

Transcriptome-wide investigation of genomic imprinting in chicken

Laure Frésard^{1,2}, Sophie Leroux^{1,2}, Bertrand Servin^{1,2}, David Gourichon³, Patrice Dehais^{1,2,4}, Magali San Cristobal^{1,2}, Nathalie Marsaud⁵, Florence Vignoles^{1,2}, Bertrand Bed'hom^{6,7}, Jean-Luc Coville^{6,7}, Farhad Hormozdiari⁸, Catherine Beaumont⁹, Tatiana Zerjal^{6,7}, Alain Vignal^{1,2}, Mireille Morisson^{1,2}, Sandrine Lagarrigue¹⁰ and Frédérique Pitel^{1,2,*}

¹INRA, UMR444 Laboratoire de Génétique Cellulaire, Castanet-Tolosan F-31326, France, ²ENVT, UMR444 Laboratoire de Génétique Cellulaire, Toulouse F-31076, France, ³INRA, PEAT Pôle d'Expérimentation Avicole de Tours, Nouzilly F- 37380, France, ⁴INRA, Sigene UR875 Biométrie et Intelligence Artificielle, Castanet-Tolosan F-31326, France, ⁵INRA, GeT-PlaGe Genotoul, Castanet-Tolosan F-31326, France, ⁶INRA, UMR1313 Génétique animale et biologie intégrative, Jouy en Josas F-78350, France, ⁷AgroParisTech, UMR1313 Génétique animale et biologie intégrative, Jouy en Josas F-78350, France, ⁸Department of Computer Sciences, University of California, Los Angeles, CA 90095, USA, ⁹INRA, UR83 Recherche Avicoles, Nouzilly F- 37380, France and ¹⁰Agrocampus Ouest, UMR1348 Physiologie, Environnement et Génétique pour l'Animal et les Systèmes d'Élevage, Animal Genetics Laboratory, Rennes F-35000, France

Received November 14, 2013; Revised December 17, 2013; Accepted December 18, 2013

ABSTRACT

Genomic imprinting is an epigenetic mechanism by which alleles of some specific genes are expressed in a parent-of-origin manner. It has been observed in mammals and marsupials, but not in birds. Until now, only a few genes orthologous to mammalian imprinted ones have been analyzed in chicken and did not demonstrate any evidence of imprinting in this species. However, several published observations such as imprinted-like QTL in poultry or reciprocal effects keep the question open. Our main objective was thus to screen the entire chicken genome for parental-allele-specific differential expression on whole embryonic transcriptomes, using high-throughput sequencing. To identify the parental origin of each observed haplotype, two chicken experimental populations were used, as inbred and as genetically distant as possible. Two families were produced from two reciprocal crosses. Transcripts from 20 embryos were sequenced using NGS technology, producing ~200 Gb of sequences. This allowed the detection of 79 potentially imprinted SNPs, through an analysis method that we validated by detecting imprinting from mouse data already published.

However, out of 23 candidates tested by pyrosequencing, none could be confirmed. These results come together, without *a priori*, with previous statements and phylogenetic considerations assessing the absence of genomic imprinting in chicken.

INTRODUCTION

Parental genomic imprinting is a process that leads to the differential expression of alleles depending on their parental origin. This phenomenon has been described in eutherian mammals, marsupials, plants and insects, but has never been observed in birds (1,2).

Even if not the only one (3), the main theory proposed to explain this phenomenon is the parental conflict hypothesis, also known as the kinship theory. Briefly, imprinting is leading to an imbalance of parental allele contribution, and the parental conflict hypothesis argues that the genes controlling the allocation of resources to the embryo should be particularly affected, with the maternal genome restricting the use of resources, allowing to preserve the mother and future progeny, and the paternal genome favoring growth of his offspring (4–8). Readers are referred to (9) for an in-depth discussion of the kinship theory. According to this theory, imprinting should be restricted to organisms in which maternal resources directly affect the embryo.

*To whom correspondence should be addressed. Tel: +33 561 28 54 35; Fax: +33 561 28 53 08; Email: frederique.pitel@toulouse.inra.fr

It would therefore be unlikely to observe this phenomenon in oviparous animals (10), where the amount of resources is limited by the egg size, controlled by the mother. Genomic imprinting evolved both in angiosperm plants and placental mammals (11) but is also seen in insects, and it has even been shown that it arose several times during mammals' evolution due to different selective pressure at several loci (12). Regardless of the kinship theory, the appearance of genomic imprinting in nonmammalian animals is thus not implausible. Genomic imprinting was studied in birds, but until now, only the expression of genes known to be imprinted in mammals has been analyzed in chickens, most of them showing biallelic expression and thus, leading to the conclusion of the lack of imprinting in this species: several studies have been conducted on *IGF2*, known to be paternally expressed in mice and humans (13), all of them except one (14) leading to the conclusion that it was biallelically expressed (15–19). In this genomic region, the H19 imprinting center, controlling an imprinted cluster including *IGF2*, is shown to be absent from the chicken genome (19). Other genes known to be imprinted in mammals have been shown to be biallelically expressed in chicken, such as *ASCL2/MASH2*, *M6PR/IGF2R*, *DLK1* and *UBE3A* (19–21). However, parent-of-origin QTLs have been detected in chicken and quail (22–24). Whereas some of them may finally not be considered as relevant to genomic imprinting due to linkage disequilibrium or bias generated by the experimental design (25,26), others appeared to be consistent, when using appropriate experimental animal population designs and methodology (25,27). Moreover, reciprocal effects, of great importance in poultry production, may be partly explained by imprinting. Together with several molecular features, such as the absence of *DNMT3L* (key factor of the establishment of DNA methylation) in the chicken genome (28) and the possible existence of imprinting mechanisms other than DNA methylation (29,30), these contradictory observations encourage to definitely answer the question (31).

Advances in sequencing technologies now allow to study genomic imprinting through whole transcriptome sequencing (32–34), avoiding any *a priori* choice of target genomic regions, such as the chicken homologous of imprinted regions in mammals. We therefore chose Next Generation Sequencing (NGS) technology to analyze imprinting in chicken, without any preconception on whether it exists. Chicken is a model species for numerous studies in birds, and the availability of inbred lines is a great advantage for our purpose: we set up an experimental design consisting in two families from two reciprocal crosses between chicken lines, chosen so as to allow the identification of the parental origin of each allele in embryonic offspring. We sequenced whole embryos transcriptomes, and the relative expression level of each parental allele was compared, for the informative loci, between both cross ways.

MATERIALS AND METHODS

Ethics statement

Chickens were bred at INRA, UE1295 Pôle d'Expérimentation Avicole de Tours, F-37380 Nouzilly,

in accordance with European Union Guidelines for animal care. The farm is registered by the Ministry of Agriculture with the license number C37–175–1 for animal experimentation. The experiment was realized under authorization 37–002 delivered to D. Gourichon.

Chicken lines and crosses

To maximize the probability of identifying the parental origin of an allele in F1 individuals from two reciprocal crosses, we selected 2 chicken lines among 10 (Supplementary Table S1), available at INRA experimental farm (UE1295 PEAT, F-37380 Nouzilly, France). Genotyping was performed on 3–12 individuals from each population with an Illumina 57K Single Nucleotide Polymorphism (SNP) array [57 636 SNPs (35)]. SNPs were filtered (call freq > 0.9, GenTrainScore > 0.5, Minor Allele Frequency (MAF) > 0.01) and allele frequencies within each line calculated with plink, option—freq (36). Genetic similarities between lines (homozygosities for a given line) were then calculated (Figure 1) from the 52 189 filtered SNPs. The most genetically distant pair of lines together with the highest homozygosity each were selected, i.e. the Line 6 (37) and the Line R⁻ (38). A male and a female from each line were reciprocally crossed, and 24 F1 embryos, 12 from each cross (6 males and 6 females per cross), were harvested from the same batch at embryonic day 4.5 (stage 26). Genomic DNA and total RNA were concurrently extracted from the same samples of crushed whole embryos without extraembryonic membranes (AllPrep DNA/RNA Mini Kit, Qiagen). RNA quality was measured by a BioAnalyzer (Agilent); all samples had a RNA Integrity Number \geq 9.9.

Parental genomic DNA was extracted from blood samples through a high-salt extraction method (39).

Sequencing

Chicken data

RNA sequencing. Libraries with a mean insert size of 200 bp were prepared following Illumina instruction for RNA-Seq analysis, by selecting polyA⁺ fragments (TruSeq RNA Sample Prep Kit) from each sample. Samples were tagged to allow subsequent identification, amplified by polymerase chain reaction (PCR) and quantified by quantitative PCR (Agilent QPCR Library Quantification Kit) and then sequenced (paired ends, 100 bp) in triplicate on an Illumina HiSeq 2000 sequencer (Illumina, TruSeq PE Cluster Kit v3, cBot and TruSeq SBS Kit v3) by randomizing their position in six different sequencing lanes.

DNA sequencing. Parental DNA was sequenced on four lanes of Illumina HiSeq 2000 to allow the detection of SNPs discriminating the two lines. Library preparation (mean insert size of 300 bp), DNA quantification and sequencing (paired ends, 100 bp) were performed according to the manufacturer instructions (TruSeq DNA Sample Prep Kit Illumina, Agilent QPCR Library Quantification Kit, TruSeq PE Cluster Kit v3 cBot TruSeq SBS Kit v3).

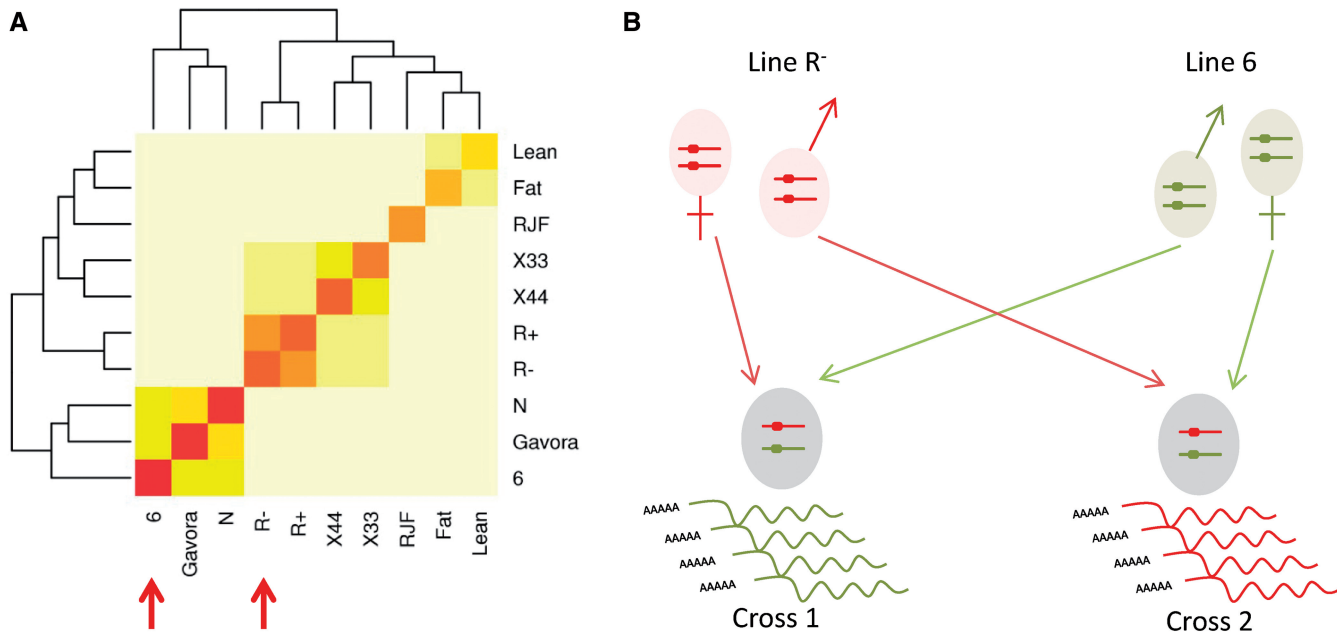


Figure 1. Experimental design. (A) Heatmap of genetic similarities between lines. Colors correspond to the inbred level of tested individuals (red is inbred). Lines 6 and R⁻ (red arrows) were chosen to be as inbred and as genetically distant as possible. (B) Reciprocal cross between the two selected lines. In case of genomic imprinting, embryos are preferentially expressing one of their parents' allele (this figure shows a case of paternally expressed gene). Detection of such event is possible with polymorphisms differentiating the lines.

Mouse data

RNA-seq sequences were collected from a previous study (40). They correspond to liver mRNA from 18 F1 males and females DBA/2J and C57BL/6J reciprocally crossed mice (two pools of nine mice). These crosses are termed as DXB and BXD in the article.

Computational analysis

When not specified, analyses were made with homemade Perl, Python and R scripts.

Genomic sequences analyses

Reads from DNA sequencing of parents were aligned on the last assembly of chicken genome (Galgal4) using the bwa program version 0.6.1, option aln (41).

Identification of SNP discriminating parents from both lines

Allelic counts for each base position on the genome were determined. These counts allowed selecting only loci discriminating parents, i.e. positions where both parents from each line were homozygous for distinct alleles.

The functional consequence of each SNP in each transcript was predicted using the Ensembl Variant Effect Predictor version 71 (42).

Transcriptome characterization

Chicken data. Tophat 2.0.5 software (43) was used to align transcript sequences from the F1 embryos on Galgal4 chicken assembly. Reads were first mapped allowing multiple alignments, with the intron length parameter set between 3 and 25000bp to limit the number of false positive while discovering most junctions, the deletion length was limited to 1 instead of 3, and the

mate inner distance in our case corresponded to 200. The option read-realign-edit-dist was set to 0 to map every read in all the Tophat mapping steps to get the best possible alignment. Finally, a micro exon search was applied. Other alignment options were set as default.

To remove potential PCR reads duplicates from the analysis, the program samtools rmdup (41) was run on each individual after alignment. Afterward, reads with potential multiple locations were discarded.

Samtools flagstat command (41) was used to generate statistics on aligned reads, notably the total number of reads mapped after applying our criteria.

To characterize the general level of expression in all the embryos, cufflinks software 2.0.0 was used on Galgal4.72 GTF file to quantify expression on known genes (44).

Mouse data. Sequencing reads were aligned with Tophat 2.0.5 on an artificial reference genome where all SNPs known between DBA/2J and C57BL/6J (45) were converted to 'N' to prevent the preferential mapping of the reference allele. In order not to affect to a large extent the alignment efficiency, no penalty was set for reads mapping on N in this new reference. A maximum of three mismatches was allowed for a read to be kept. Finally, a filter was applied to keep only reads uniquely mapped to the reference genome.

Identification of SNP with biased expression

Transcripts alignments were merged by parental cross for further analysis. Allelic counts allowed to select SNPs represented with at least 10 reads in each cross, with an inverted allelic ratio between both crosses (Fold Change ≥ 2.5).

To test the expression bias, a Fisher Exact Test was made on every transcribed loci and the significance

threshold was fixed at False Discovery Rate (FDR) < 0.05 using the R package 'qvalue' (46).

Validation assays

Pyrosequencing

Twenty-three markers were tested on a Qiagen PyroMark Q24 sequencer. Primers were designed using PyroMark Assay Design software (Supplementary Table S2). PCR samples were produced using PyroMark PCR Kit (Qiagen). Runs were analyzed by PyroMark Q24 1.0.10 software with default analysis parameters.

Droplet Digital PCR

In total, three markers were tested on Bio-Rad QX100™ Droplet Digital PCR (ddPCR) System machine. Primers and probes were ordered at Sigma (Supplementary Table S3).

RESULTS

Constitution of an informative device

To detect genomic imprinting for a given gene, it is required that we are able to discriminate the parental origin of its expressed allele in the F1 embryos. The principal purpose of the lines' choice was thus to maximize the chances to identify from which parent the expressed allele had been transmitted to the embryos. To select parents as inbred (to reduce intra-population polymorphism) and as genetically distant (to maximize inter-population polymorphism) as possible, 3–12 animals from 10 chicken populations were genotyped on an Illumina 57K SNP array (35). Line R⁻ and Line 6 were the two that fitted the best these criteria (Figure 1A) and were reciprocally crossed: one male from Line 6 was mated with a female from Line R⁻ and one R⁻ male was crossed with one female from Line 6 (Figure 1B). We selected 12 embryos from each reciprocal cross, with balanced ratio of six males and six females from each cross.

To detect SNPs discriminating the parents from both lines, genomic DNA from all four parents was sequenced. A total of 897 648 484 paired-ends reads (100 bp) were produced from an Illumina HiSeq 2000 sequencer and 813 407 461 (91%) aligned to the Galgal4 reference genome, which was thus covered at 98%, with a mean depth of 20×.

We found 2 298 622 SNPs discriminating the parents from both lines on genomic DNA, amounting to 2.1 ± 0.7 SNP/kb of sequence. When considering the number of annotated genes with at least one discriminating SNP, we found that 12 303 genes were represented. It corresponds to 72% of chicken annotated genes (Supplementary Table S4). Among the observed discriminating SNPs, 54.9% were localized in intergenic regions (Supplementary Figure S1), and 27.4% of the SNP detected in the coding region were nonsynonymous, which is similar to results previously obtained in chicken (47,48).

To have a better idea of their density in the potentially expressed regions, SNPs within exons were considered, that is 55 016 SNPs (i.e. 2% of detected SNPs). It corresponds to a density of 1.3 SNP/kb of coding regions, which is sufficient to have a reasonable screening of the exome.

Nevertheless, we used the total set of available discriminating SNPs in the study, to include nonannotated regions in the analysis.

The fraction of these SNPs that will be observed in the RNA-Seq data represents the number of opportunities to link the allelic imbalance of transcripts detected in embryos with their parental origin.

Transcriptome characterization

To link the parental origin of the alleles to their expression pattern in embryos, we performed RNA-Seq on 12 embryos from each cross. We obtained reliable sequencing results from 20 embryos (SRA accession number SRP033603), 11 from Cross 1 (5 females and 6 males) and 9 from Cross 2 (4 females and 5 males); the other 4 were discarded because of sequencing tags problems.

From those 20 embryos, 1.33×10^9 and 6.07×10^8 sequences were produced from Cross 1 and Cross 2 individuals, respectively (200 Gb of sequence data altogether), corresponding on average to 121.1 million reads and 67.4 million reads for each individual in Cross 1 and Cross 2, respectively.

After alignment to the chicken assembly, removing of PCR duplicates and allowing only one possible hit, 74% of reads were mapped to the reference genome, which corresponds to 42% coverage of the genome (39% for Cross 1 and 35% for Cross 2, Supplementary Figure S2). We set a minimal depth of 10 reads to consider a region as covered, to avoid background of low expressed regions.

Regarding the general expression in chicken embryo, a minimal FPKM of 1 was expected (one fragment per kilobase per million reads) to call a gene as 'expressed'. Among 17 108 genes already described in chicken (available from galgal4.72 GTF file), 11 416 genes, i.e. 67% of them, had a FPKM superior or equal to 1 in our data set.

Filters—bias in expression

RNA sequences were filtered to detect allelic imbalance expression at loci discriminating parents from both lines. Around 300 million bases were counted as expressed in both crosses, meaning that at least one read was observed at this position in each cross. To avoid bias in expression linked to a lack of depth, a threshold of 10 reads had to be reached in each cross (which corresponds to a minimal depth of 20 in total), leading to 123 532 053 positions, i.e. 40% of base positions expressed in both crosses.

Among these loci, 5 388 148 were polymorphic, and therefore potentially informative to test allelic imbalance. When crossing these expressed positions to discriminating SNPs (201 209 positions), we found that 7731 of 11 416 expressed genes (68%) were informative, i.e. containing at least one SNP discriminating parents of embryos in both crosses (Supplementary Table S4). A Fisher Exact Test was applied on these 5 million SNPs to test if the expression profile was different between Cross 1 and Cross 2. At a 5% FDR threshold, we observed 117 740 loci for which expression profiles were significantly different between Cross 1 and Cross 2, including positions with an allelic bias in only one of the crosses. Only positions

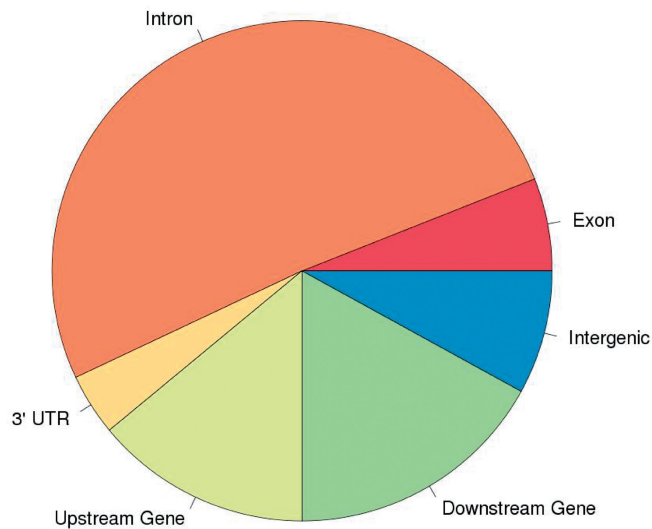


Figure 2. Genomic contexts of the candidate loci on the chicken genome.

where the allelic ratio was inverted between both crosses were potentially candidates, i.e. the minor allele in one cross is overrepresented in the reciprocal cross. Because the study was made on the whole embryo (a mixture of different tissues) a complete extinction of one allele was not necessarily expected, and we set a minimal fold change of 2.5 for this filter. A total of 4904 SNPs met these criteria. They could either result from polymorphisms in parental sequences or from an effective bias in expression. A further filtering was done removing the SNPs that, even if significant, were heterozygous in the parents and therefore not informative to trace back the allele parental origin. For that reason, linking these candidate SNPs to the parental origin of the allele was only possible at loci discriminating parents from both crosses. To do so, we crossed filtered data with the 2 298 622 discriminating SNPs detected in parallel on parents' genomic DNA. We checked that the detected allelic imbalance was similar for most of the pooled embryos in each cross, to avoid any bias due to a high monoallelic expression in few embryos (never observed in our candidates). At the end, we ended up with 79 potentially imprinted SNPs. These SNPs were mostly located in introns (51% of them); the others were upstream or downstream of annotated genes (31%), exons (6%), 3'-UTR (4%) or intergenic regions (8%) (Figure 2, Table 1). This distribution is similar to that of discriminating SNPs in sequences declared as expressed in both crosses, where 11% of these SNPs are located in known annotated transcripts (exons and UTR regions). This attribution was based on the known chicken annotation (Ensembl release 72, http://www.ensembl.org/Gallus_gallus/Info/Index). We considered expressed intergenic regions as unannotated regions.

Validation assays

We tested 23 candidate SNPs to confirm the Illumina HiSeq results with pyrosequencing as an independent validation method (49,50). A total of 16 embryos (8 per cross)

were tested. Tested candidates were chosen to get a panel of different situations highlighted after sequences analysis: intronic, exonic, regulatory regions and unannotated regions were looked at. Furthermore, candidates with different depth profiles were tested. To select the positions to be tested, we preferentially chose loci where SNPs surrounding the candidate position had an allelic profile similarly unbalanced: taking into account these additional positions—even if not passing our filters due to a lack of depth or an allelic ratio <2.5 —ensure to avoid artifactual SNPs.

First, to avoid any bias due to PCR amplification of an allele or to genomic duplication in one cross, we checked that the pattern obtained by pyrosequencing on genomic DNA from the embryos truly corresponds to a heterozygous profile. Then we quantified the allelic ratio observed on cDNA.

Unfortunately, while confirming the heterozygous status of embryos in each case, we could not observe on cDNA any allelic imbalance corresponding to an imprinting profile (Figure 3).

Three markers were also tested by droplet digital PCR as another complementary quantitative approach. This confirmed the pyrosequencing results, further invalidating our HiSeq results (Figure 3).

Method validation with RNA sequencing data from mouse

As parental genomic imprinting is thought to be absent in chicken, we used as 'positive control' the mouse data set produced by Lagarrigue and collaborators (40) to detect imprinted genes in mouse liver, to validate the approach we proposed on chicken. In the mouse study, a reciprocal cross was made between DBA/2J (further called D) and C57BL/6J (further called B) mouse strains and RNA was sequenced from liver samples of 18 F1 mice.

Out of 285 166 513 reads generated by RNA-seq in both mouse crosses, 145 412 592 reads (51%) were mapped uniquely on an artificial reference genome where SNPs between both mouse strains were replaced by 'N'. It corresponds to 74 733 541 reads for one cross (DXB) and 70 679 051 reads for the reciprocal cross (BXD).

These sequences were then filtered for bias in expression following the same pipeline as for chicken embryos. At a 5% FDR, we found 20 candidate SNPs with an inverted ratio between both crosses (minimum Fold Change of 2.5, Table 2). Of these, five are located in three genes known to be imprinted in mouse: *H13* (33), *Peg13* (51), both detected in the original study, and *Snurf* (52), not revealed by the first analysis. A fourth gene, *Sgce*, also detected by Lagarrigue *et al.*, was identified when less stringent conditions on sequencing depth threshold were applied (i.e. when accepting nine reads in each cross).

These results confirm the efficiency of our pipeline to detect genomic imprinting from RNA-Seq data that allowed not only confirming three genes identified by Lagarrigue *et al.* but also demonstrating the presence of another imprinted gene, *Snurf*, which had not been detected in their study. The newly discovered loci have to be further analyzed.

Table 1. Candidate positions, allelic counts and statistical results

Chromosome	Position	P-value	Allelic ratio Cross 1	Allelic ratio Cross 2	Read count for Cross 1		Read count for Cross 2		Gene	Gene short name	FPKM All embryos	Transcript	Ensembl annotation
					Line R- count (female)	Line R- count (male)	Line R- count (female)	Line R- count (male)					
1	12338856	1.08E-04	0.36	2.71	12	35	18	6	ENSGALG00000008328	RSBNIL	12.17	ENSGALT000000013555	INTRON
1	12338917	2.00E-04	0.30	2.50	9	32	14	14	ENSGALG00000008328	RSBNIL	12.17	ENSGALT000000013555	INTRON
1	32563608	0.00E+00	15.95	0.09	37	605	2053	193	-	-	-	-	INTERGENIC
1	33595711	5.01E-07	20.00	0.31	19	0	4	15	ENSGALG00000009835	XPO7	51.98	ENSGALT000000015988	DOWNSTREAM GENE
1	37687954	8.07E-06	0.34	2.50	49	16	9	24	ENSGALG00000010208	CAFS2	0.95	ENSGALT000000016598	EXON
1	39422816	9.63E-05	0.39	5.67	6	17	16	2	ENSGALG00000010320	PAWS2	20.45	ENSGALT000000016802	INTRON
1	44707119	8.16E-06	0.23	2.67	25	5	8	23	ENSGALG000000011285	-	64.82	ENSGALT000000018413	INTRON
1	44752101	1.60E-20	4.46	0.22	57	12	26	123	ENSGALG000000011293	-	44.72	ENSGALT000000018422	DOWNSTREAM GENE
1	47566268	6.13E-04	3.50	0.40	20	5	5	14	ENSGALG000000011651	ARL1	70.20	ENSGALT000000019040	INTRON
1	83083794	6.38E-05	2.57	0.38	35	13	8	23	ENSGALG000000011643	UTP20	11.36	ENSGALT000000019028	DOWNSTREAM GENE
1	103304118	3.08E-05	2.75	0.24	21	7	4	20	ENSGALG000000015804	BZ1F1	11.24	ENSGALT000000024570	INTRON
1	104233850	2.56E-04	3.25	0.29	12	3	6	23	ENSGALG000000015887	URB1	8.22	ENSGALT000000025483	INTRON
1	113276853	6.44E-04	0.30	2.67	5	19	15	5	ENSGALG000000027198	TMEM47	20.96	ENSGALT000000045448	INTRON
1	120338773	1.47E-06	3.75	0.10	14	3	1	19	ENSGALG000000016541	BEND2	1.28	ENSGALT000000026692	INTRON
1	120338806	3.54E-06	4.00	0.15	15	3	2	19	ENSGALG000000016541	BEND2	1.28	ENSGALT000000026692	INTRON
1	129355531	5.49E-04	0.36	5.00	21	7	1	9	ENSGALG000000016702	PPP2R3B	5.25	ENSGALT000000026955	3' UTR
1	154366155	8.90E-05	0.30	4.00	32	9	2	11	ENSGALG000000016716	-	8.91	ENSGALT000000026974	INTRON
1	154366157	2.04E-05	3.30	0.27	32	9	2	10	ENSGALG000000016925	TBC1D4	8.53	ENSGALT000000027361	INTRON
1	194707275	2.94E-04	0.30	5.33	5	19	15	2	ENSGALG000000017322	TBC1D4	8.53	ENSGALT000000027361	INTRON
2	42801285	8.05E-04	2.78	0.36	8	24	13	4	ENSGALG000000017321	PLEKH81	9.43	ENSGALT000000027993	UPSTREAM GENE
2	64390351	3.84E-06	3.00	0.20	17	5	5	29	ENSGALG000000011654	ACPP	0.51	ENSGALT000000019043	UPSTREAM GENE
2	85818148	7.42E-06	0.40	4.22	7	19	37	8	ENSGALG0000000212781	EFIE1E1	64.98	ENSGALT000000020851	INTRON
2	107189780	9.83E-04	0.17	1.11	11	11	3	3	ENSGALG000000013387	CIPTM1L	29.23	ENSGALT000000021546	INTRON
2	107698638	6.38E-04	3.00	0.36	7	23	13	4	ENSGALG000000012914	PKMDC	6.38	ENSGALT000000021060	UPSTREAM GENE
2	116079172	1.68E-04	5.50	0.21	10	1	2	13	ENSGALG000000015574	MCM4	53.61	ENSGALT000000021064	INTRON
3	9834854	2.77E-12	4.70	0.33	46	9	21	66	ENSGALG000000008811	SIC5A1	7.98	ENSGALT000000025131	INTRON
3	14807328	1.15E-06	2.76	0.32	20	57	24	7	ENSGALG000000008702	SIC1A4	13.62	ENSGALT000000014322	DOWNSTREAM GENE
3	39864572	3.61E-04	2.80	0.20	13	4	2	14	ENSGALG000000008698	CZORF72	29.81	ENSGALT000000014157	INTRON
3	65470116	6.44E-04	3.33	0.38	19	5	5	15	ENSGALG000000016419	C1orf198	81.24	ENSGALT000000045254	INTRON
3	95715441	1.86E-04	2.83	0.20	16	5	2	14	ENSGALG000000016704	ASAP2	15.96	ENSGALT000000026486	INTRON
3	108502770	6.08E-05	8.33	0.40	24	2	3	9	ENSGALG000000006877	DC51	62.40	ENSGALT000000036694	INTRON
4	8064555	3.43E-06	0.09	5.00	21	1	1	9	ENSGALG000000006919	KLHL4	7.62	ENSGALT000000011141	INTRON
4	8649670	6.74E-16	0.04	4.76	26	0	6	80	ENSGALG000000006919	POF1B	3.18	ENSGALT000000011214	INTRON
4	8688872	1.81E-05	0.30	10.50	9	2	1	20	ENSGALG000000007993	-	-	-	INTERGENIC
4	13252457	4.84E-04	0.36	6.00	13	4	1	11	ENSGALG000000007993	DCX	29.76	ENSGALT000000012977	INTRON
4	51035600	1.58E-04	0.36	2.56	27	9	8	22	ENSGALG000000011762	UBA6	16.21	ENSGALT000000019202	INTRON
5	26479524	9.25E-04	0.28	3.00	28	7	2	8	-	-	-	-	INTERGENIC
5	26993012	5.39E-44	4.20	0.21	152	642	101	20	ENSGALG000000009415	SMOC1	11.86	ENSGALT000000016210	3' UTR
5	54981852	8.55E-05	2.82	0.39	30	10	2	8	ENSGALG000000009415	HEAT6A	11.50	ENSGALT000000016210	INTRON
5	58239517	7.71E-04	0.27	3.67	14	3	2	10	ENSGALG000000012426	DDHD1	13.43	ENSGALT000000020311	INTRON
6	2283386	9.83E-04	0.15	2.75	12	1	3	10	ENSGALG00000002117	PAR6	40.62	ENSGALT000000030306	INTRON
6	3936793	2.57E-05	0.21	2.50	422	88	3	9	ENSGALG00000002479	MATIA	4.13	ENSGALT00000003908	DOWNSTREAM GENE
6	3936829	2.08E-06	3.89	0.25	353	90	2	11	ENSGALG00000002479	MATIA	3.43	ENSGALT00000003908	EXON
6	8511844	7.21E-05	8.50	0.30	1	16	9	2	ENSGALG000000002493	-	4.13	ENSGALT00000000927	EXON
6	8511848	7.21E-05	8.50	0.30	1	16	9	2	ENSGALG000000003135	CHANK3	7.25	ENSGALT000000003135	INTRON

(continued)

Table 1. Continued

Chromosome	Position	P-value	Allelic ratio Cross 1		Allelic ratio Cross 2		Read count for Cross 1		Read count for Cross 2		Gene	Gene short name	FPKM All embryos	Transcript	Ensembl annotation
			Allelic ratio Cross 1	Allelic ratio Cross 2	Line 6 count (female)	Line R-count (male)	Line 6 count (female)	Line R-count (male)	Line 6 count (male)	Line R-count (female)					
7	2773655	6.4E-04	3.67	0.18	10	2	2	10	10	2	ENSGALG00000002921	LANCL1	31.73	ENSGALT00000004610	INTRON
7	16539884	7.29E-04	0.36	3.00	7	7	21	14	4	4	ENSGALG00000009325	OIA1	77.00	ENSGALT00000015187	INTRON
7	21996208	6.53E-04	2.50	0.23	12	2	7	12	12	12	ENSGALG00000011402	RQCD1	80.61	ENSGALT00000018598	INTRON
7	23060451	4.93E-04	0.31	2.67	19	3	3	14	39	39	ENSGALG00000011493	TMEM169	26.11	ENSGALT000000038136	INTRON
7	23060984	7.18E-07	0.32	3.46	24	7	7	12	44	44	ENSGALG00000011493	TMEM169	26.11	ENSGALT000000038136	INTRON
8	14872186	4.52E-05	0.40	3.75	23	59	59	14	3	3	ENSGALG00000006284	SH3GLB1	81.17	ENSGALT00000007186	DOWNSTREAM GENE
10	5799143	1.09E-05	0.50	0.21	34	9	9	2	13	13	ENSGALG00000004059	LRRC49	4.48	ENSGALT00000006457	DOWNSTREAM GENE
11	1067297	7.82E-06	16.00	0.27	15	0	0	2	10	10	ENSGALG00000005140	RRAD	2.96	ENSGALT00000008251	UPSTREAM GENE
12	11697037	4.83E-10	10.00	0.36	24	249	249	13	4	4	ENSGALG00000028227	PKRARA	31.01	ENSGALT000000044834	3' UTR
12	11716237	1.98E-05	2.84	0.17	18	53	53	11	1	1	ENSGALG00000007023	PKRARA	46.62	ENSGALT00000013388	UPSTREAM GENE
12	12137467	2.04E-04	0.22	2.60	22	4	4	2	12	12	ENSGALG00000007172	FHIT	2.28	ENSGALT00000011614	DOWNSTREAM GENE
13	6552018	4.02E-06	0.15	4.67	19	2	2	13	13	13	ENSGALG00000001698	GABRA1	1.15	ENSGALT00000002609	INTRON
13	8459982	4.04E-04	2.86	0.29	19	6	6	3	3	3	ENSGALG00000002863	ERGIC1	95.63	ENSGALT00000004517	INTRON
13	9544586	2.95E-04	0.24	2.60	24	5	5	4	12	12	ENSGALG00000002971	NSD1	14.72	ENSGALT00000004699	INTRON
13	15977261	5.19E-04	2.70	0.29	26	9	9	3	13	13	ENSGALG00000000592	-	2.92	ENSGALT000000034019	INTRON
14	1136098	4.82E-04	2.50	0.09	9	3	3	0	10	10	-	-	-	-	INTERGENIC
14	6472859	2.20E-04	3.83	0.27	5	22	22	10	2	2	ENSGALG00000005878	MILST8	51.45	ENSGALT00000009446	UPSTREAM GENE
14	12699258	1.04E-04	0.06	2.67	0	15	15	7	2	2	ENSGALG00000005861	HEXDC	1.39	ENSGALT000000009423	UPSTREAM GENE
15	2820777	1.98E-07	2.85	0.39	56	19	19	14	37	37	ENSGALG00000007801	COR7	10.33	ENSGALT000000039543	INTRON
15	4948428	1.78E-07	5.80	0.36	4	28	28	27	9	9	ENSGALG00000025981	PAM16	39.32	ENSGALT00000012630	UPSTREAM GENE
15	6380069	2.43E-10	0.05	3.06	20	0	0	17	54	54	ENSGALG00000003307	TMED2	12.04	ENSGALT00000005224	UPSTREAM GENE
15	6380069	2.43E-10	0.05	3.06	20	0	0	17	54	54	ENSGALG00000004818	DDX55	12.04	ENSGALT00000005224	UPSTREAM GENE
15	6516240	6.88E-04	3.20	0.37	15	4	4	6	18	18	ENSGALG00000004869	RPL6	823.48	ENSGALT00000007692	INTRON
16	205646	1.60E-58	0.22	14.43	199	891	891	100	6	6	ENSGALG00000004821	ISCU	64.14	ENSGALT00000007773	INTRON
17	2382552	7.14E-04	0.33	11.00	8	2	2	0	10	10	ENSGALG00000004893	TMEM119	30.51	ENSGALT00000007813	DOWNSTREAM GENE
17	6208752	8.83E-05	4.25	0.17	16	3	3	1	11	11	ENSGALG00000003776	SART3	33.23	ENSGALT000000038897	UPSTREAM GENE
17	7602149	8.26E-05	0.40	3.33	11	29	29	19	5	5	ENSGALG00000001503	TBCD	24.35	ENSGALT000000039250	INTRON
18	3220361	9.91E-05	0.29	3.00	13	48	48	11	3	3	ENSGALG00000001503	TBCD	24.35	ENSGALT00000002283	INTRON
20	11987559	1.78E-05	0.29	4.50	5	20	20	17	3	3	ENSGALG0000000207653	RAE1	11.30	ENSGALT00000012397	INTRON
21	949389	6.40E-05	5.80	0.36	4	28	28	10	3	3	ENSGALG00000001006	TP73	3.27	ENSGALT00000004487	INTRON
26	895745	6.00E-11	8.40	0.39	41	4	4	15	40	40	ENSGALG00000000225	MGAT4C	2.46	ENSGALT000000040513	DOWNSTREAM GENE
27	2380730	1.23E-04	0.31	3.50	8	28	28	13	3	3	ENSGALG00000000226	TMEM9	5.33	ENSGALT00000000804	DOWNSTREAM GENE
27	4047630	8.49E-14	29.00	0.31	28	0	0	18	60	60	ENSGALG000000017850	TLK2	20.36	ENSGALT00000000559	3' UTR
28	2538432	1.43E-06	18.00	0.25	17	0	0	2	11	11	ENSGALG00000001634	SNORA21	112.61	ENSGALT00000002857	INTRON
											ENSGALG00000001235	RPL23	840.73	ENSGALT00000002496	INTRON
											ENSGALG00000001874	ANKRD24	1.00	ENSGALT000000001874	EXON

In bold, positions tested by pyrosequencing. Allelic ratios = allele 1 counts/allele 2 counts, for each cross.

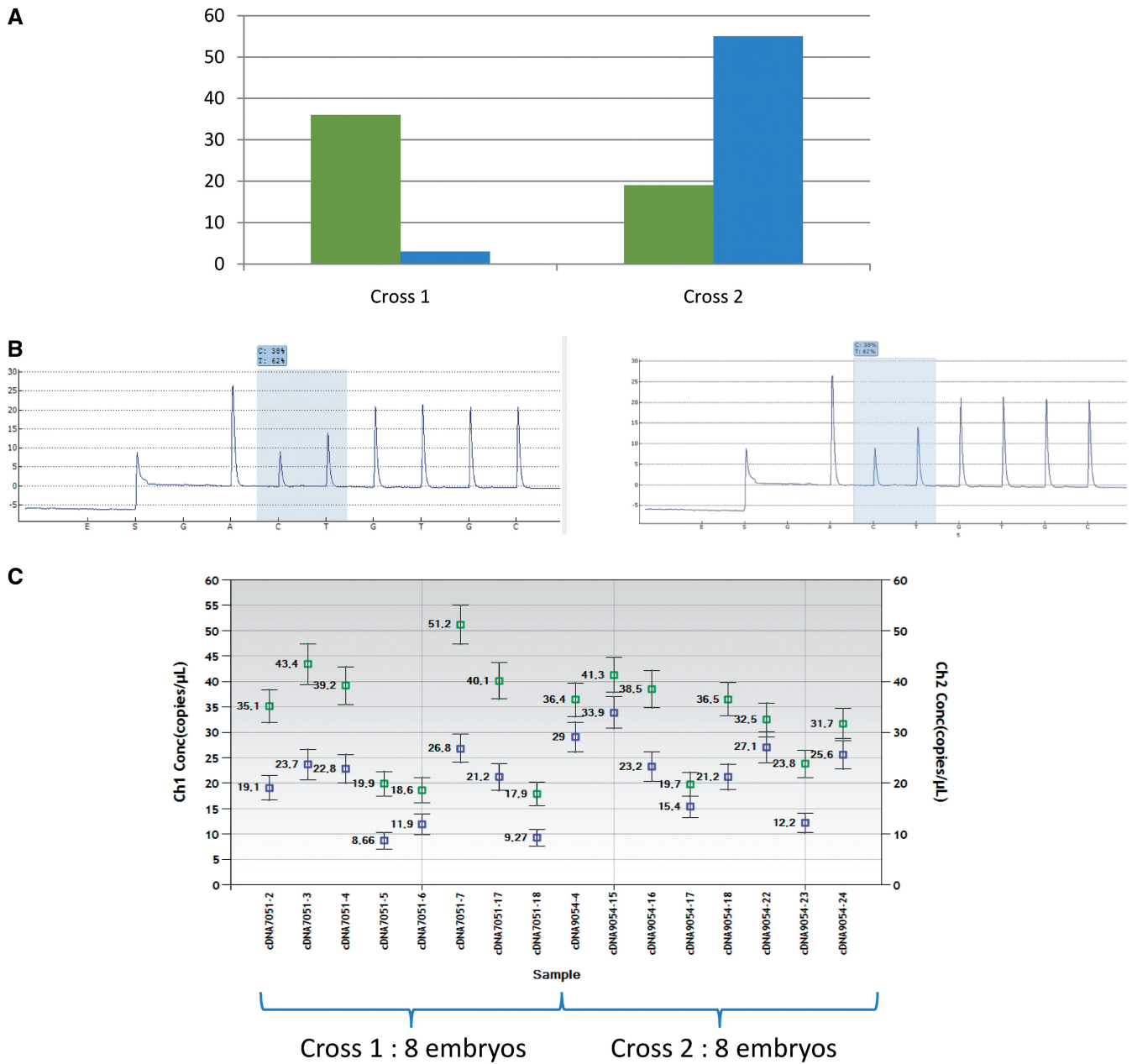


Figure 3. Comparison of results on a candidate gene obtained through HiSeq, droplet digital PCR and pyrosequencing. Colors are differentiating both alleles in sequencing results and ddPCR. (A) HiSeq counts of both alleles at the candidate locus (A allele in green, G allele in blue). (B) Pyrosequencing results from one sample of each cross (chosen as representative of average results). Analyses were made on the reverse strand. Allele A is overexpressed in both crosses (left: Cross 1 results from one embryo's cDNA, right: Cross 2 results from one embryo's cDNA). The highlighted peaks correspond to the SNP where the relative proportion is quantified. (C) Droplet digital PCR on the candidate locus from eight samples of each cross (A allele in green, G allele in blue).

DISCUSSION

This work represents, to our knowledge, the first transcriptome-wide investigation of genomic imprinting in birds. Previous studies in chicken were made on targeted genes, known to be imprinted in mammals, and they concluded to the nonexistence of this phenomenon in birds (15–18,20,53). To get free of any *a priori*, investigations should not be restricted to targeted loci but consider every single chicken gene. RNA sequencing technology is an approach meeting these expectations (54,55).

Transcriptome sequencing has already been used with this intention in other species, especially in mouse (32,33,54). It not only allowed the confirmation of most mammalian imprinted genes but also the discovery of putative new ones.

We thus set up an experimental cross designed for the search of parental genomic imprinting in chicken, by using RNA-Seq on whole embryos from reciprocal crosses between chicken lines, as inbred and genetically distant as possible. Our results identified 79 potentially imprinted

Table 2. Candidate position obtained from the mouse data set

Chromosome	Position	Gene short name	Localization	Imprinting status ^a	<i>P</i> -value Fisher exact test	Allelic ratio Cross 1	Allelic ratio Cross 2
1	24619956	Not annotated			1.55e-117	8.64	0.004
2	152530393	H13	Exon	Maternal	1.25e-62	3.91	0.30
4	60672634	Mup2 or Mup11	Exon	Unknown	1.06e-26	792	0.31
4	142807839	Not annotated			2.70e-05	3.83	0.34
4	144508797	Dhrs3 or AK147071	Intron	Unknown	4.44e-07	2.82	0.07
4	61746350	Mup3	Intron	Unknown	1.22e-14	0.03	4
5	87427181	Not annotated			3.94e-26	0.31	103
7	31513450	Gapdhs or Tmem147	Intron	Unknown	6.15e-15	19	0.13
7	67133548	Snurf	Exon	Paternal	5.64e-32	85	0.03
7	67131688	Snurf	Exon	Paternal	3.36e-40	0.01	47
11	83318343	Taf15	Exon	Unknown	4.04e-08	10.5	0.37
11	98911391	Igfbp4	Intron	Unknown	1.75e-55	5.11	0.01
11	98911393	Igfbp4	Intron	Unknown	5.69e-73	0.12	90
12	105186516	Serpina1e	Intron	Unknown	1.86e-13	0.05	14
14	22266157	Adk	Intron	Unknown	2.17e-05	0.07	2.8
14	30847794	Chdh	intron	Unknown	1.33e-13	0.02	10
15	72637471	Peg13	Exon	Paternal	1.77e-09	0.06	18
15	72640049	Peg13	Exon	Paternal	1.11e-21	0.02	18
17	35534271	Bat1a	Intron	Unknown	4.63e-12	0.37	4.1
19	46726575	D19Wsu16 or Opal1	Intron	Unknown	3.33e-05	0.09	3.4

^aKnown imprinting status in mouse in the literature.

In bold, genes known to be imprinted in the literature.

Allelic ratios = allele 1 counts/allele 2 counts, for each cross.

SNPs, mainly located in introns from distinct genes, not shown to be imprinted when analyzing their coding portion. It has been demonstrated that imprinted retrogenes could be hosted in introns of biallelically expressed genes in mouse (56), leading us to consider these loci as possibly interesting. But in the presence of genes truly imprinted, one could expect a higher number of candidate SNPs detected in known coding regions, and these results, with a large proportion of SNPs outside the coding regions, strengthened the need of experimental validations of these results.

These positions out of the coding sequence raise the question of the functional significance of the transcribed portion of the genome. There is a strong debate in the literature about the proportion of the genome that is 'functional' (57–60), following the ENCODE results, which assigned biochemical functions to 80.4% of the human genome, for which the cumulative coverage is 62.1% for processed transcripts (61). In our analysis, only ~40% of the genome is considered to be transcribed, with a minimal coverage of 10 reads in the pool of all 20 embryos data at a single-base scale. This discrepancy may be partly explained by our selection of only polyadenylated mRNA. Nevertheless, the transcribed portion of the genome far exceeds its exonic subset, a more and more admitted characteristic of the genome (62,63).

Before validation, we ensured that candidate SNPs were not declared biased because of only one highly expressed embryo (Supplementary Figure S3). None of the candidate loci tested through validation experiments was confirmed as imprinted. These candidates being probably false positives, we thus validated our methodology by reanalyzing a mouse data set, as a 'positive control',

where imprinting should be detected. By confirming the three previously identified imprinted genes [*H13*, *Peg13* and *Sgce* (40)], and by discovering one more in these data, *Snurf*, previously known to be imprinted (52), we confirmed the reliability of the method. These four genes found imprinted in mouse are a severe underestimate of the total set of genes known to be imprinted in this species. This low number of genes is discussed in the original study by Lagarrigue and coworkers (40), notably highlighting the fact that only a fraction of imprinted genes are expressed in a given tissue at a precise developmental stage. Their study was the first, to our knowledge, to examine the parent-of-origin specific gene expression in adult liver in mouse, and there may be few imprinted genes in this tissue. The amount of data is probably the main reason for this low number of detected genes. Our chicken data set contains 10 times more sequences aligned and analyzed than the mouse data set. Considering that the mouse has a slightly higher number of annotated genes, expression can be better estimated in our chicken RNA-Seq sequences. With a higher sequencing depth, more genes could potentially have been detected. Nevertheless, two analysis methods from the same data set lead to the detection of imprinted genes in mouse. The differences between the results obtained here (four genes) and in the previous analysis (three genes) from the same data have two main origins: first, the minimal depth was originally set to 10 for each allele, which discarded *Snurf*, almost totally imprinted, as a candidate locus. Moreover, the original study did not take into account the noncoding region (explaining the discovery of only exonic imprinted SNPs in the first analysis). The biological reality of the newly discovered loci has to be determined, but this analysis confirmed that imprinting,

when existing, is detected with our method. This validation is reinforced by the fact that we analyzed 200 Gb of chicken sequence (versus 3 Gb for the mouse study), leading to the assumption that we should have observed significant results if imprinting did exist in chicken. On the other hand, the use of whole embryos in our study (versus only liver samples in mouse) might have affected the final results. But our choice was, for the same amount of data, to reach a sufficient depth to reliably discover multitissue parental imprinting, without having to *a priori* select one single tissue. The choice of the embryo was determined by the high number of imprinted genes already known to be imprinted in the embryo, compared with single adult tissues (64). Moreover, others studies conducted on the whole embryo allowed the detection of imprinting in mouse (65). If imprinting does exist in chicken at this developmental stage, but is observed only in one tissue, or if a locus is paternally expressed in some tissues but maternally expressed in others (66), our device might have not allowed to detect such cases. While such profiles were already observed, most genes are imprinted in multiple tissue types (67). Nevertheless, it would be interesting to confirm the absence of imprinting in chicken by studying other tissues, extraembryonic membranes in particular: most of the known imprinted genes in mammals are expressed in the placenta (68,69), notably regulating nutritional resources to the embryo during development (70,71).

Another question concerning our model's choice was the embryonic stage of development. Stage 26 is a developmental stage from which a sufficient quantity of RNA can be extracted for RNA-Seq and validation experiments, and the expression of imprinted genes has already been described at the equivalent stage in the mouse embryo (64). It would also have been interesting to check the imprinting status of genes expressed in adult brain: by applying the kinship theory to a social context, the existence of imprinting would be possible in a species showing maternal care, as previously proposed (72). Incidentally, several genes have been shown to be imprinted in the mouse brain, notably related to social behavior (73). But the best bird lines to study would have been in this case from species with strong parental care behavior, and inbred chicken lines, despite their suitability for the RNA-Seq strategy used, are probably not the best model in this area.

Besides practical reasons as availability of inbred lines and of a reliable genome assembly, the choice of the chicken for this study has been dictated by several arguments: parent-of-origin QTLs have been detected in this species (27), and chicken is an example of farm animals used in crossbreeding, where genomic imprinting could be important for the utilization of reciprocal effects. It is also an important animal model in developmental biology and phylogenetics, and is used as a model for other avian species, especially in genetics and genomics (47).

A last comment about the device can be made about the sequenced transcripts: we limited the screening to polyA RNA, and thus we could not conclude on the potential imprinted status of noncoding RNAs, some of which are also known to be imprinted in mammals (74).

With a reliable analysis method but no evidence of imprinting in our chicken data set, our results seem to confirm the absence of imprinting in chicken, at least in the whole embryo at embryonic day 4.5. This analysis is a first step in the genome-wide search of genomic imprinting in birds. The absence of this phenomenon will be definitely confirmed—taking into account the difficulty of claiming the absence of such a multifarious mechanism—only when other studies will complete this first one: other transcribed parts of the genome (such as non-coding RNA), other tissues at other developmental stages must be analyzed, as well as other chicken lines and, most importantly, other bird species. But given previous work on several candidate genes (15–21), and absence of differential methylated regions according to the parental allele in the chicken genome (75), it gives a strong argument, through a genome-wide study, in agreement with previous studies. This absence would be in accordance with the most shared hypothesis, the kinship theory, which leads to the assumption that imprinting does not exist in oviparous species (10). Another theory concerning the evolution of genomic imprinting mechanisms is the 'host defence' hypothesis, assuming that imprinting appeared consequently to mechanisms aimed at silencing foreign DNA elements inserted into the genome, and evolved in parallel with accumulation of repeats in certain genomic regions (12,70,76,77). This is again not in favor of the existence of this phenomenon in chicken, a species with a low level of repeats and active transposable elements, compared with mammals (78).

The absence of this phenomenon would lead to alternative explanations of reciprocal effects. These effects represent ~20% of the phenotypic variability of body weight in broiler and of egg viability in layers (79), and one cause could have been genomic imprinting. But other mechanisms can lead to such effects, as direct maternal effects, sex-linked genes or mitochondrial DNA transmission. Regarding QTLs detected as showing parent-of-origin effects, most of them can result from biases due to the animal device, as previously pointed out (25,26,80). Rowe and coworkers, in a paper describing a maternally expressed QTL obtained in a pedigree avoiding such biases, comment their results by stating they 'have insufficient evidence to confirm that these are truly imprinted effects'. Taking into account the historical discussion about the absence of genomic imprinting in chicken and our results, it is possible that no QTL truly showing parent-of-origin effect does exist in chicken.

To understand why these false imprinted candidates were detected, we tried to investigate the possible biases and issues encountered during the analysis. This is not a unique situation in RNA sequencing data analysis, as some of the results from imprinting analysis through transcriptome sequencing had also been contested in mouse (81,82), highlighting many possible biases inherent to RNA sequencing data, from experiments to analyses. In particular, some alleles can be underrepresented through library preparation, several SNPs can appear to be artificial and read alignment can be biased due to a better mapping of the reference allele (83). Because we were looking for inverted allelic ratios between reciprocal

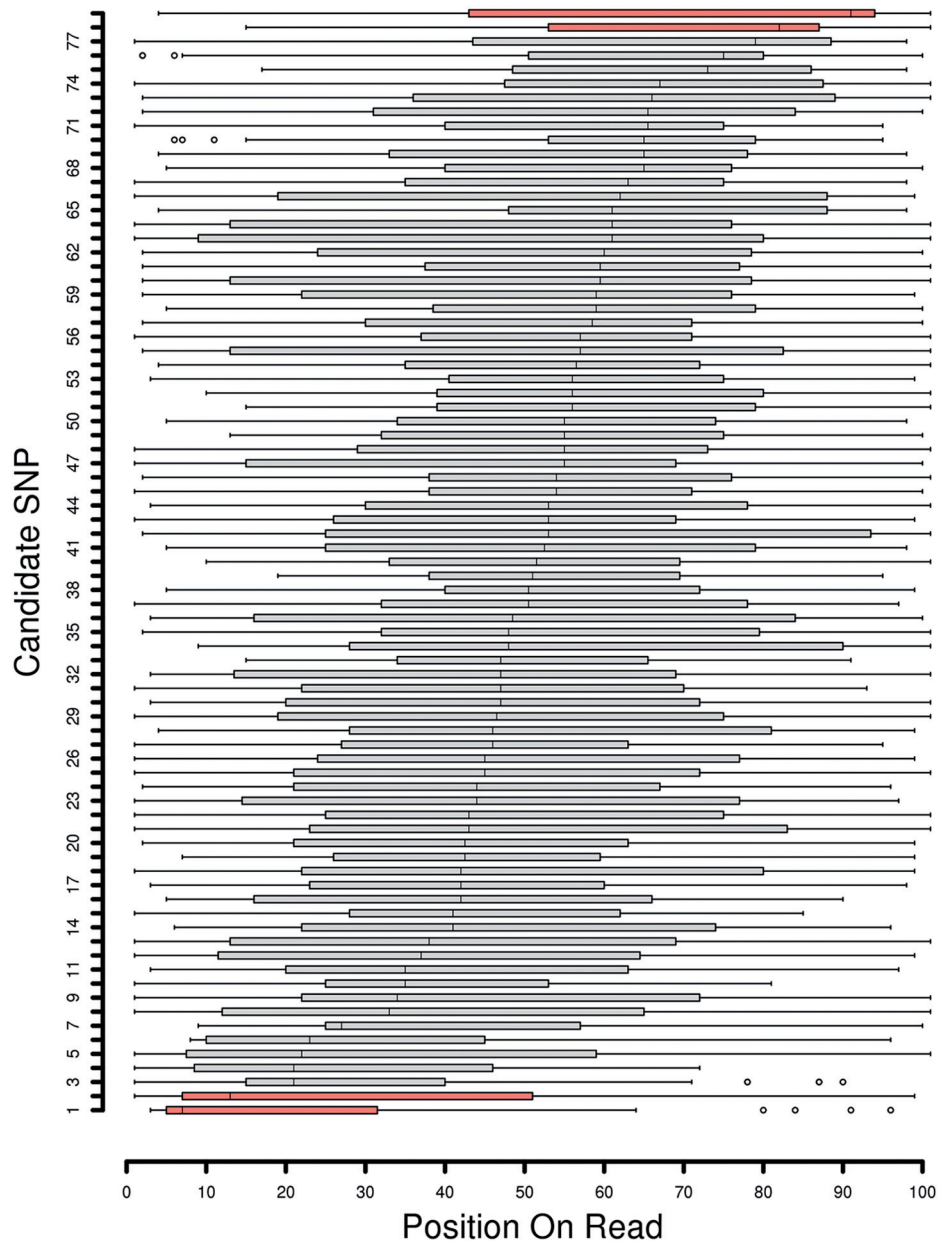


Figure 4. Candidate SNP positions among sequenced reads. Boxplot of SNP localization among reads for each candidate ordered by median. Orange color represents candidates with a median position in the 20% extremities of reads.

crosses in our study, this last issue could not explain our results because regions where the reference allele was preferentially mapped were avoided. We considered that the possible problem of artifactual SNPs was also absent from our data because we crossed the SNPs observed from the RNA-Seq data with the polymorphisms detected between the parents by genomic sequencing. Furthermore, all the candidate SNPs analyzed through other methods, as Sanger sequencing (data not shown) or pyrosequencing, confirmed the heterozygous status of the embryo's DNA. The possibility of underrepresentation of an allele due to the library preparation remains, even if less likely due to the use of reciprocal crosses: to generate such a bias, an allele should be 'randomly' retained in one cross, while the

alternative allele should be 'randomly' kept in the other cross during the library preparations.

Likewise, there is a possible dependence between nucleotide frequency at a SNP and its position on sequence reads (84). This bias can originate from the sequencing platform or from a positional bias of the reads across the transcripts. In any case, correcting this bias can improve expression estimates (85). All 79 positions were studied for this putative bias, but only four of them do have a median position among reads in the 20% extremities (Figure 4). The sequencing depth may be another cause of the median deviation from the theoretical value of 50 (average position on the read). While location of a SNP on a read may randomly vary, lower expression

level increases the probability of the median position of the SNP differing from this medium position. By removing candidate SNPs being in extremities of reads, we either remove low expressed loci or effectively biased loci. Excluding these four candidates is thus a conservative choice. But the unconfirmed allelic imbalance of the 75 left ones cannot be explained this way. Moreover, if we consider the proportion of forward and reverse reads for each candidate, it appears that no overrepresentation of one direction of sequencing was detected (no candidate SNP with a proportion of forward read $<20\%$ or $>80\%$). This potential bias was thus not the source of false positive in our case.

If the discovery of false-positive SNPs does not directly come from alignment or position bias, it can also be linked with analysis criteria. One of the main issues in such studies is the sequencing depth threshold to declare allelic imbalance. Studies made until now used different threshold in their analyses (32,33,86). This issue was underlined in many allelic imbalance studies using RNA sequences data, but until now, no consensus was made on a reliable threshold (32,86,87). We set ours to a minimum read depth of 10 in each cross, which corresponds to a minimal coverage of 20 reads for each locus considered in the analysis. This threshold was appropriate to highlight imprinted loci in the mouse data set. Thus, taking into account the lower sequencing depth in the mouse study compared with the chicken one, we should have detected imprinted loci in chicken if present: even a higher threshold would have brought to light such loci from our data set.

We thus looked at the candidates with regard to the general distribution of sequencing depth. Most of them have a low expression profile (Figure 5), which could lead to false-positive candidates. We can raise the hypothesis that at some low sequencing depth, i.e. at low expression levels, genes might express one allele preferentially, or that it is not technically possible to properly amplify both alleles when the gene is expressed under a certain limit. Under these conditions, the principal null hypothesis used in this study could be modified: instead of a 50:50 ratio between both alleles in transcripts of heterozygous embryos, the null hypothesis would depend on the expression rate of the genes. This hypothesis is in accordance with previous studies, highlighting the technical issues leading to allelic imbalance (86,88,89).

By looking to possible biases in our analysis, we tried to understand why no genomic imprinting could be observed on the candidate SNPs that had been highlighted. This leads to several considerations: firstly, results' validation is required through a reliable quantitative method, as we did mainly by pyrosequencing; secondly, possible experimental and analytical biases should be considered in the final results interpretation and care should be taken regarding general sequencing depth, possible alignment biases and threshold choices, as previously underlined (82,90).

This study was designed to detect parental genomic imprinting in birds, with chicken as a model species. Our results show that, at least at this developmental stage, there is no genomic region imprinted at the whole

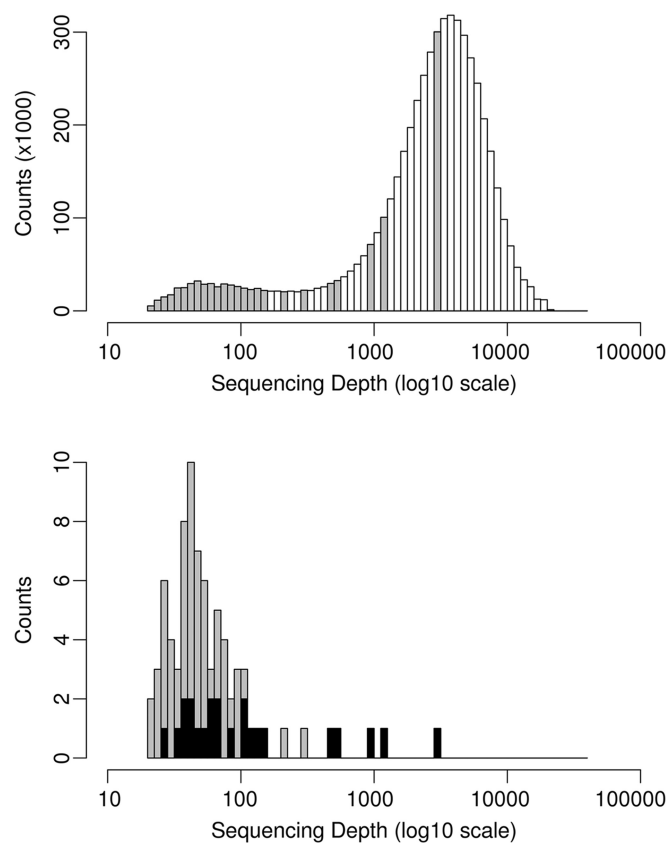


Figure 5. Sequencing depth distribution for all SNPs (top), with classes containing candidate SNPs in gray, and for candidate SNPs (bottom), with tested SNPs in black.

embryo level. This does not exclude specific events occurring in a small subset of tissues or at different developmental stages that would most probably not have been detectable in the combination of tissues used in this work. Nevertheless, this study proposes a performing method to investigate the existence of genomic imprinting at a genome scale, without *a priori* choice of targets genes known to be imprinted in mammals. Our results support the fact that, while the evolution of imprinting has occurred in a convergent manner for several clades, from plants to mammals, it is probably missing in birds.

ACCESSION NUMBERS

NCBI_SRA accessions number: SRP033603.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful for support by the PEAT team in managing animals. We thank GeT-PlaGe Genotoul platform for performing the HiSeq sequencing, and the

Bio-Rad team for their help in the ddPCR analyses. We are grateful to Déborah Bourc'his, Jérôme Cavailé, Hervé Acloque and Julie Demars for informal discussions throughout this work.

FUNDING

Région Midi-Pyrénées (<http://www.midipyrenees.fr/>) and Animal Genetics Division (INRA, <http://www.ga.inra.fr/en/>) (to L.F.); French 'Agence Nationale de la Recherche' EpiBird grant [ANR-009-GENM-004, <http://www.agence-nationale-recherche.fr/>]. Funding for open access charge: INRA, UMR444 Laboratoire de Génétique Cellulaire, Castanet-Tolosan, F-31326, France.

Conflict of interest statement. None declared.

REFERENCES

- Das, R., Anderson, N., Koran, M.I., Weidman, J.R., Mikkelsen, T.S., Kamal, M., Murphy, S.K., Linblad-Toh, K., Grealis, J.M. and Jirtle, R.L. (2012) Convergent and divergent evolution of genomic imprinting in the marsupial *Monodelphis domestica*. *BMC Genomics*, **13**, 394.
- da Rocha, S.T. and Ferguson-Smith, A.C. (2004) Genomic imprinting. *Curr. Biol.*, **14**, R646–R649.
- Holman, L. and Kokko, H. (2013) The evolution of genomic imprinting: costs, benefits and long-term consequences. *Biol. Rev. Camb. Philos. Soc.*, (doi: 10.1111/brv.12069; epub ahead of print, October 28, 2013).
- Haig, D. and Graham, C. (1991) Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell*, **64**, 1045–1046.
- Moore, T. and Haig, D. (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.*, **7**, 45–49.
- Haig, D. (1997) Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proc. Biol. Sci.*, **264**, 1657–1662.
- Haig, D. (2000) The kinship theory of genomic imprinting. *Ann. Rev. Ecol. Evol. Syst.*, **31**, 9–32.
- Ashbrook, D.G. and Hager, R. (2013) Empirical testing of hypotheses about the evolution of genomic imprinting in mammals. *Front. Neuroanat.*, **7**, 6.
- Haig, D. (2013) Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity (Edinb.)*, (doi: 10.1038/hdy.2013.97; epub ahead of print, October 16, 2013).
- Iwasa, Y. (1998) The conflict theory of genomic imprinting: how much can be explained? *Current Topics in Developmental Biology*, **40**, 255–293.
- Feil, R. and Berger, F. (2007) Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet.*, **23**, 192–199.
- Renfree, M.B., Suzuki, S. and Kaneko-Ishino, T. (2013) The origin and evolution of genomic imprinting and viviparity in mammals. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, **368**, 20120151.
- Giannoukakis, N., Deal, C., Paquette, J., Goodyer, C.G. and Polychronakos, C. (1993) Parental genomic imprinting of the human IGF2 gene. *Nat. Genet.*, **4**, 98–101.
- Koski, L.B., Sasaki, E., Roberts, R.D., Gibson, J. and Etches, R.J. (2000) Monoallelic transcription of the insulin-like growth factor-II gene (Igf2) in chick embryos. *Mol. Reprod. Dev.*, **56**, 345–352.
- Nolan, C.M., Killian, J.K., Petite, J.N. and Jirtle, R.L. (2001) Imprint status of M6P/IGF2R and IGF2 in chickens. *Dev. Genes Evol.*, **211**, 179–183.
- O'Neill, M.J., Ingram, R.S., Vrana, P.B. and Tilghman, S.M. (2000) Allelic expression of IGF2 in marsupials and birds. *Dev. Genes Evol.*, **210**, 18–20.
- Wang, G., Yan, B., Deng, X., Li, C., Hu, X. and Li, N. (2005) Insulin-like growth factor 2 as a candidate gene influencing growth and carcass traits and its biallelic expression in chicken. *Sci. China C Life Sci.*, **48**, 187–194.
- Yokomine, T., Kuroiwa, A., Tanaka, K., Tsudzuki, M., Matsuda, Y. and Sasaki, H. (2001) Sequence polymorphisms, allelic expression status and chromosome locations of the chicken IGF2 and MPR1 genes. *Cytogenet. Genome Res.*, **93**, 109–113.
- Yokomine, T., Shirohzu, H., Purbowasito, W., Toyoda, A., Iwama, H., Ikeo, K., Hori, T., Mizuno, S., Tsudzuki, M., Matsuda, Y. et al. (2005) Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian *Ascl2/Mash-Igf2-H19* region. *Genome Res.*, **15**, 154–165.
- Shin, S., Han, J.Y. and Lee, K. (2010) Cloning of avian Delta-like 1 homolog gene: the biallelic expression of Delta-like 1 homolog in avian species. *Poult. Sci.*, **89**, 948–955.
- Colosi, D.C., Martin, D., More, K. and Lalande, M. (2006) Genomic organization and allelic expression of UBE3A in chicken. *Gene*, **383**, 93–98.
- Minvielle, F., Kayang, B.B., Inoue-Murayama, M., Miwa, M., Vignal, A., Gourichon, D., Neau, A., Monvoisin, J.L. and Ito, S. (2005) Microsatellite mapping of QTL affecting growth, feed consumption, egg production, tonic immobility and body temperature of Japanese quail. *BMC Genomics*, **6**, 87.
- Tuiskula-Haavisto, M., De Koning, D.J., Honkatukia, M., Schulman, N.F., Mäki-Tanila, A. and Vilkki, J. (2004) Quantitative trait loci with parent-of-origin effects in chicken. *Genet. Res.*, **84**, 57–66.
- Tuiskula-Haavisto, M. and Vilkki, J. (2007) Parent-of-origin specific QTL—a possibility towards understanding reciprocal effects in chicken and the origin of imprinting. *Cytogenet. Genome Res.*, **117**, 305–312.
- de Koning, D.J., Bovenhuis, H. and van Arendonk, J.A. (2002) On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. *Genetics*, **161**, 931–938.
- Sandor, C. and Georges, M. (2008) On the detection of imprinted quantitative trait loci in line crosses: effect of linkage disequilibrium. *Genetics*, **180**, 1167–1175.
- Rowe, S.J., Pong-Wong, R., Haley, C.S., Knott, S.A. and de Koning, D.J. (2009) Detecting parent of origin and dominant QTL in a two-generation commercial poultry pedigree using variance component methodology. *Genet. Sel. Evol.*, **41**, 6.
- Yokomine, T., Hata, K., Tsudzuki, M. and Sasaki, H. (2006) Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenet. Genome Res.*, **113**, 75–80.
- Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R. and Reik, W. (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat. Genet.*, **36**, 1291–1295.
- Peters, J. and Williamson, C.M. (2007) Control of imprinting at the Gnas cluster. *Epigenetics*, **2**, 207–213.
- Frésard, L., Morisson, M., Brun, J.M., Collin, A., Pain, B., Minvielle, F. and Pitel, F. (2013) Epigenetics and phenotypic variability: some interesting insights from birds. *Genet. Sel. Evol.*, **45**, 16.
- Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D. and Dulac, C. (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*, **329**, 643–648.
- Wang, X., Soloway, P.D. and Clark, A.G. (2011) A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics*, **189**, 109–122.
- Wang, X., Miller, D.C., Harman, R., Antczak, D.F. and Clark, A.G. (2013) Paternally expressed genes predominate in the placenta. *Proc. Natl Acad. Sci. USA*, **110**, 10705–10710.
- Groenen, M.A., Megens, H.J., Zare, Y., Warren, W.C., Hillier, L.W., Crooijmans, R.P., Vereijken, A., Okimoto, R., Muir, W.M. and Cheng, H.H. (2011) The development and characterization of a 60K SNP chip for chicken. *BMC Genomics*, **12**, 274.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J. Hum. Genet.*, **81**, 559–575.
- Bumstead, N. and Barrow, P.A. (1988) Genetics of resistance to *Salmonella typhimurium* in newly hatched chicks. *Br. Poult. Sci.*, **29**, 521–529.

38. Bordas, A., Tixier-Boichard, M. and Merat, P. (1992) Direct and correlated responses to divergent selection for residual food intake in Rhode Island Red laying hens. *Br. Poult. Sci.*, **33**, 741–754.
39. Roussot, O., Feve, K., Plisson-Petit, F., Pitel, F., Faure, J.M., Beaumont, C. and Vignal, A. (2003) AFLP linkage map of the Japanese quail *Coturnix japonica*. *Genet. Sel. Evol.*, **35**, 559–572.
40. Lagarrigue, S., Martin, L.J., Hormozdiari, F., Roux, P.F., Pan, C., van Nas, A., Demeure, O., Cantor, R., Ghazalpour, A., Eskin, E. *et al.* (2013) Analysis of allele specific expression in mouse liver by RNA-seq: a comparison with “Cis”-eQTL identified using genetic linkage. *Genetics*, **195**, 1157–1166.
41. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
42. McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, **26**, 2069–2070.
43. Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
44. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, **7**, 562–578.
45. Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M. *et al.* (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, **477**, 289–294.
46. Dabney, A. and Storey, J.D. (2009) qvalue: Q-value estimation for false discovery rate control, R package, v1.24.0 ed. <http://CRAN.R-project.org/package=qvalue>.
47. Burt, D.W. and White, S.J. (2007) Avian genomics in the 21st century. *Cytogenet. Genome Res.*, **117**, 6–13.
48. Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P. *et al.* (2013) Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*, **14**, 59.
49. Sun, A., Ge, J., Siffert, W. and Frey, U.H. (2005) Quantification of allele-specific G-protein beta3 subunit mRNA transcripts in different human cells and tissues by Pyrosequencing. *Euro. J. Hum. Genet.*, **13**, 361–369.
50. Wang, H. and Elbein, S.C. (2007) Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol. Biol.*, **373**, 157–176.
51. Smith, R.J., Dean, W., Konfortova, G. and Kelsey, G. (2003) Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res.*, **13**, 558–569.
52. Le Meur, E., Watrin, F., Landers, M., Sturny, R., Lalande, M. and Muscatelli, F. (2005) Dynamic developmental regulation of the large non-coding RNA associated with the mouse 7C imprinted chromosomal region. *Dev. Biol.*, **286**, 587–600.
53. Yokomine, T., Shirohzu, H., Purbowasito, W., Toyoda, A., Iwama, H., Ikeo, K., Hori, T., Mizuno, S., Tsudzuki, M., Matsuda, Y. *et al.* (2005) Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian *Ascl2/Mash2-Igf2-H19* region. *Genome Res.*, **15**, 154–165.
54. Babak, T., DeVeale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M. and Lim, L.P. (2008) Global survey of genomic imprinting by transcriptome sequencing. *Curr. Biol.*, **18**, 1735–1741.
55. Babak, T. (2012) Identification of imprinted Loci by transcriptome sequencing. *Methods Mol. Biol.*, **925**, 79–88.
56. Monk, D., Arnaud, P., Frost, J.M., Wood, A.J., Cowley, M., Martin-Trujillo, A., Guillaumet-Adkins, A., Iglesias Platas, I., Camprubi, C., Bourc’his, D. *et al.* (2011) Human imprinted retrogenes exhibit non-canonical imprint chromatin signatures and reside in non-imprinted host genes. *Nucleic Acids Res.*, **39**, 4577–4586.
57. Doolittle, W.F. (2013) Is junk DNA bunk? A critique of ENCODE. *Proc. Natl Acad. Sci. USA*, **110**, 5294–5300.
58. Graur, D., Zheng, Y., Price, N., Azevedo, R.B., Zufall, R.A. and Elhaik, E. (2013) On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.*, **5**, 578–590.
59. Niu, D.K. and Jiang, L. (2013) Can ENCODE tell us how much junk DNA we carry in our genome? *Biochem. Biophys. Res. Commun.*, **430**, 1340–1343.
60. Eddy, S.R. (2013) The ENCODE project: missteps overshadowing a success. *Curr. Biol.*, **23**, R259–R261.
61. The ENCODE Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
62. Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A. *et al.* (2011) The reality of pervasive transcription. *PLoS Biol.*, **9**, e1000625; discussion e1001102.
63. Stamatoyannopoulos, J.A. (2012) What does our genome encode? *Genome Res.*, **22**, 1602–1611.
64. Schulz, R., Woodfine, K., Menhenniott, T.R., Bourc’his, D., Bestor, T. and Oakey, R.J. (2008) WAMIDEX: a web atlas of murine genomic imprinting and differential expression. *Epigenetics*, **3**, 89–96.
65. Schulz, R., Menhenniott, T.R., Woodfine, K., Wood, A.J., Choi, J.D. and Oakey, R.J. (2006) Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. *Nucleic Acids Res.*, **34**, e88.
66. Garfield, A.S., Cowley, M., Smith, F.M., Moorwood, K., Stewart-Cox, J.E., Gilroy, K., Baker, S., Xia, J., Dalley, J.W., Hurst, L.D. *et al.* (2011) Distinct physiological and behavioural functions for parental alleles of imprinted *Grb10*. *Nature*, **469**, 534–538.
67. Prickett, A.R. and Oakey, R.J. (2012) A survey of tissue-specific genomic imprinting in mammals. *Mol. Genet. Genomics*, **287**, 621–630.
68. Barbaux, S., Gascoin-Lachambre, G., Buffat, C., Monnier, P., Mondon, F., Tonanny, M.B., Pinard, A., Auer, J., Bessieres, B., Barlier, A. *et al.* (2012) A genome-wide approach reveals novel imprinted genes expressed in the human placenta. *Epigenetics*, **7**, 1079–1090.
69. Reik, W. and Walter, J. (2001) Genomic imprinting: parental influence on the genome. *Nat. Rev.*, **2**, 21–32.
70. Ferguson-Smith, A.C. (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev.*, **12**
71. Gutierrez-Marcos, J.F., Constancia, M. and Burton, G.J. (2012) Maternal to offspring resource allocation in plants and mammals. *Placenta*, **33(Suppl. 2)**, e3–e10.
72. Ubeda, F. and Gardner, A. (2010) A model for genomic imprinting in the social brain: juveniles. *Evolution*, **64**, 2587–2600.
73. Curley, J.P. (2011) Is there a genomically imprinted social brain? *Bioessays*, **33**, 662–668.
74. Peters, J. and Robson, J.E. (2008) Imprinted noncoding RNAs. *Mamm. Genome*, **19**, 493–502.
75. Li, Q., Li, N., Hu, X., Li, J., Du, Z., Chen, L., Yin, G., Duan, J., Zhang, H., Zhao, Y. *et al.* (2011) Genome-wide mapping of DNA methylation in chicken. *PLoS One*, **6**, e19428.
76. Barlow, D.P. (1993) Methylation and imprinting: from host defense to gene regulation? *Science*, **260**, 309–310.
77. Renfree, M.B., Hore, T.A., Shaw, G., Marshall Graves, J.A. and Pask, A.J. (2009) Evolution of genomic imprinting: insights from marsupials and monotremes. *Annu. Rev. Genomics Hum. Genet.*, **10**, 241–262.
78. Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E. *et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695–716.
79. Fairfull, R.W. (1990) In: Crawford, R.D. (ed.), *Poultry Breeding and Genetics*. Elsevier Science Publishers, Amsterdam, pp. 913–933.
80. Lawson, H.A., Cheverud, J.M. and Wolf, J.B. (2013) Genomic imprinting and parent-of-origin effects on complex traits. *Nat. Rev. Genet.*, **14**, 609–617.
81. Kelsey, G. and Bartolomei, M.S. (2012) Imprinted Genes... and the Number Is? *PLoS Genetics*, **8**, e1002601.

82. DeVeale, B., van der Kooy, D. and Babak, T. (2012) Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.*, **8**, e1002600.
83. Degner, J.F., Marioni, J.C., Pai, A.A., Pickrell, J.K., Nkadori, E., Gilad, Y. and Pritchard, J.K. (2009) Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, **25**, 3207–3212.
84. Hansen, K.D., Brenner, S.E. and Dudoit, S. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.*, **38**, e131.
85. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L. and Pachter, L. (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.*, **12**, R22.
86. Heap, G.A., Yang, J.H., Downes, K., Healy, B.C., Hunt, K.A., Bockett, N., Franke, L., Dubois, P.C., Mein, C.A., Dobson, R.J. *et al.* (2010) Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum. Mol. Genet.*, **19**, 122–134.
87. Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D. and Clark, A.G. (2008) Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One*, **3**, e3839.
88. Heinrich, V., Stange, J., Dickhaus, T., Imkeller, P., Kruger, U., Bauer, S., Mundlos, S., Robinson, P.N., Hecht, J. and Krawitz, P.M. (2012) The allele distribution in next-generation sequencing data sets is accurately described as the result of a stochastic branching process. *Nucleic Acids Res.*, **40**, 2426–2431.
89. Nothnagel, M., Wolf, A., Herrmann, A., Szafranski, K., Vater, I., Brosch, M., Huse, K., Siebert, R., Platzer, M., Hampe, J. *et al.* (2011) Statistical inference of allelic imbalance from transcriptome data. *Hum. Mutat.*, **32**, 98–106.
90. Coolon, J.D., Stevenson, K.R., McManus, C.J., Graveley, B.R. and Wittkopp, P.J. (2012) Genomic imprinting absent in *Drosophila melanogaster* adult females. *Cell Rep.*, **2**, 69–75.