

Identification and resolution of artifacts in the interpretation of imprinted gene expression

Charlotte Proudhon and Déborah Bourc'his

Advance Access publication date 8 September 2010

Abstract

Genomic imprinting refers to genes that are epigenetically programmed in the germline to express exclusively or preferentially one allele in a parent-of-origin manner. Expression-based genome-wide screening for the identification of imprinted genes has failed to uncover a significant number of new imprinted genes, probably because of the high tissue- and developmental-stage specificity of imprinted gene expression. A very large number of technical and biological artifacts can also lead to the erroneous evidence of imprinted gene expression. In this article, we focus on three common sources of potential confounding effects: (i) random monoallelic expression in monoclonal cell populations, (ii) genetically determined monoallelic expression and (iii) contamination or infiltration of embryonic tissues with maternal material. This last situation specifically applies to genes that occur as maternally expressed in the placenta. Beside the use of reciprocal crosses that are instrumental to confirm the parental specificity of expression, we provide additional methods for the detection and elimination of these situations that can be misinterpreted as cases of imprinted expression.

Keywords: *genomic imprinting; DNA methylation; monoallelic expression; germline; placenta*

INTRODUCTION

Imprinted genes are defined as genes that are expressed from one of the two parental alleles in a parent-of-origin specific manner. The parental determinism of imprinted expression is strictly epigenetic. Differential methylation marks are established in the female and the male germlines by *de novo* DNA methyltransferases and their cofactor DNMT3L [1, 2], creating an epigenetic asymmetry at *cis*-regulatory elements referred to as imprinted control regions (ICRs). Parental differences in allelic methylation are subsequently maintained after fertilization and are joined by differential histone modifications to build an euchromatic state on the active allele and an heterochromatic state on the inactive allele. To date, ~130 imprinted genes mapping to 25 genomic regions have been identified in mouse and/or humans (http://www.har.mrc.ac.uk/research/genomic_imprinting). Current estimates of their

total number range from 100 to 2000 genes, depending on the prediction method [3].

Computational methods of prediction of imprinting use algorithms trained on the simple DNA features of known imprinted genes [4, 5], thereby ignoring the epigenetic determinism of imprinted expression. Experimental methods can be based on allelic assessment of expression or DNA methylation [6, 7], or in the recognition of regions with overlapping euchromatin and heterochromatin marks, respectively, H3 Lys4 trimethylation and H3 Lys9 trimethylation [8, 9]. Practically, most of the known imprinted genes were uncovered through the genome-wide transcriptome analysis of models of global or local imprinting anomalies, such as uniparental conceptuses (parthenogenotes and androgenotes) and mice carrying uniparental isodisomies for informative chromosomes, performed on microarrays [10, 11]. More recently, the sensitivity of

Corresponding author. Déborah Bourc'his, INSERM U934/UMR 3215, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France. Tel.: +33 1 56 24 68 25; Fax: +33 1 56 24 69 39; E-mail: deborah.bourchis@curie.fr

Déborah Bourc'his is a team leader at the Institut Curie in Paris, France. Her research interests are focused on the epigenetic regulation of reproduction in mammals.

Charlotte Proudhon is a PhD student in the group of Déborah Bourc'his.

Table 1: Recommended experimental procedures to avoid confounding effects in the identification or confirmation of imprinted genes

Artifact	Procedure
Random monoallelic expression rather than parent-specific expression	Use non-monoclonal cell populations
Allele-specific PCR amplification rather than allele-specific expression	Test unbiased primers on genomic DNA of hybrid animals or on 50:50 premixed RNA from the two backgrounds
Genetically determined allele-specific expression (a.k.a. strain bias or eQTL) rather than parent-specific expression	Use reciprocal hybrid crosses in mouse or opposite polymorphic combination in human studies
Maternally expressed placenta-only imprinted genes: contamination or infiltration with maternal cells rather than expression from the maternal allele	Confirm in backcrosses or perform embryo transfer in a foster mother with unrelated strain background

genome-wide transcription approaches was improved by the use of high-throughput sequencing, for the quantification of allelic ratios of heterozygous single nucleotide polymorphisms (SNPs) in RNA of reciprocal hybrid F1 mice [12, 13]. Screening for differentially methylated regions (DMRs) have also succeeded in identifying new imprinted genes [6], but they have not been combined yet with new sequencing technologies.

Difficulties in identifying imprinted genes reside in the intrinsic biological properties of these genes. Genomic imprinting is not an ‘all-or-none’ phenomenon, but occurs in a continuum from complete monoparental expression to slight but significant biased expression toward one parental allele [14]. These cases of more subtle preferential expression will be hardly detectable in poorly quantitative methods. Moreover, imprinted expression is often tissue- and/or stage-specific. Most of imprinted genes are expressed in a monoallelic manner before birth, in the developing embryo and the placenta, and after birth in the brain [15]. Although systematic studies have not been performed, it seems that imprinted genes tend to be turned off or to alternatively acquire a biallelic status of expression in other somatic tissues. Expression screening in non-relevant tissues and developmental stages will fail to detect imprinted genes, and unfortunately, most of the human studies are performed on easily accessible material such as lymphoblastoid cell lines [16]. In this regard, while high throughput screening performed on whole-mount neonatal brain did not uncover a significant number of new imprinted genes [13], careful dissection of brain sub-structures led to the prediction of ~1300 candidates, illustrating the very fine spatio-temporal regulation of imprinted expression [17].

While a high rate of false negatives is a well-appreciated drawback in imprinted gene screening or validation, false positives represent also a major issue in the robust identification of imprinted genes. A variety of mechanisms can lead to biased expression of one allele, ranging from technical problems to biologically determined reasons. In this review, we address the most commonly encountered artifacts that can be erroneously taken as evidence for imprinting, and provide guidelines to detect and avoid such confounding effects (Table 1). We focus on three main sources of imprinting misinterpretation: (i) random monoallelic expression in monoclonal cell populations, (ii) genetically determined monoallelic expression and (iii) contamination by maternal cells. Rigorous testing of the parental origin of expression in reciprocal F1 crosses can eliminate the confusing effects of random and genetically determined monoallelic expression in mouse. Complex patterns of expression will, however, occur in cases where allele-specific expression (ASE) is under the dual control of genomic imprinting and genetic polymorphisms. Problems of maternal contamination specifically apply to those genes that are only found imprinted in the placenta and have a maternal-specific pattern of expression. We propose here some experimental procedures to detect this type of artifact and significantly eliminate false positives.

MONOALLELIC EXPRESSION: PARENT-SPECIFIC OR STOCHASTIC?

Imprinted genes are characterized by the preferential to exclusive expression of one of the two parental alleles, always the same. By contrast, random

monoallelic expression describes a situation where some cells transcribe one allele of a gene, whereas other cells transcribe the other allele [18]. Imprinted and randomly expressed genes share a number of epigenetic characteristics. Once determined, the allelic choice is maintained and stably propagated throughout cell divisions. Differential DNA methylation, histone modifications and replication timing play a key role in distinguishing the active and the inactive alleles. Moreover, imprinted genes and randomly expressed genes share a restriction of H3 Lys4 dimethylation on the promoter of the active allele, while this mark tends to spread into the body of biallelically expressed genes [19]. However, randomly expressed genes will appear as biallelically expressed when cell populations with different allelic choices are analyzed, while one parental allele will be clearly overrepresented in cases of imprinted genes in such mixed-cell populations.

A paradigm of random monoallelic expression concerns X-linked genes subject to dosage compensation in mammalian females. In this case, a cross-talk between the two X chromosomes at the time of embryonic differentiation breaks the symmetry between the two chromosomes [20]. This leads to one of the chromosome to upregulate the inactivation determining Xist RNA, which induces *in cis* the silencing of linked genes. Random monoallelic expression also exists for autosomal genes and includes olfactory and pheromone receptor genes, as well as various immune response genes (immunoglobulins, interleukins, Toll-like and NK receptors) [21–25]. In the case of interleukins, the genes can invariably be expressed from one allele or the other or even from both alleles simultaneously, providing some inter- and intra-individual variability in immune response.

Recently, random monoallelic expression was shown to be more widespread than originally thought and not restricted to genes involved in the nervous or immune systems. Because stochastic inactivation results in two simultaneous patterns of expression in a mixture of cells, one way to detect random monoallelic is to study clonal cell populations. A SNP study of ~4000 genes in human monoclonal lymphoblastoids derived by single-cell cloning demonstrated that 10% of them could be indistinctively expressed from one allele or the other [26]. Conservative extrapolation to the whole genome would suggest that ~1000 human genes may be subject to random monoallelic expression and would affect a wide variety of molecular

functions. Some of the genes identified in lymphoblastoids were also monoallelically expressed in fibroblast cell subclones and in small homogeneous patches of apparently clonally derived tissue from the placenta. Screening of clonal neural stem cells derived from mouse F1 hybrids also evidenced that at least five genes could be monoallelically expressed in the murine central nervous system [27].

In clonal cell populations, particularly the ones induced by extreme culture conditions, one cannot exclude that the observed monoallelic expression is not biologically genuine but rather due to some genetic or epigenetic drift. More relevant to our issue, randomly expressed genes become indistinguishable from imprinted genes when studying clonal cell populations, as both will show monoallelic expression or DNA methylation (Figure 1A). Herein, confounding effects may be encountered in any situation leading to pauci- and monoclonality. Epstein–Barr virus (EBV) transformation in itself was shown to reduce cell population diversity and to rapidly lead to monoclonality in the process of lymphoblastoid derivation [28]. About 20% of all donor lymphoblastoid cell lines may be affected by this process. On a whole tissue-scale, rapid cell-number expansion from a limited pool of progenitor cells as well as limited cell migration during tissue formation will also lead to clonal cell patches, as it is for the placenta [29]. The allelic expression analysis of different combinations of parental alleles can circumvent this confounding effect. But while this is usually done in mouse by rigorous testing of reciprocal crosses of polymorphic strains, alternate heterozygous samples may not always be available in human studies. An additional issue when using human samples is that the pedigree and therefore the parental origin of each allele is not necessarily known. Finally, it should be pointed here that the use of monoclonal cell lines is also a source of confounding effects between a true random monoallelic expression, such as the one X-linked genes undergo, and a monoallelic expression determined by genetic differences between the two alleles (see next paragraph).

In conclusion, expression or methylation data from transformed cells, micropatches of placenta and, more generally, any cell lines or tissues that are clonally restricted, are not well suited for the screening or confirmation of imprinted genes. This is particularly well illustrated in a recent genome-wide survey of differential allelic expression in human performed on genotyping SNP-microarrays.

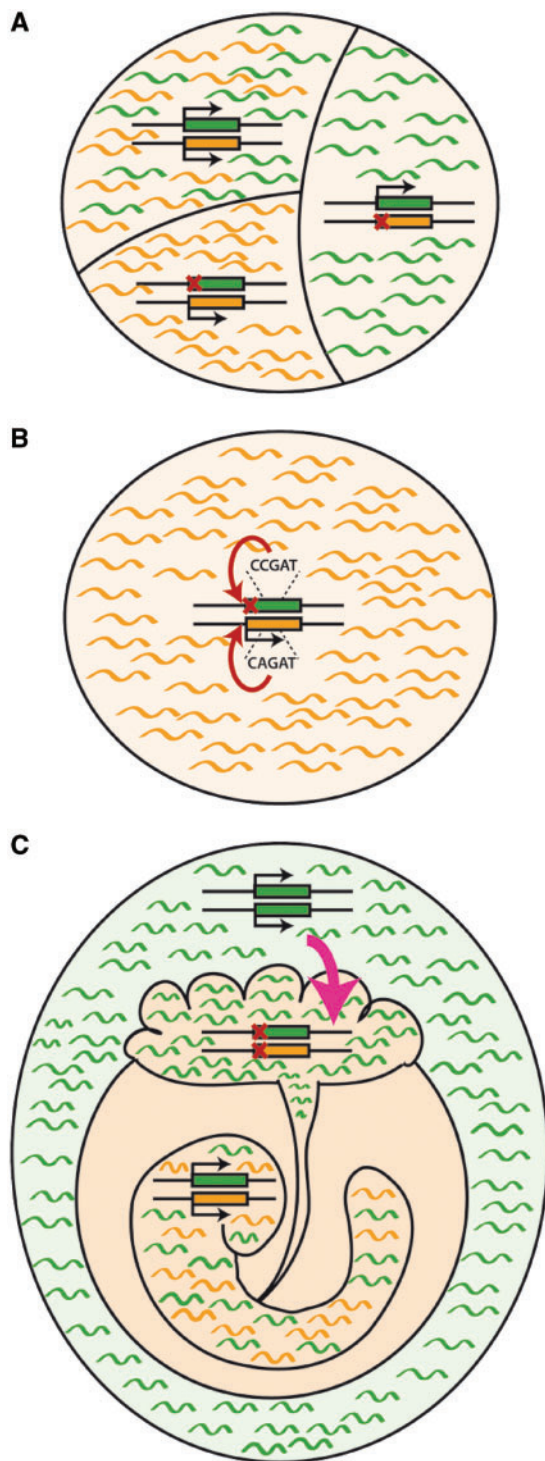


Figure 1: Three types of artifacts that can lead to a false interpretation of imprinting. All the situations are represented in the context of heterozygous samples, in which the origin of each transcript can be traced by expressed SNPs. **(A)** Random monoallelic expression in monoclonal cell populations or tissue micropatches. Some patches will express the green allele, others the orange one and some others the combination of the two alleles. **(B)** Genetically determined allele-specific

Prediction of imprinting was highly inconsistent among lymphoblastoid cell lines [16]. Moreover, analysis of non-transformed primary cell types confirmed that these genes were subject to random inactivation rather than imprinting. To avoid confusing effects, the clonal status of the cell line-based material should be verified (by investigation the expression of X-linked genes or immunoglobulin heavy-chain rearrangements, when possible), or better, bulk, non-cell line, *ex vivo* material should be preferred for the identification of imprinted genes (Table 1).

MONOALLELIC EXPRESSION: EPIGENETIC OR GENETIC DETERMINISM?

Imprinted gene expression is by definition completely independent of the nucleotide sequence carried by the maternal and the paternal allele. This has been known since the pioneered experimental construction of uniparental conceptuses in an inbred mouse strain [30, 31], where developmental failure could not be attributed to different genetic contribution of the maternal and the paternal pronuclei, but rather supported the existence of epigenetic differences between the parental genomes. It was indeed demonstrated thereafter that imprinted monoallelic expression results from differential marking by DNA methylation of maternal and paternal alleles through their passage in their respective parental germlines [32]. However, there are situations where ASE or allele-specific methylation (ASM) can be coupled to genetic polymorphisms (Figure 1B). In outbred mice or in human samples, different alleles of a number of non-imprinted genes are not found equally at the mRNA level, with a more favorable expression of one polymorphism, always the same, in all the cells of the individual or

expression. *Cis*-acting variants can influence the transcription of each allele, leading to the overrepresentation of one allele at the mRNA level. **(C)** Infiltration of the placenta by maternal material. The model represented here is based on a gene that is biallelically expressed in the embryo and not expressed in the placenta where maternal expression is nonetheless detected. In this case, observed maternal-specific expression does not originate from the expression of the maternal allele in the embryonic part of the placenta, but rather from transcripts carried by maternal cells.

a tissue. The use of reciprocal crosses in mouse and of reciprocal SNP combinations in humans (when available) can clearly distinguish between imprinted expression and genetically determined ASE: allele preference will exchange with reverse parental transmission for imprinted genes, while the same allele will be overrepresented in case of non-imprinted ASE. Genetically determined ASE can not be mistaken with random monoallelic expression in mixed-cell populations, but as mentioned above, would show exactly the same characteristics in monoclonal cell lines.

Such influence of the genetic background on allelic expression is referred as strain bias or expression quantitative trait locus (eQTL). This represents a prevalent cause of variation in gene expression dosage and in phenotypes, and a potential source of disease susceptibility. Gene-expression profiling performed by cDNA microarray in two commonly used mouse inbred strains was instrumental in highlighting how differing genetic backgrounds contribute to expression variability [33]. Different levels of expression were observed at 1% of genes of the male adult brain between the C57Bl6/J and the 129SvEv strains, ranging from 2- to ~20-fold. This initial study did not investigate whether expression differences between the two alleles were conserved in (C57Bl6/J × 129SvEv) F1 hybrids, but more recent works have clearly revealed the widespread existence of ASE in mammals. In the mouse, 7 out of 69 genes showed this type of allelic asymmetry in three tissues of adult hybrid females [34]. In human, a first report revealed ~1.3- to 4.3-fold differences in allelic expression of 6 out of 13 genes studied in a large panel of heterozygote individuals [35]. In a larger survey, ASE was detected at 18% of 129 genes that were selected for their potential function in metabolism and immunity [36]. We will review here the causes for this type of genetically determined allelic asymmetry and in which contexts they can be mistaken with or mask imprinted expression.

Strain bias in polymerase chain reaction amplification

Determination of the allelic status of gene expression is often based on RT-PCR assays that quantify the relative amplification of two expressed SNPs in RNA of hybrid animals or heterozygous human samples. However, PCR-based approaches are a common source of technical artifacts in the

interpretation of ASE, by revealing a preferential amplification of an allele, rather than a true preferential expression of this allele. SNPs can indeed affect the efficiency of the RT-PCR reaction, by creating a mismatch at the site of primer annealing or by locally creating secondary mRNA or cDNA structures. In mouse, the C57Bl6/J strain constitutes the reference genome and primers used to investigate allelic expression are designed on this background. In hybrid situations between C57Bl6/J and a different mouse strain, a more favorable amplification of the C57Bl6/J-inherited SNP is often observed as a result of additional SNPs present in the vicinity of the primers in the opposite strain, which reduce the amplification rate of the corresponding allele. In humans, similar misinterpretation can apply, with preferential amplification of the allele carrying the SNP present on the genome used as a reference for primer design.

Sequencing a few hundred base pairs around the SNP of interest in the non-C57Bl6/J strain used in hybrids will allow the identification of unreferenced SNPs that could adversely affect the PCR reaction. Primers should then be designed on confirmed SNP-free regions, to limit potential *cis*-effects on annealing efficiency or elongation rate. Also, unbiased primers should be systematically tested in contexts of equal representation of each allele, such as genomic DNA of hybrid animals or heterozygous human samples. This requires reliance on primers that can amplify both genomic DNA and cDNA, and therefore on a PCR reaction than do not span introns. Alternatively, analysis of premixed samples made of a 50:50 ratio of RNA from the two studied backgrounds can be used. Once the primers have been selected for their ability to amplify two different SNPs with a ratio close to 1:1, they can be used for investigating the allele-specific status of expression of a gene on F1 cDNA samples. Assuming that potential biases in PCR amplification were excluded, RT-PCR approaches on reciprocal F1 crosses have been instrumental in distinguishing parent-of-origin expression from strain-dependent expression in mouse [37, 38]. Amplification-free allele-specific RNase protection assays can also be used, provided that the two parental transcripts can be distinguished by their size.

Strain bias in quantitative trait loci

A combination of *cis*- and *trans*-acting SNPs are susceptible to govern genetically determined ASE or

strain bias in quantitative trait loci. *Cis*-acting regulatory variants that map to promoter regions can reduce the transcriptional efficiency of the associated allele by affecting the recruitment of transcription factors. Alternatively, *cis*-acting SNPs can also influence the strength of long distance regulatory sequences such as enhancers. *Trans*-acting SNPs on the other side could result from alteration in sequence-recognition domains of transcription factors, where the SNP will render the protein less apt to bind specific alleles.

Implication of *cis*-acting functional elements is easier to evidence, as they will become immediately apparent in quantitative assessment of allele ratios of expressed SNPs in RNA (cDNA), normalized to ratios in corresponding genomic DNA. Using this scheme, the occurrence of *cis* regulatory variants in human has been estimated on a genome-wide scale by hybridization on SNP-specific microarrays [39–41]. In lymphoblastoid cell lines, ASE was estimated to affect 20% of the genes [41], and could be linked to genetic variants in 30% of the cases [40]. Interestingly, ~30% of heterozygous SNPs were estimated to display local ASM in a survey of 16 human pluripotent and adult cell lines [42], therefore in a similar rate than ASE. This had led to the idea that *cis*-variants may influence allelic expression by first altering the propensity of each allele to be methylated [43, 44]. The notion that the majority of allelic differences in expression could be determined by methylation-sensitive *cis*-acting SNPs was also proposed in mouse studies. DNA methylation profiling of macrophages uncovered the existence of >400 non-imprinted differentiated regions between two inbred stains (C57Bl6/J and BALB/c), which were also maintained in the context of hybrid animals [45]. So, *cis*-variants seem to alter expression by changing the genetic information and also its epigenetic characteristics.

Mechanisms leading to sequence-specific ASM can be different among loci. Allele-specific affinity for DNA-binding proteins with downstream effects on DNA methylation is one possibility; direct effects on sequences targeted by DNA methylation, the CpG dinucleotides, is another one. In this regards, two recent studies showed that most of the ASM is contributed by heterozygous CpG-SNPs in the human genome [42, 46]. In summary, a large fraction of allele-specific epigenetic differences in humans and outbred mouse populations can in fact be connected to genetic variation. This contrasts

with imprinted genes or randomly expressed genes, whose monoallelic expression is independent of the genetic background.

Genetically determined ASE can, however, mask imprinted expression, for those imprinted genes that show a preferential expression of one allele rather than a strict monoallelic expression. For genes that are both ‘partially imprinted’ and influenced by strain biases, apparently contradictory results will be obtained from reciprocal crosses. The evidence for imprinting will be stronger in the reciprocal cross for which the parental allele that is preferentially expressed coincides with the strain with the highest level of expression [9]. In this cross, there will therefore be evidence for ASE. Conversely, when the partially silenced parental allele coincides with the strain with the highest level of expression, this allele will be expressed at a certain level and there will therefore be evidence for expression from the two alleles. A typical example of these potential confounding effects is illustrated by the *Th* and the *Dhrc7* genes, which map to the boundary of an imprinted domain of the mouse distal chromosome 7 and have been classified as non-robust imprinted genes, based on the conflicting results obtained with different hybrid mouse crosses [7]. This interfering effect of the genetic sequence on the imprinting status is likely to culminate in highly genetically divergent crosses. Variable loss of imprinting is indeed known to occur in the offspring of inter-specific matings between distant mouse strains, depending on the orientation of the cross [47, 48].

PLACENTA-ONLY IMPRINTED GENES: MATERNAL EXPRESSION OR MATERNAL CONTAMINATION?

Genomic imprinting co-evolved with the emergence of a placental mode of reproduction in mammals, ~160 million years ago [49]. Functional relationships between genomic imprinting and the placenta are also evident in regard of the recurrent placental phenotypes observed after genetic inactivation of imprinted genes and in imprinting-deficient mouse models [50, 51]. Studies of the stage and temporal expression of imprinted genes are also very informative in this regard [52]. Among the hundred genes that are known to be imprinted, at least half of them are expressed in the placenta and

Table 2: List of placenta-imprinted only genes in mouse

Gene	Mouse chromosome region	Expressed allele	Regional DMR/ICR	Name
<i>Gatm</i>	Prox2	M	?	Glycine amidinotransferase (L-arginine:glycine amidinotransferase)
<i>Tfpi2</i>	Prox6	M	<i>Peg10</i> ICR	Tissue factor pathway inhibitor 2
<i>Ppplr9a</i>	Prox6	M	<i>Peg10</i> ICR	Protein phosphatase 1, regulatory (inhibitor) subunit 9A
<i>Pon3</i>	Prox6	M	<i>Peg10</i> ICR	Paraoxonase 3
<i>Pon2</i>	Prox6	M	<i>Peg10</i> ICR	Paraoxonase 2
<i>Th</i>	Dist7	M	KvDMR and <i>H19</i> DMD	Tyrosine hydroxylase
<i>Ascl2/Mash2</i>	Dist7	M	KvDMR	Achaete-scute complex homolog 2
<i>Tspan32</i>	Dist7	M	KvDMR	Tetraspanin 32
<i>Cd81</i>	Dist7	M	KvDMR	CD81 antigen
<i>Tssc4</i>	Dist7	M	KvDMR	Tumor-suppressing subchromosomal transferable fragment 4
<i>Nap114</i>	Dist7	M	KvDMR	Nucleosome assembly protein 1-like 4
<i>Tnfrsf23</i>	Dist7	M	KvDMR	Tumor necrosis factor receptor superfamily, member 23
<i>Osbp15</i>	Dist7	M	KvDMR	Oxysterol binding protein-like 5
<i>Dcn</i>	Dist10	M	?	Decorin
<i>Slc38a4/Ata3</i>	Dist15	P	<i>Slc38a4</i>	Solute carrier family 38, member 4
<i>Slc22a3</i>	Prox17	M	<i>Igf2r</i> ICE	Solute carrier family 22 (organic cation transporter), member 3
<i>Slc22a2</i>	Prox17	M	<i>Igf2r</i> ICE	Solute carrier family 22 (organic cation transporter), member 2

The genes with a doubtful imprinted status are indicated in Bold. M: maternal expression; P: paternal expression.

extraembryonic tissues (62/132), and more than half of these (35/62) are expressed in a parent-of-origin manner in these tissues [51–53]. Among these placentally imprinted genes, 18 are also imprinted in the embryo proper and some adult tissues. The remaining 17 can also be expressed in other tissues, but they show all an imprinted expression exclusive to the placenta, and are thus referred to as placenta-only imprinted genes (Table 2).

The extraembryonic tissues, which include the placenta, constitute an interface between the mother and the embryo/fetus. They derive from the trophoblast and the primitive endoderm lineages that emerge at ~3.5 and 4.5 dpc during mouse development and whose role is to support the implantation of the embryo in the uterine wall, by promoting invasion and angiogenesis [54]. The maternal deciduum appears as an inflammatory response to the implantation of the embryo and surrounds the embryo in the pregnant uterus, in close contact with the most external layer of extraembryonic cells, the trophoblast giant cells. After a period of autonomous growth, the embryo becomes dependent upon maternal resource allocation, and establishes a connection with the mother at ~8.5 dpc to form the precursor of the placenta, the chorion. The placenta is composed of these juxtaposed fetal and maternal tissues that converge at the site of embryonic

implantation. Maternal and fetal blood vessels are interconnected within the middle layer of the placenta called the labyrinth [55]. Several studies have reported that maternal cells can be found in the fetal part of the placenta and even in fetal organs or tissues, illustrating that not only molecule exchanges but also a true cell trafficking occur between the two entities [56, 57]. Moreover, it has been reported that most of placental samples will be contaminated with maternal cells after 8.5 dpc, because of the technical difficulty in separating intermingled maternal and fetal tissues during dissection [58].

Strikingly, all but one of the placenta-only imprinted genes are maternally expressed in the placenta from hybrid F1 crosses [52]. Among the 17 genes listed in Table 2, only *Slc38a4* is a paternally expressed gene. Considering those close physical relationships between maternal and embryonic tissues in the placenta and in the trophoblast giant cell layer, it seems very likely that maternal-specific detection in these tissues may result from a dissection-based contamination with maternal cells or an infiltration of maternal cells (Figure 1C), and not from the expression of the maternal allele in the embryonic part of the placenta. This confounding effect will be exacerbated for genes that are highly expressed in the maternal deciduum or maternal blood cells. In this regard, the placenta-only imprinted gene *Tnfrh1* is

mildly expressed in most of embryonic and adult tissues where it is not imprinted. It is, however, very strongly expressed at the materno–fetal interface and is coincidentally found to be maternally expressed in the most external layer of extraembryonic cells that are directly in contact with the maternal deciduum [59]. This type of example raises a serious doubt about the real imprinted status of this category of genes.

As a whole, 15 genes display a maternal-specific expression in tissues that are highly prone to maternal contamination or infiltration. Other criteria converge toward the idea that 13 of these placenta-only imprinted genes may be false positives: (i) unknown dependence upon a *cis-acting* DMR: none of them are associated with a differentially methylated DMR mapping to their promoter, two of them map to a region where no DMR has been defined (*Gatm* and *Dcn*) and the dependence of the others upon the local ICR has not been determined; (ii) unusual genomic organization: two of them are present as singletons (*Gatm* and *Dcn*), while most of imprinted genes are organized in clusters, around an imprinting control region that carries the parent-specific methylation mark inherited from the germline and can act upon several megabases; (iii) lack of placental phenotype or lack of data concerning placental function: genetic inactivation of these genes does not lead to a placental phenotype, or alternatively, data concerning their functional investigation in the placenta are not available and (iv) lack of evolutionary conservation: placental physiology greatly differs between mice and humans. This may result in a different degree of maternal infiltration in these species, and in this regard, the imprinting status of these genes is, most of the time, not conserved in humans [60]. Each of this argument does not exclude a real imprinted status when taken single-handedly. Indeed, it has been argued that imprinted expression is independent of DNA methylation in the placenta and may rather only rely on histone modifications marks in this organ [61]. Similarly, differences in placental physiology may actually impose different gene networks and mode of regulation between mice and humans [62]. However, accumulation of these four features motivates a careful reexamination of the status of these genes we currently consider as maternally expressed in the placenta.

The classical (AxB) F1 hybrid progenies used to screen or validate imprinted genes do not allow to make the distinction between a maternal transcript

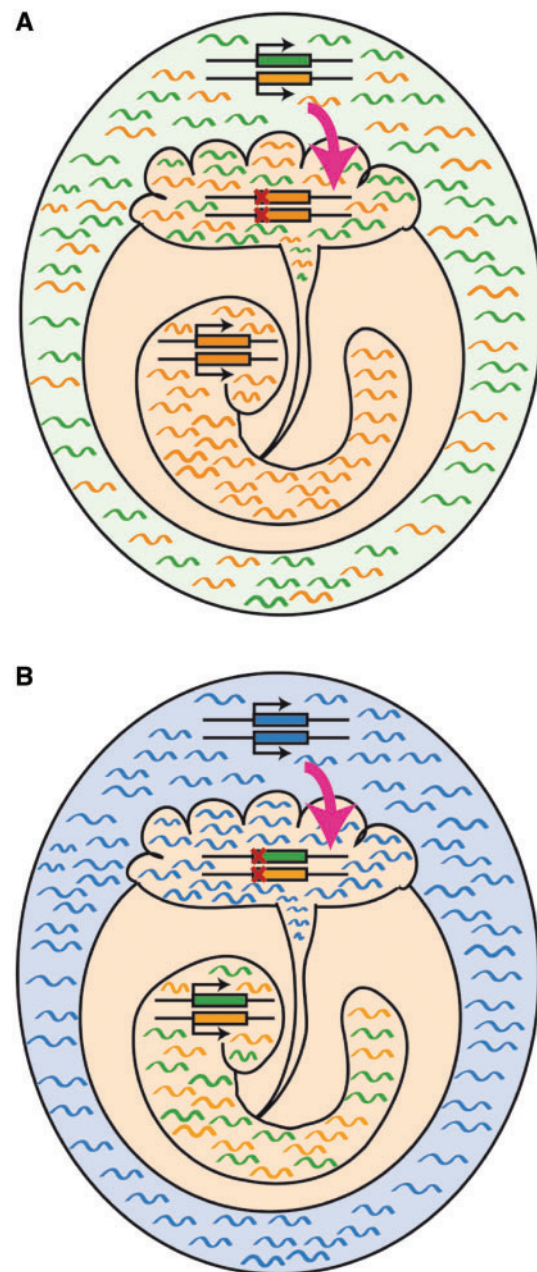


Figure 2: Experimental guidelines to detect maternal cells in the placenta. For simplification, we chose to illustrate a situation where there is very little expression from the placenta. **(A)** Perform $(A \times B) \times B$ backcrosses to generate a $(B \times B)$ homozygous embryo in an $(A \times B)$ heterozygous mother. The presence of A transcripts in the placenta reveals an infiltration with maternal material. **(B)** Transfer $(A \times B)$ embryos in the uterus of a $(C \times C)$ homozygous mother. Presence of C transcripts in the placenta will provide evidence for a transfer of maternal cells from the recipient mother ($C \times C$, blue transcripts), and not expression from the maternal allele provided by the biological mother (A, green transcripts, Orange: B transcripts, Blue: C transcripts).

coming from the inbred mother, which will be of the A type, or from the maternal allele of the F1 embryonic placenta, which will also be from the A type. To rule out a maternal contamination or infiltration, two experimental designs can be used and they both rely on the ability to distinguish a transcript provided by maternal cells, from a transcript derived from the maternal allele of the embryo (Table 1). First, backcrosses such as $(A \times B) \times B$ can be used to generate a homozygous $(B \times B)$ embryo in an $(A \times B)$ heterozygous maternal environment (Figure 2A). Any detection of the A allele that is not present in the genome of the embryo will signal the existence of a transfer of maternal material in the placenta, rather than expression from the maternal allele. Detection of the B allele does not however allow the distinction between the two possibilities. Same principles will apply for $(A \times B) \times A$ reverse backcrosses. Second, transfer of a heterozygous $(A \times B)$ embryo in a foster mother of a third allelic composition, C, is also possible, provided that three nucleotides exist at the same SNP position between the three involved strains (Figure 2B). In this case, expression of the C allele in the $(A \times B)$ placenta will also reveal maternal contamination/infiltration from the recipient mother, and not expression from the allele transmitted by the biological mother. Unfortunately, these crucial validating experiments have been very scarcely performed in the literature [9, 59].

CONCLUSION

A high number of confounding effects exist in the identification of imprinted genes, especially when screening is based on differential expression of parental alleles. Random monoallelic expression and *cis*-acting variants can not only be misinterpreted as imprinted expression when appropriate reciprocal crosses are not performed, but they can also mask a real imprinted status. Moreover, placental tissues, which are evolutionary and functionally relevant to the biology of imprinting, are highly prone to maternal contamination and a potential source of false evidence of maternal-specific expression. Experimental approaches should be carefully designed to avoid the use of monoclonal cell populations and combine reciprocal crosses, but also backcrosses, to increase the robustness of screening and validation of imprinted expression.

Key Points

- Evidence for monoallelic expression is not systematically synonymous of imprinted expression.
- Monoclonal cell populations as well as placental tissues are particularly prone to display imbalance in the representation of specific parental alleles at the RNA level.
- As a conclusion, screening approaches based on differential allelic methylation in biparental material or in comparison with models of imprinting deficiency are likely more informative in the detection of imprinted loci.

Acknowledgements

We apologize to colleagues whose work could not be cited due to space limitations.

FUNDING

Agence pour la Recherche contre le Cancer (ARC) (to C.P.) and an European Young Investigator (EURYI) award (to D.B.).

References

1. Bourc'his D, Xu GL, Lin CS, *et al.* Dnmt3L and the establishment of maternal genomic imprints. *Science* 2001;**294**: 2536–9.
2. Trasler JM. Gamete imprinting: setting epigenetic patterns for the next generation. *Reprod Fertil Dev* 2006;**18**: 63–9.
3. Henckel A, Arnaud P. Genome-wide identification of new imprinted genes. *Brief Funct Genomics* 2010;**9**:304–14.
4. Luedi PP, Hartemink AJ, Jirtle RL. Genome-wide prediction of imprinted murine genes. *Genome Res* 2005;**15**: 875–84.
5. Luedi PP, Dietrich FS, Weidman JR, *et al.* Computational and experimental identification of novel human imprinted genes. *Genome Res* 2007;**17**:1723–30.
6. Smith RJ, Dean W, Konfortova G, *et al.* Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res* 2003;**13**:558–69.
7. Schulz R, Menhenniott TR, Woodfine K, *et al.* Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. *Nucleic Acids Res* 2006;**34**:e88.
8. Wen B, Wu H, Bjornsson H, *et al.* Overlapping euchromatin/heterochromatin-associated marks are enriched in imprinted gene regions and predict allele-specific modification. *Genome Res* 2008;**18**:1806–13.
9. Brideau CM, Eilertson KE, Hagarman JA, *et al.* Successful computational prediction of novel imprinted genes from epigenomic features. *Mol Cell Biol* 2010;**30**:3357–70.
10. Ishino F, Kuroiwa Y, Miyoshi N, *et al.* Subtraction-hybridization method for the identification of imprinted genes. *Methods Mol Biol* 2001;**181**:101–12.
11. Oakey RJ, Beechey CV. Imprinted genes: identification by chromosome rearrangements and post-genomic strategies. *Trends Genet* 2002;**18**:359–66.

12. Babak T, Deveale B, Armour C, *et al.* Global survey of genomic imprinting by transcriptome sequencing. *Curr Biol* 2008;**18**:1735–41.
13. Wang X, Sun Q, McGrath SD, *et al.* Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One* 2008;**3**:e3839.
14. Khatib H. Is it genomic imprinting or preferential expression? *Bioessays* 2007;**29**:1022–8.
15. Wood AJ, Oakey RJ. Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet* 2006;**2**:e147.
16. Pollard KS, Serre D, Wang X, *et al.* A genome-wide approach to identifying novel-imprinted genes. *Hum Genet* 2008;**122**:625–34.
17. Gregg C, Zhang J, Weissbourd B, *et al.* High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 2010;**329**:643–8.
18. Krueger C, Morison IM. Random monoallelic expression: making a choice. *Trends Genet* 2008;**24**:257–9.
19. Rougeulle C, Navarro P, Avner P. Promoter-restricted H3 Lys 4 di-methylation is an epigenetic mark for monoallelic expression. *Hum Mol Genet* 2003;**12**:3343–8.
20. Nora EP, Heard E. X chromosome inactivation: when dosage counts. *Cell* 2009;**139**:865–7.
21. Chess A, Simon I, Cedar H, *et al.* Allelic inactivation regulates olfactory receptor gene expression. *Cell* 1994;**78**:823–34.
22. Rhoades KL, Singh N, Simon I, *et al.* Allele-specific expression patterns of interleukin-2 and Pax-5 revealed by a sensitive single-cell RT-PCR analysis. *Curr Biol* 2000;**10**:789–92.
23. Hollander GA, Zuklys S, Morel C, *et al.* Monoallelic expression of the interleukin-2 locus. *Science* 1998;**279**:2118–21.
24. Held W, Roland J, Raulet DH. Allelic exclusion of Ly49-family genes encoding class I MHC-specific receptors on NK cells. *Nature* 1995;**376**:355–8.
25. Pereira JP, Girard R, Chaby R, *et al.* Monoallelic expression of the murine gene encoding Toll-like receptor 4. *Nat Immunol* 2003;**4**:464–70.
26. Gimelbrant A, Hutchinson JN, Thompson BR, *et al.* Widespread monoallelic expression on human autosomes. *Science* 2007;**318**:1136–40.
27. Wang J, Valo Z, Smith D, *et al.* Monoallelic expression of multiple genes in the CNS. *PLoS One* 2007;**2**:e1293.
28. Pagnol V, Uz E, Wallace C, *et al.* Extreme clonality in lymphoblastoid cell lines with implications for allele specific expression analyses. *PLoS One* 2008;**3**:e2966.
29. Moreira de Mello JC, de Araujo ES, Stabellini R, *et al.* Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. *PLoS One* 2010;**5**:e10947.
30. McGrath J, Solter D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 1984;**37**:179–83.
31. Barton SC, Surani MA, Norris ML. Role of paternal and maternal genomes in mouse development. *Nature* 1984;**311**:374–6.
32. Tucker KL, Beard C, Dausmann J, *et al.* Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev* 1996;**10**:1008–20.
33. Sandberg R, Yasuda R, Pankratz DG, *et al.* Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc Natl Acad Sci USA* 2000;**97**:11038–43.
34. Cowles CR, Hirschhorn JN, Altshuler D, *et al.* Detection of regulatory variation in mouse genes. *Nat Genet* 2002;**32**:432–7.
35. Yan H, Yuan W, Velculescu VE, *et al.* Allelic variation in human gene expression. *Science* 2002;**297**:1143.
36. Pastinen T, Sladek R, Gurd S, *et al.* A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* 2004;**16**:184–93.
37. Ruf N, Dunzinger U, Brinckmann A, *et al.* Expression profiling of uniparental mouse embryos is inefficient in identifying novel imprinted genes. *Genomics* 2006;**87**:509–19.
38. Ruf N, Bähring S, Galetzka D, *et al.* Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. *Hum Mol Genet* 2007;**16**:2591–9.
39. Milani L, Lundmark A, Kiialainen A, *et al.* DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. *Blood* 2010;**115**:1214–25.
40. Ge B, Pokholok DK, Kwan T, *et al.* Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* 2009;**41**:1216–22.
41. Serre D, Gurd S, Ge B, *et al.* Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet* 2008;**4**:e1000006.
42. Shoemaker R, Deng J, Wang W, *et al.* Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* 2010;**27**:883–9.
43. Kerkel K, Spadola A, Yuan E, *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 2008;**40**:904–8.
44. Zhang Y, Rohde C, Reinhardt R, *et al.* Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol* 2009;**10**:R138.
45. Schilling E, El Chartouni C, Rehli M. Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome Res* 2009;**19**:2028–35.
46. Hellman A, Chess A. Extensive sequence-influenced DNA methylation polymorphism in the human genome. *Epigenetics Chromatin* 2010;**3**:11.
47. Vrana PB, Fossella JA, Matteson P, *et al.* Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in *Peromyscus*. *Nat Genet* 2000;**25**:120–4.
48. Shi W, Lefebvre L, Yu Y, *et al.* Loss-of-imprinting of Peg1 in mouse interspecies hybrids is correlated with altered growth. *Genesis* 2004;**39**:65–72.
49. Renfree MB, Hore TA, Shaw G, *et al.* Evolution of genomic imprinting: insights from marsupials and monotremes. *Annu Rev Genomics Hum Genet* 2009;**10**:241–62.
50. Bourc'his D, Proudhon C. Sexual dimorphism in parental imprint ontogeny and contribution to embryonic development. *Mol Cell Endocrinol* 2008;**282**:87–94.

51. Reik W, Constancia M, Fowden A, *et al.* Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. *J Physiol* 2003;**547**:35–44.
52. Hudson QJ, Kulinski TM, Huetter SP, *et al.* Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. *Heredity* 2010;**105**:45–56.
53. Coan PM, Burton GJ, Ferguson-Smith AC. Imprinted genes in the placenta—a review. *Placenta* 2005;**26**(Suppl A):S10–20.
54. Rossant J, Cross JC. Placental development: lessons from mouse mutants. *Nat Rev Genet* 2001;**2**:538–48.
55. Rinkenberger J, Werb Z. The labyrinthine placenta. *Nat Genet* 2000;**25**:248–50.
56. Piotrowski P, Croy BA. Maternal cells are widely distributed in murine fetuses *in utero*. *Biol Reprod* 1996;**54**:1103–10.
57. Vernochet C, Caucheteux SM, Kanellopoulos-Langevin C. Bi-directional cell trafficking between mother and fetus in mouse placenta. *Placenta* 2007;**28**:639–49.
58. Higashimoto K, Soejima H, Yatsuki H, *et al.* Characterization and imprinting status of OBPH1/Obph1 gene: implications for an extended imprinting domain in human and mouse. *Genomics* 2002;**80**:575–84.
59. Clark L, Wei M, Cattoretti G, *et al.* The Tnfrh1 (Tnfrsf23) gene is weakly imprinted in several organs and expressed at the trophoblast-decidua interface. *BMC Genet* 2002;**3**:11.
60. Monk D, Arnaud P, Apostolidou S, *et al.* Limited evolutionary conservation of imprinting in the human placenta. *Proc Natl Acad Sci USA* 2006;**103**:6623–8.
61. Lewis A, Mitsuya K, Umlauf D, *et al.* Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat Genet* 2004;**36**:1291–5.
62. Frost JM, Moore GE. The importance of imprinting in the human placenta. *PLoS Genet* 2010;**6**:e1001015.