348,277-62,349,927	Ndn	—	pat > 0.99	pat > 0.95	Known	—
chr7:62,376,979-62,381,640	Magel2	—	pat > 0.99	—	Known	—
chr7:128,439,777-128,461,513	Tial1	+	mat > 0.65	+	New	—
chr7:128,611,365-128,696,436	Inpp5f	+	pat > 0.8	+	Known	—
chr7:142,575,532-142,578,146	H19	—	mat > 0.99	—	Known	—
chr7:142,650,768-142,658,804	Igf2	—	mat > 0.55	pat > 0.8	Known	—
chr7:143,107,254-143,427,042	Kcnq1	—	—	pat > 0.99	Known	—
chr7:143,458,339-143,461,050	Cdkn1c	—	—	mat > 0.95	Known	—
chr9:89,909,775-90,026,979	Rasgrf1	—	pat > 0.95	—	Known	—
chr10:13,090,788-13,131,695	Plagl1	—	pat > 0.99	pat > 0.99	Known	—
chr11:11,814,101-11,890,408	Ddc	+	mat > 0.55	—	Known	—
chr11:11,930,499-12,037,420	Grb10	—	pat > 0.95	mat > 0.95	Known	—
chr11:22,972,005-22,976,496	Zrsr1	pat > 0.95	pat > 0.95	pat > 0.95	Known	—
chr11:22,899,728-22,982,284	Commd1/Murr	+	mat > 0.85	+	Known	—
chr12:108,860,030-108,893,211	Wars	+	pat > 0.55	+	Known[a,b]	—
chr12:109,032,182-109,068,217	Begain	—	pat > 0.75	—	Known	—

Table 1. Continued

Region in mm10	Gene Name	LIV	HYP	VNO	Status	SNPs F_{st} > 0.8
chr12:109,453,455-109,463,336	Dlk1	—	pat > 0.95	—	Known	—
chr12:109,542,023-109,568,594	Meg3	—	mat > 0.95	mat > 0.95	Known	—
chr12:109,589,193-109,600,330	Rtl1	—	mat > 0.8	—	Known	—
chr12:109,603,945-109,661,711	Rian	—	mat > 0.99	mat > 0.99	Known	—
chr12:109,734,825-109,749,457	Mirg	—	mat > 0.99	—	Known	—
chr13:107,413,865-107,414,767	ENSMUST00000061241	mat > 0.65	mat > 0.65	+	New	
chr15:72,506,991-72,508,007	AK039650	—	mat > 0.95	—	Known[a]	—
chr15:72,589,620-73,061,204	Trappc9	+	mat > 0.7	+	Known	—
chr15:72,805,600-72,810,324	Peg13	pat > 0.99	pat > 0.99	pat > 0.99	Known	5
chr15:73,098,490-73,099,318	DQ715667	—	mat > 0.7	+	Known[a]	—
chr15:73,101,625-73,184,947	Eif2c2	+	mat > 0.7	+	Known[a]	—
chr17:77,674,376-77,674,702	ENSMUST00000168236	mat > 0.55	mat > 0.55	mat > 0.6	New	—
chr18:12,972,252-12,992,948	Impact	pat > 0.8	pat > 0.95	pat > 0.75	Known	—

NOTE.—Expression status is designed as "+" when expressed but not imprinted in a given tissue, as "−" when not expressed, and as the degree of maternal (mat) or paternal (pat) bias when imprinted. Chromosomal regions with clustered transcripts are separated by horizontal lines.
[a]First described by DeVeale et al. (2012).
[b]Confirmed by pyrosequencing in DeVeale et al. (2012).
[c]Confirmed by pyrosequencing in the present study.

Prader–Willi/Angelman syndromes in humans. Several of these are nested within other transcription units, such as AK038761 and A230073K19Rik within Ube3a. In fact, there appear to be even further imprinted transcripts in the region, because we found SNPs that do not correspond to the annotated named transcripts on which we have focused the analysis. However, we refrain from discussing these, because it is currently difficult to assess whether these might simply correspond to splicing variants of the annotated genes.

Differentiation between Populations

The populations used in our previous study of mating preferences (Montero et al. 2013) are both from the subspecies M. m. domesticus that have been separated since no more than 3,000 years. Although these populations are genetically well differentiated (Ihle et al. 2006; Teschke et al. 2008; Staubach et al. 2012), they harbor only few differentially fixed SNPs (Staubach et al. 2012) that could be used for a systematic assessment of the imprinting status. However, as we are specifically interested in loci that show rapid divergence between the two populations, we have sequenced total HYP RNA from eight animals from each of the two populations to confirm the presence of the newly discovered transcripts in these wild populations. In addition, we assessed which of the transcripts show a particularly high genetic differentiation, that is, which may contribute to the observed differential mate recognition between the populations.

We aligned the reads to the mouse transcriptome and confirmed that all transcripts recovered from the inbred strains were also present in the wild-derived mice. We also assessed whether any of them is differentially expressed between the populations, based on read count coverage, but we did not find significant differences (not shown). We then calculated F_{st} for all transcripts and checked whether any of the imprinted transcripts had a high F_{st}, that is, shows strong genetic differentiation. The average F_{st} across all SNPs in the

transcripts was 0.28, and 2.5% of the SNPs had an F_{st} > 0.8. An F_{st} > 0.8 implies a particularly high differentiation between the populations, that is, this is the pattern we were looking for. Using this value as a cutoff for highly differentiated SNPs, we checked all imprinted transcripts for whether they harbor SNPs with an F_{st} above this cutoff. We find five such transcripts, which are all paternally expressed (table 1).

Two of the transcripts, D7Ertd715e on chromosome 7 and Peg13 on chromosome 15, show several of such highly differentiated SNPs with F_{st} > 0.8 (table 1). In addition, three transcripts in the vicinity of D7Ertd715e, namely A230073 K19Rik, AK139082, and Snrpn/Snurf, show single SNPs with F_{st} > 0.8. In fact, two other transcripts in the region, Ube3a and AK03876, harbor one SNP each with F_{st} > 0.7; this suggests that the whole region between Ube3a and Snrpn/Snurf is highly differentiated between the two mouse populations.

Given the high differentiation of these genes between the two populations, it was possible to design pyrosequencing assays for given SNPs to allow the verification of their imprinting status in individuals of both populations and their reciprocal crosses. For this we extracted HYP RNA of three biological replicates per cross. First, we confirmed the pyrosequencing approach for testing WSB/PWD reciprocal crosses for all four transcripts (fig. 2) and assessed the differential expression status in pure individuals of both wild-caught populations. Finally, we confirmed for each of the four tested transcripts in reciprocal crosses of both natural populations that they show the same expression bias as found in the WSBxPWD crosses (fig. 2).

Discussion

There is a growing realization that parent-of-origin effects play a significant role in the development and behavior of mammals (Wolf and Hager 2009; Curley 2011; Garfield et al. 2011). This corresponds to an increasing number of identified genes that show a parent-of-origin expression bias, that is, imprinting. Using next-generation sequencing approaches, it has
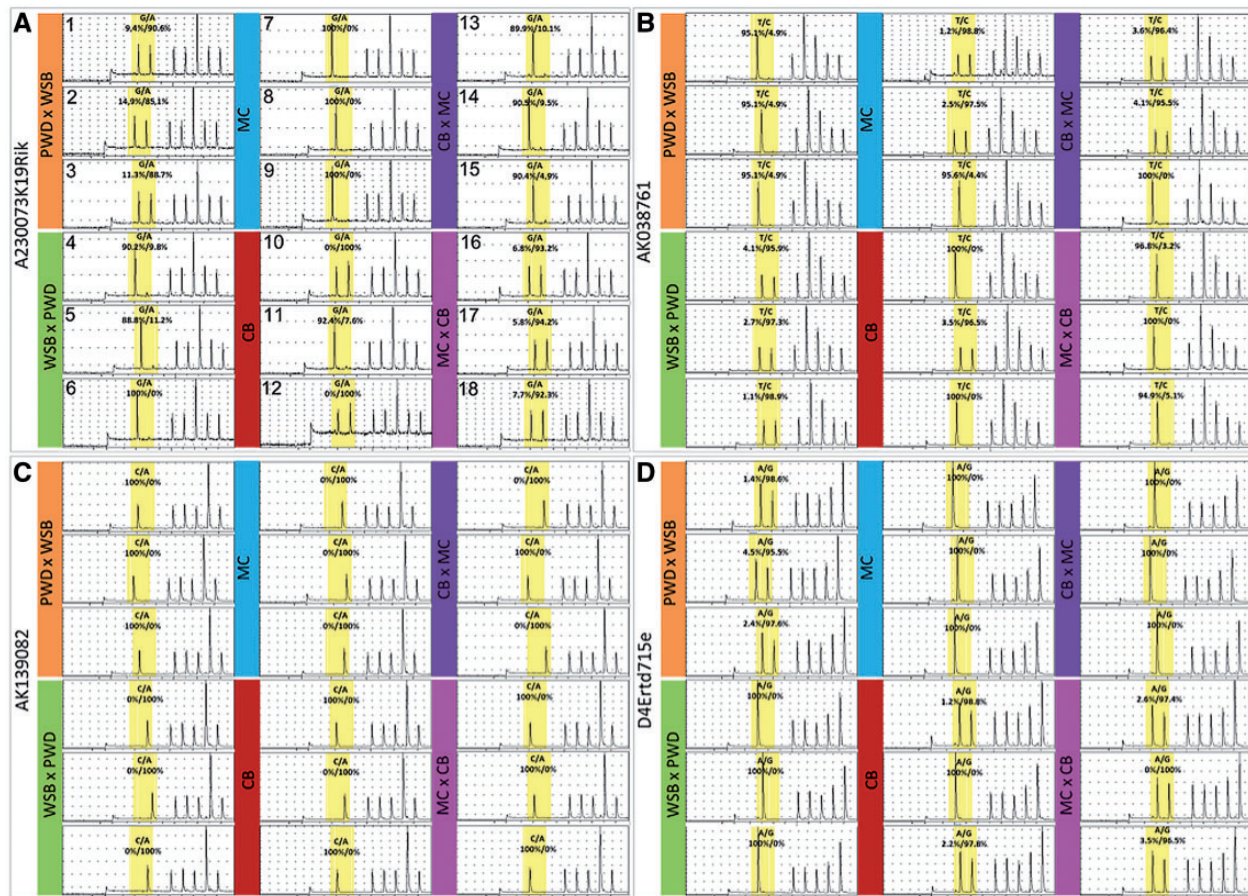
Fig. 2. Pyrograms confirming the imprinting status at four loci in six groups of mice. The loci tested were (*A*) A230073K19Rik, (*B*/) AK038761, (*C*) AK139082, and (*D*) D7Ertd715e. Reciprocal crosses of WSB and PWD inbred strains as well as pure outbred French (MC), pure outbred German (CB) individuals, and reciprocal crosses of these strains were used with three samples per group (female is always listed first in the cross designations). The percent contribution of the one or other allele is listed in each field. Note that not all SNPs are completely diagnostic, but the data are consistent for the respective crosses. In cases where two peaks show the same height but the percent contribution confirms the skew of allelic expression, this is due to the amplicon sequence given for pyro-sequencing. If a SNP is preceded by a nucleotide being the same as one of the two SNP states, three scenarios can occur: 1) equal expression results in the one peak being twice as high as the other with normal height (not present in our data set), 2) unique expression of the allele preceded by the same nucleotide results in a peak twice as high as a normal peak and no peak for the second allele (e.g., individual 7), and 3) if just the second allele is expressed, this results in the first peak from the detected SNP-preceding nucleotide and the second normal height peak resulting from the expressed allele. Amplicon sequences are listed in supplementary table S3, Supplementary Material online.

become easier to generate data to identify imprinted genes, but these techniques have their own biases that can lead to artifacts when not properly controlled for. Our experimental design and data analysis procedure avoids technical and other biases by following the general recommendations proposed by Wang and Clark (2014). In addition, we have applied the statistical approach proposed by DeVeale et al. (2012) and used mock comparisons within a given data set to control for false positives. Our results confirm the observations of DeVeale et al. (2012) that this leads to a fairly reliable exclusion of false positives while retaining good sensitivity, and therefore we recommend this as a standard procedure for future analyses.

The problem with overcalling candidate loci exists mostly for cases with weakly biased expression, for example, 45% of the transcript from one parental allele versus 55% from the other. Wang and Clark (2014) suggest to raise the cutoff to 35–65% to avoid this uncertainty zone all together. On the other hand, some genes with only a small bias (*Adam 23*, *Bcl2l1*, *Klhdc10*, and *Wars*) were independently discovered in our data set and the DeVeale et al. (2012) data set. Two of them (*Adam23* and *Wars*) were actually confirmed by quantitative pyrosequencing in the DeVeale et al. (2012) study. This gives confidence that the derivation of an FDR from the data is a suitable alternative to setting a more stringent cutoff. But independent of this, most of the newly discovered imprinted transcripts in our study would actually be called by both cutoff criteria (table 1); hence, we can consider them as real candidates for new imprinted genes in the mouse.

Many of the newly identified imprinted transcripts are located within a highly studied prototypic imprinted region

that consists of a 3 Mb large cluster of genes located on chromosome 7. A partial deletion of the region (mostly encompassing *Ube3a*) in mice results in impaired learning and altered ultrasonic communication (Jiang et al. 2010). These are phenotypes that we had also implicated in the explanation in our mating preference study (Montero et al. 2013). We found that recognizing the mate from the appropriate populations involved a learning phase, that is, mice coming directly from cages into the seminatural environment were not yet primed for showing population-specific mate choice. Further, akin to what is known about behavioral imprinting of songs in birds, we suggested that ultrasonic vocalization (USV) could be involved in the mate choice decisions as well (Montero et al. 2013). In fact, in a parallel study on USV we found differences in song patterns between the two populations (von Merten et al. 2014).

Intriguingly, the same general region including *Ube3a* and *Snrpn/Snurf* harbors a cluster of four of the five detected transcripts with particularly high genetic differentiation between the populations. It represents a size of 1.1 Mb and thus appears to belong to the larger regions of high differentiation between these populations. In a previous genome-wide study based on high-density SNP chips to trace selective sweeps, we found about 300 similarly high differentiated regions with an average size of about 80 kb and a maximum size of up to 1.4 Mb between the German and French populations (Staubach et al. 2012). However, the *Ube3a/Snrpn/Snurf* region was not found as a candidate in this previous study, because it is not well represented by SNPs on the microarray, that is, it escaped the strict statistical criteria that we had applied in the study. A more detailed analysis of the region will become possible through analyzing full-genome resequencing data for these populations.

High genetic differentiation can be caused by positive selection in one or both populations or by random fixation along the lineages. However, the latter is less likely since regions of this size would be broken up by recombination under conditions of random coalescence processes. Still, although we cannot so far distinguish between selective and random events in this case with confidence, the very fact that it is a highly differentiated region implies that it could contribute to behavioral differences between the populations.

The whole PWS region that encompasses the genes *Ube3a* and *Snrpn/Snurf* has been first identified in humans as a sporadic disorder in which affected children have developmental delay, a characteristic behavioral profile, and a behaviorally caused obesity problem (Cassidy et al. 2012). They also suffer cognitive and speech and language impairments. The molecular pathogenesis and the roles of specific PWS genes in the complex phenotype are still poorly understood. At least 11 genes were found to be inactivated in PWS, most often because of a sporadic deletion of the paternal allele combined with the normal process of genomic imprinting that silences the maternal allele of these genes. The mouse PWS homologous region on chromosome 7 has been targeted in more than 30 mouse models of PWS that partly mimic the human phenotypes (Resnick et al. 2013). Most genes in the PWS region are paternally expressed; only *Ube3a* is maternally

expressed. Hence, our finding that most of the newly discovered HYP transcripts in the region are also paternally expressed is in line with these patterns. The paternal expression is controlled by an imprinting center, and its deletion abolishes paternal expression of the genes and results in PWS in humans and a comparable phenotype in mice (Yang et al. 1998).

The evolution of the PWS region has been mostly studied at a larger evolutionary scale so far. Imprinting in the region appears to have evolved more than a 100 Ma after the fusion of two originally nonimprinted regions that contained *Ube3a* and *Snrpn* (Rapkins et al. 2006). The region has then acquired additional transcripts at least partly by retrotransposition (Chai et al. 2001; Rapkins et al. 2006) with rodent-specific (Chai et al. 2001) and human-specific variants (Neumann et al. 2014). Hence, the region appears to be generally actively evolving.

The second highly differentiated transcript, *Peg13*, has not been functionally studied so far but is located in the intron of a maternally expressed gene (*Trappc9*) that has been implicated in humans with intellectual disability disorders (Kakar et al. 2012; Marangi et al. 2013) and may be involved in its regulation (Court et al. 2014). Interestingly, patients with loss-of-function alleles show symptoms very similar to Prader–Willi-like phenotypes (Marangi et al. 2013). Although it seems too early to speculate about interactions of genetic pathways between these regions, it seems interesting that the two most highly differentiated paternally expressed transcripts that we found in our screen are related to regions involved in cognitive ability syndromes.

Three other previously studied genes with a paternal expression bias and a behavioral function could have been relevant for the effects that we have observed in the German versus French populations studied in Montero et al. (2013). These are *Grb10*, *Peg3*, and *Rasgrf1*. *Grb10* was implicated in adult social dominance behavior (Garfield et al. 2011), and *Peg3* was suggested to have evolved to regulate males experience-dependent preference for receptive females (Swaney et al. 2008). *Rasgrf1* has been implicated in learning and memory, but the paternal expression appears to affect mostly olfactory learning and memory in neonatal mice (Drake et al. 2011). All three transcripts are neither differentiated between the populations, nor do they show signs of differential expression (as assessed by read counts in the HYP RNAseq data). Accordingly, we consider it as less likely that they contribute to the differential recognition effect that we had seen between members of the population. The same is true for the MUPs that are produced in the LIV and for which a transgenerational regulatory effect has been suggested (Nelson et al. 2013). Although we do see them expressed in our data, they show no signs of imprinting or enhanced differentiation. However, as MUPs represent a complex multi-gene family with variable copy numbers (Logan et al. 2008; Karn and Laukaitis 2012), there could also be an annotation and mapping problem that will need further investigation.

It has been shown that imprinted genes can influence adult behavior, including mating (Curley 2011). Our results point to the PWS region genes as candidates to be

investigated for their role in mating behavior. However, because this behavior is complex, it remains to be further investigated whether imprinted genes are directly or only indirectly involved in explaining the paternally determined preference behavior observed by Montero et al. (2013). Investigating the architecture of parent-of-origin effects in mice, Mott et al. (2014) suggest that these are not necessarily controlled directly by imprinted genes but may depend on interactions with nonimprinted genes. Hence, any other of the hundreds of highly differentiated regions between the French and German population (Staubach et al. 2012) could also be involved in the population-specific mate choice. For example, we also found for two genes involved in USV population-specific selective sweeps (von Merten et al. 2014). Still, based on the cumulative evidence discussed above it would seem that the genes within the *Ube3a–Snrpn* imprinted region on chromosome 7, possibly together with the *Trappc9* region on chromosome 13, are good candidates for being involved in population-specific mate choice decisions.

Materials and Methods

Crosses and Samples

For an initial screen for imprinted transcripts in our target tissues in house mice, we generated crosses between inbred stains of *M. m. domesticus* (strain WSB/EiJ) and *M. m. musculus* (PWD/PhJ). Reciprocal crosses were set up and female offspring (3–6 individuals per genotype and cross direction, supplementary table S1 Supplementary Material online) were sacrificed at the age of 11–13 weeks to prepare the HYP, the VNO, and the LIV from the same individuals. All dissections were done by the same person (I.M.) following personal instructions and standardized protocols designed by Valery Grinevich (MPI for Medical Research Heidelberg for HYP) and Masayo Omura (MPI for Biophysics, Frankfurt for VNO). For the HYP, brains were removed and further dissection was performed on wet ice using gross anatomical borders (i.e., optic chiasm arms, optic tracts, and mammillary body) as markers to remove the HYP. For the VNO, we followed the first steps of the procedure described in Meeks and Holy (2009). Prepared tissues were immediately frozen on dry ice and kept at −70° until RNA preparation.

Sequencing libraries were constructed from the polyadenylated fraction of the RNA and sequenced as 100 bp paired-ends on an Illumina HiSeq sequencer. Libraries were prepared with the Illumina TruSeq RNA Sample Preparation Kit v2. To confirm the occurrence of the newly identified candidate transcripts in the wild type populations and to quantify allele frequencies of differentiating SNPs, we sequenced HYP RNA from 16 individuals from two wild-caught outbred *M. m. domesticus* populations that consisted of eight unrelated individuals each from France and Germany. Supplementary table S1, Supplementary Material online, provides the read statistics for both experiments. More details on the derivation and molecular analysis of these populations can be found in Ihle et al. (2006), Teschke et al. (2008), and Staubach et al. (2012). To confirm the imprinting status for a subset of the transcripts, we generated reciprocal crosses from these

populations and prepared HYP RNA for pyrosequencing assays. Part of the preparations for this experiment were performed by a second person trained by I.M.

SNP Identification

Reads were aligned by TopHat 2.0.4 to the mm10 mouse reference genome (GRCm38) using the transcriptome defined by Mus_musculus.GRCm38.68. Up to four mismatches (up to three for German and French population samples) were allowed. Only uniquely mapping reads were kept and duplicated reads were discarded as PCR/optical artifacts with Picard tools. Reads were realigned around indels and recalibrated according to known SNPs with GATK by using SNPs derived from sequenced PWD RNA samples (supplementary table S1, Supplementary Material online) as well as available SNPs for the strains PWD, WSB, and B6. SNP calling was performed with Unified Genotyper from Genome AnalysisTK-2.1-11. The tissues were analyzed separately. PWD SNPs were called from PWD samples and WSB SNPs were called from PWDxWSB and WSBxPWD samples together. At least eight reads per sample were required for SNP calling, and only genotypes without strand bias (GATK FS parameter > 60) and that were consistent in all individuals from one cross direction were kept.

To call SNPs between French and German populations, deduplicated uniquely mapping reads with a depth of at least ten per sample and no strand bias were used. A subset of SNPs where only PWD or WSB differ from the reference genomes, or both differ but in different ways, was identified as SNPs between WSB and PWD. To minimize strain bias in mapping, two strain-specific genomes were then constructed in silico by introducing all identified PWD or WSB SNPs into the reference genome. All raw reads from F1s were remapped to both reference genomes (as described above), and a maximum of one mismatch was allowed. If a read was mapping to different coordinates in the PWD and WSB genomes, the better quality mapping was kept. Uniquely mapping deduplicated reads were used for quantification of allelic expression.

Detection of Imprinting

For all SNPs between WSB and PWD, we counted remapped reads in each of the parental versions. Reads from all individuals from a given cross direction were summed. Significance of allele-specific expression (ASE) was tested by a Chi-square test against equal proportions. Loci with the same parental ASE bias in reciprocal crosses were considered to be imprinted. To set an appropriate significance cutoff, we followed the procedure described by DeVeale et al. (2012). Briefly, we combined reads from three samples from a given cross direction and compared allelic expression for every SNP with pooled reads depth ≥24. We also performed the same analysis with reads from the same cross direction (WSBxPWD) split into two groups. Such mock comparison should not detect any imprinting, and therefore estimates a cutoff that limits the FDR to the desired level (see fig. 1 for further explanation). Three sets of 60 million reads each were tested in

this way. The ratios of significant SNPs for the mock comparison and the real comparison were then plotted for a range of cutoff values, and the value representing an FDR of ≤ 0.05 was used for the identification of imprinted transcripts (fig. 1). All identified SNPs and their allelic imbalance values are provided in supplementary table S2, Supplementary Material online. The transcripts listed as imprinted in table 1 are annotated and named UCSC genes from the mm10 reference annotation. Additional SNPs were seen outside of annotated gene regions and usually correspond to expressed sequence tags that so far lack annotations. We have refrained from adding these to the list to remain conservative.

Assessment of Genetic Differentiation

SNPs were called for HYP transcripts as described above in individuals from French and German populations. Variance components and fixation indices (F_{st}) between the two populations were estimated according to Weir and Cockerham (1984) as implemented in R by Eva Chan (http://www.eva-chan.org/rscripts.html, last accessed September 5, 2014). As we focus on imprinted transcripts, we used a "single-parent" correction: Heterozygous positions were replaced with a randomly chosen allele and F_{st} was computed. The median value of F_{st} computed from 100 draws was used as the final value for any given position.

Confirmation by Pyrosequencing

Diagnostic SNPs were chosen from four candidate transcripts to design primers for their respective pyrosequencing assay (supplementary table S3, Supplementary Material online) using the Pyromark Assay Design software 2.0, assay scores > 97. Hypothalamus RNA from pure breeds and reciprocal crosses, from three individuals each, was used to generate biotinylated amplicons using the Pyromark OneStep RT-PCR kit (Qiagen). Pyrosequencing was done on a Pyromark ID PSQ 96 MA Sequencer (Qiagen) following the manufacturer's instructions, and allele frequencies were determined using the instruments' software.

Supplementary Material

Supplementary tables S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

References

Arnaud P, Monk D, Hitchins M, Gordon E, Dean W, Beechey CV, Peters J, Craigen W, Preece M, Stanier P, et al. 2003. Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. *Hum Mol Genet.* 12:1005–1019.

Cassidy SB, Schwartz S, Miller JL, Driscoll DJ. 2012. Prader–Willi syndrome. *Genet Med.* 14:10–26.

Chai JH, Locke DP, Ohta T, Greally JM, Nicholls RD. 2001. Retrotransposed genes such as Frat3 in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. *Mamm Genome.* 12:813–821.

Chao W, D'Amore PA. 2008. IGF2: epigenetic regulation and role in development and disease. *Cytokine Growth Factor Rev.* 19: 111–120.

Court F, Camprubi C, Garcia CV, Guillaumet-Adkins A, Sparago A, Seruggia D, Sandoval J, Esteller M, Martin-Trujillo A, Riccio A, et al. 2014. The PEG13-DMR and brain-specific enhancers dictate imprinted expression within the 8q24 intellectual disability risk locus. *Epigenetics Chromatin* 7:5.

Curley JP. 2011. Is there a genomically imprinted social brain? *Bioessays* 33:662–668.

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

Drake NM, DeVito LM, Cleland TA, Soloway PD. 2011. Imprinted Rasgrf1 expression in neonatal mice affects olfactory learning and memory. *Genes Brain Behav.* 10:392–403.

Garfield AS, Cowley M, Smith FM, Moorwood K, Stewart-Cox JE, Gilroy K, Baker S, Xia J, Dalley JW, Hurst LD, et al. 2011. Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature* 469:534–538.

Gregg C, Zhang JW, Butler JE, Haig D, Dulac C. 2010. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* 329: 682–685.

Gregg C, Zhang JW, Weissbourd B, Luo SJ, Schroth GP, Haig D, Dulac C. 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329:643–648.

Horsthemke B, Wagstaff J. 2008. Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am J Med Genet A.* 146A: 2041–2052.

Ihle S, Ravaoarimanana I, Thomas M, Tautz D. 2006. An analysis of signatures of selective sweeps in natural populations of the house mouse. *Mol Biol Evol.* 23:790–797.

Janotova K, Stopka P. 2011. The level of major urinary proteins is socially regulated in wild *Mus musculus musculus. J Chem Ecol.* 37:647–656.

Jiang YH, Pan YZ, Zhu L, Landa L, Yoo J, Spencer C, Lorenzo I, Brilliant M, Noebels J, Beaudet AL. 2010. Altered ultrasonic vocalization and impaired learning and memory in Angelman syndrome mouse model with a large maternal deletion from Ube3a to Gabrb3. *PLoS One* 5:e12278.

Kakar N, Goebel I, Daud S, Nurnberg G, Agha N, Ahmad A, Nurnberg P, Kubisch C, Ahmad J, Borck G. 2012. A homozygous splice site mutation in TRAPPC9 causes intellectual disability and microcephaly. *Eur J Med Genet.* 55:727–731.

Karn RC, Laukaitis CM. 2012. The roles of gene duplication, gene conversion and positive selection in rodent Esp and Mup pheromone gene families with comparison to the Abp family. *PLoS One* 7: e47697.

Leinders-Zufall T, Ishii T, Mombaerts P, Zufall F, Boehm T. 2009. Structural requirements for the activation of vomeronasal sensory neurons by MHC peptides. *Nat Neurosci.* 12:1551–U1598.

Logan DW, Marton TF, Stowers L. 2008. Species specificity in major urinary proteins by parallel evolution. *PLoS One* 3:e3280.

Marangi G, Leuzzi V, Manti F, Lattante S, Orteschi D, Pecile V, Neri G, Zollino M. 2013. TRAPPC9-related autosomal recessive intellectual disability: report of a new mutation and clinical phenotype. *Eur J Hum Genet.* 21:229–232.

Meeks JP, Holy TE. 2009. An ex vivo preparation of the intact mouse vomeronasal organ and accessory olfactory bulb. *J. Neurosci Methods.* 177:440–447.

Montero I, Teschke M, Tautz D. 2013. Paternal imprinting of mating preferences between natural populations of house mice (*Mus musculus domesticus*). *Mol Ecol.* 22:2549–2562.

MBE

Mott R, Yuan W, Kaisaki P, Gan XC, Cleak J, Edwards A, Baud A, Flint J. 2014. The architecture of parent-of-origin effects in mice. *Cell* 156: 332–342.

Nelson AC, Cauceglia JW, Merkley SD, Youngson NA, Oler AJ, Nelson RJ, Cairns BR, Whitelaw E, Potts WK. 2013. Reintroducing domesticated wild mice to sociality induces adaptive transgenerational effects on MUP expression. *Proc Natl Acad Sci U S A*. 110:19848–19853.

Neumann LC, Feiner N, Meyer A, Buiting K, Horsthemke B. 2014. The imprinted NPAP1 gene in the Prader-Willi syndrome region belongs to a POM121-related family of retrogenes. *Genome Biol Evol*. 6: 344–351.

Rapkins RW, Hore T, Smithwick M, Ager E, Pask AJ, Renfree MB, Kohn M, Hameister H, Nicholls RD, Deakin JE, et al. 2006. Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet*. 2:1666–1675.

Resnick JL, Nicholls RD, Wevrick R. 2013. Recommendations for the investigation of animal models of Prader-Willi syndrome. *Mamm Genome*. 24:165–178.

Roberts SA, Simpson DM, Armstrong SD, Davidson AJ, Robertson DH, McLean L, Beynon RJ, Hurst JL. 2010. Darcin: a male pheromone that stimulates female memory and sexual attraction to an individual male's odour. *BMC Biol*. 8:75.

Staubach F, Lorenc A, Messer PW, Tang K, Petrov DA, Tautz D. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genet*. 8:e1002891.

Stockley P, Bottell L, Hurst JL. 2013. Wake up and smell the conflict: odour signals in female competition. *Philos Trans R Soc Lond B Biol Sci*. 368:20130082.

Swaney WT, Curley JP, Champagne FA, Keverne EB. 2008. The paternally expressed gene Peg3 regulates sexual experience-dependent preferences for estrous odors. *Behav Neurosci*. 122:963–973.

Teschke M, Mukabayire O, Wiehe T, Tautz D. 2008. Identification of selective sweeps in closely related populations of the house mouse based on microsatellite scans. *Genetics* 180:1537–1545.

von Merten S, Hojer S, Pfeifle C, Tautz D. 2014. A role for ultrasonic vocalisation in social communication and divergence of natural populations of the house mouse (*Mus musculus domesticus*). *PLoS One* 9(5):e97244.

Wang X, Clark AG. 2014. Using next-generation RNA sequencing to identify imprinted genes. *Heredity* 113(2):156–166.

Wang X, Soloway PD, Clark AG. 2010. Paternally biased X inactivation in mouse neonatal brain. *Genome Biol*. 11:R79.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population-structure. *Evolution* 38:1358–1370.

Wilkins JF, Haig D. 2003. What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet*. 4: 359–368.

Wolf JB, Cheverud JM, Roseman C, Hager R. 2008. Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLoS Genet*. 4:e1000091.

Wolf JB, Hager R. 2009. Selective abortion and the evolution of genomic imprinting. *J Evol Biol*. 22:2519–2523.

Wood AJ, Oakey RJ. 2006. Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet*. 2:1677–1685.

Wood AJ, Schulz R, Woodfine K, Koltowska K, Beechey CV, Peters J, Bourc'his D, Oakey RJ. 2008. Regulation of alternative polyadenylation by genomic imprinting. *Genes Dev*. 22:1141–1146.

Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, Jenkins NA, Copeland NG, Brannan CI. 1998. A mouse model for Prader–Willi syndrome imprinting-centre mutations. *Nat Genet*. 19:25–31.