BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 9. NO 4. 304–314 doi:10.1093/bfgp/elq016

Genome-wide identification of new imprinted genes

Amandine Henckel and Philippe Arnaud

Advance Access publication date 29 June 2010

Abstract

In the mid-1980s, elegant studies on mouse embryos revealed that both parental genomes are required for normal development leading to the discovery of genomic imprinting. Imprinting is a parent-of-origin-dependent epigenetic mechanism whereby a subset of autosomal genes is expressed from only one of the parental alleles. Imprinting control involves both DNA- and histone-methylation, which differentially mark the parental alleles. More than a hundred imprinted genes have been identified so far, many of which play important roles in the regulation of growth and development. Nonetheless, the full extent of imprinting and its biological functions remain underestimated. In this review, we describe recently developed strategies to identify novel imprinted genes and highlight the potential of combining several high throughput approaches. By integrating databases obtained from epigenome- and transcriptome-wide analyses, we now have the unique opportunity to identify all the imprinted genes in the human/mouse genomes.

Keywords: *imprinted genes; whole-genome screen; allelic-transcription; chromatin signature*

INTRODUCTION

In the mid-1980s, embryological studies in the mouse demonstrated the functional non-equivalence of the maternal and paternal genomes. Specifically, conceptuses derived from zygotes that contained two sets of the maternal chromosomes (called gynogenotes) or two sets of the paternal chromosomes (called androgenotes) failed to develop beyond mid-gestation [1, 2]. These findings established that diploidy alone was not sufficient to sustain embryonic development and suggested the existence of genes that were expressed only on the allele inherited from the mother or the father. These genes were called 'imprinted genes'. In the following years this assumption was confirmed through the characterization of insulin-like growth factor 2 (*Igf2*) as a paternally-expressed and of *H19* and *Igf2r* (IGF2-receptor) as maternally-expressed genes in the mouse [3–5].

Since then, more than hundred imprinted genes have been identified in the mouse and about half of them are expressed from the maternal allele and the other half from the paternal allele (an up-to-date list is available at http://www.har.mrc.ac.uk/research/genomic_imprinting/). Although for the majority of them the imprinting status is conserved also in their human orthologues, some failed to show imprinted expression in humans [6, 7]. More rarely, some are imprinted in the human only (see http://igc.otago.ac.nz/home.html for details). Nevertheless, the number of identified imprinted genes is unlikely to be complete as the precise extent of imprinting in the mouse or human genome is not fully known. Estimates of their actual total number range from 100 to 2100 [7–9].

As suggested by the developmental failure of uniparental conceptuses [1, 2], many of these genes are involved in regulation of cell proliferation and growth in both placenta and embryo. Others play key roles in neurological processes and in behaviour [10, 11]. As a consequence, perturbations in imprinted gene expression are an important cause of several growth and behavioural syndromes in humans including the Beckwith–Wiedemann,

Corresponding author. Philippe Arnaud, Institut de Génétique Moléculaire, CNRS UMR5535 and Universités de Montpellier, 1919, route de Mende, 34090 Montpellier, France. Tel: +33 467613661; Fax: +33 467040231; E-mail: philippe.arnaud@igmm.cnrs.fr

Amandine Henckel has worked, during her PhD, on the epigenetic regulation of genomic imprinting. She is currently a postdoctoral researcher at the Max Planck Institute in Freiburg, Germany, where she is interested in the histone H1 and its role in chromatin organisation.

Philippe Arnaud is a research scientist in the 'Institut de Génétique Moléculaire', CNRS. His research interests are on the role of DNA methylation and histone modifications in the regulation of genomic imprinting in mammals.

Prader–Willi and Angelman syndromes [12], as well as in cancer [13].

Despite the major advances of the past 25 years, the picture of the biological functions in which imprinting plays a role is probably not completed yet and therefore the accurate identification of the whole set of imprinted genes present in mammalian genomes is crucially needed.

In this review, after a brief characterization of imprinted genes, we describe the new approaches that are currently used to identify novel imprinted genes with a strong emphasis on the potential of recently developed high throughout approaches.

MAIN FEATURES OF IMPRINTED GENES

Imprinted genes are distributed throughout the genome. Although some are isolated and others occur in pairs, the majority are organized in clusters that are up to one megabase in size and are structurally conserved between human and mouse. These imprinted domains contain both maternally and paternally expressed genes as well as non-imprinted genes and at least one non-coding RNA in most cases (Figure 1).

The allele-specific expression of imprinted genes (and more broadly of imprinted loci) is regulated by epigenetic modifications of which DNA methylation is a major component. Notably, the discrete *cis*-acting regions that orchestrate imprinted monoallelic expression (referred to as ICRs for imprinting control regions) all coincide with differentially methylated region (DMR) harbouring allelic DNA methylation inherited from the male or female gamete (the germline DMR) [14, 15]. Germline acquisition of DNA methylation at ICRs requires the DNA methyltransferase 3A (DNMT3A) [16] in cooperation with the DNA methyltransferase 3 Like (DNMT3L) [14, 17]. Once acquired, DNA methylation imprints are maintained in all the somatic lineages throughout development and adult life. These imprinting marks are 'read' in different ways to ensure appropriate parental allele-specific expression [12]. Schematically, differential DNA methylation of ICRs is thought to initiate a sequence of events that will lead to the coordinated allele-specific expression of entire clusters of imprinted genes. Importantly, for many imprinted genes these 'reading' mechanisms and the consequent expression of the imprinted genes are limited to specific tissues or developmental stages, despite the constitutive presence of the allelic DNA methylation imprint at the ICR (Figure 1). Beside germline DMRs at ICRs, imprinted domains can also contain DMRs at which parental allele-specific methylation is established after fertilization only. These so called 'somatic DMRs' are found at promoters of some imprinted genes and can be tissue-specific (Figure 1).

IDENTIFICATION OF IMPRINTED GENES

Beside the data obtained using uniparental conceptuses, the existence of imprinted genes was further supported by genetic studies on mouse embryos that were uniparentally disomic for individual chromosomes, or for chromosomal regions (i.e. mice that have inherited two maternal copies and no paternal copy, or vice versa, of a chromosome or portion of a chromosome). Mice characterized by uniparental disomy (UPD) of different chromosomal regions to cover almost the entire mouse genome were used to carry out a phenotypic screen for imprinting effects [18]. By identifying a number of regions in which UPD caused a broad range of phenotypic abnormalities, including abnormal growth and embryonic lethality, this approach allowed the delineation of genomic regions that are likely to contain imprinted genes. The results of this screen constitute the 'backbone' of the current imprinting map (http://www.har.mrc.ac.uk/research/genomic_imprinting/) [19] and many of the imprinted genes identified so far map to the 'imprinted' regions that have been delineated by using UPD mice. However, not all imprinted genes are necessarily in these regions. For instance, regions containing imprinted genes, the deregulation of which leads to a 'subtle' developmental effect, might not be revealed by this kind of phenotypic screen or their identification as imprinted regions can be made difficult by earlier lethality or phenotypes on the same chromosome. The evaluation of the full biological function of imprinted genes requires thus genome-wide approaches that enable the identification of imprinted genes independently of their genomic localization. To this end, dedicated screens that are based on three main specific features of imprinted loci (i.e. DNA sequence characteristics, imprinted expression and epigenetic signature) have been developed (Table 1).

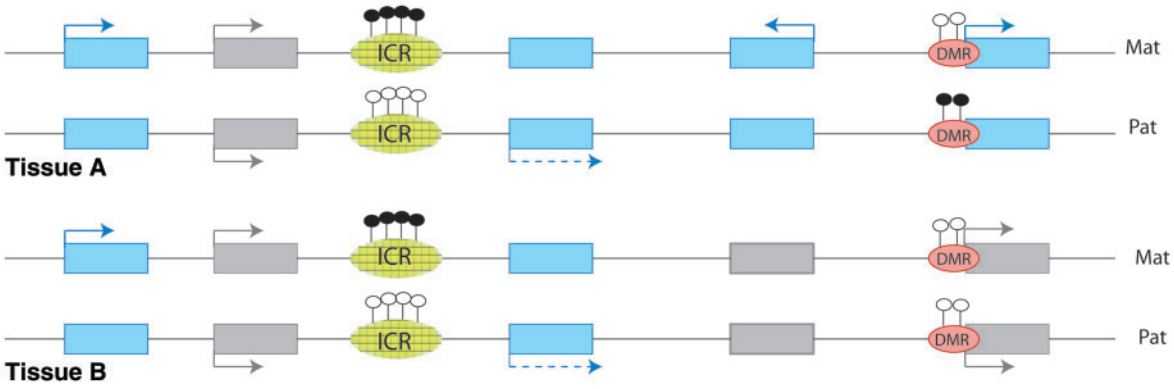

Figure 1: Canonical view of an imprinted domain. Schematic representation of a putative imprinted domain that comprises imprinted (maternally- and paternally-expressed) and non-imprinted coding genes as well as an imprinted non-coding RNA (dotted line). Imprinted expression is controlled by a maternally methylated (filled lollipops) germline DMR/ICR (hatched oval) and further relies, for a single transcript, on a tissue-specific paternally methylated region (somatic DMR). The epigenetic signature at the ICR is maintained during all somatic development, although imprinted expression can vary from one tissue and/or developmental stage to another (compare tissue A and B).

Table 1: Comparison of the different approaches used to identify new imprinted genes

Screen based on	Advantages	Drawbacks	Techniques
Sequence analysis	Availability of whole genome sequences	Screen based on sequence features of known imprinted genes, potentially discarding imprinted genes controlled in a different manner. It requires experimental validation	Computational analysis (machine learning algorithms)
Transcription	By focusing on parental-allele-specific expression can virtually identified all imprinted genes	Identification limited to genes expressed in the tissue and at the developmental stage analysed. Mono-allelic expression is not an exclusive characteristic of imprinted genes	Subtractive hybridization Differential display cDNA hybridization on microarrays RNA-seq ChIP-SNP for transcription factors ChIP-seq for histone marks associated with transcription
Chromatin signature	Feature constitutively maintained through all somatic development (at germline DMRs)	More likely to delineate imprinted domains rather than to identify individual imprinted genes	*DNA methylation* RLGS, Me-RDA MeDIP combined with microarray, BS-seq *Histone modifications* ChIP-chip, ChIP-seq

IMPRINTED GENE PREDICTION THROUGH DNA SEQUENCE CHARACTERISTICS

The identification of an important number of imprinted genes offers opportunities for the computational identification of candidate imprinted genes. Computational approaches are based on the identification of specific sequence characteristics that are shared by imprinted genes [20–24]. Specifically, human imprinted regions significantly lack short interspersed transposable elements (SINEs) in comparison to non-imprinted loci [20, 21]. The first large-scale prediction of imprinted genes based on DNA sequence characteristics alone was performed by Luedi *et al.* [25]. They conducted a comparative analysis of DNA sequence features in 44 known imprinted genes and 530 genes presumed, but not experimentally determined, to be non-imprinted in the mouse genome. This analysis included statistics on the distribution of several families of repetitive elements, transcription factor binding sites and CpG islands. Among the most

significant characteristics identified, the authors confirmed that low concentration of SINEs is a significant feature of imprinted regions and determined that their orientation relative to the imprinted gene has a high discriminatory value. In addition, the presence within introns of endogenous retrovirus elements (ERVs) and long interspersed element (LINE) L1s is also likely to be important to predict the imprinting status of a gene. Subsequently, these 'predictive' features were used to train a classifier to predict the imprinted or non-imprinted status of a gene and from which parental allele a candidate imprinted gene should be expressed. Application of this classifier to a total of 23 788 annotated autosomal genes resulted in the identification of 600 (2.5%) candidate imprinted genes, of which 384 (64%) were predicted to be maternally expressed [25]. A similar, but more conservative and selective approach based on the combined use of four classifiers was then used to analyse the human genome [26]. This strategy allowed the prediction of 156 imprinted genes (0.75%) out of the 20 770 annotated autosomal genes, and 88 (56%) of them were predicted to be expressed on the maternal allele [26]. The difference in the percentage of predicted imprinted genes in the mouse and human genomes (2.5 and 0.75%, respectively) can be explained by the use of a more stringent approach, but may also highlight the fact that the human genome contains fewer imprinted genes than the mouse genome [26]. Furthermore, the number of predicted imprinted genes in both genomes was probably underestimated as these studies were largely restricted to protein-coding genes.

Finally, among the 156 candidates in the human genome, only disk-large associated protein 2 (*DLGAP2*) and potassium channel, subfamily K, member 9 (*KCNK9*), which were selected due to their high probability to be imprinted, were experimentally demonstrated to be imprinted and to be expressed from the predicted parental allele [26]. Interestingly, *Kcnk9* was previously shown to be imprinted with maternal and brain-specific expression also in the mouse [27]. In this study, 16 promising candidate imprinted genes were selected among the 600 predicted by Luedi *et al.* [25] in the mouse genome and their imprinted status was tested in E11.5 mouse embryos. With the exception of *Kcnk9,* the other 15 genes did not show evidence of imprinted expression at this developmental stage [27].

Beside prediction based on LINE and SINE density, an original study has identified novel imprinted genes (e.g. *Mcts2)* through a hypothesis-driven search for intronic X-chromosome derived retro-elements [24].

The analysis of all these studies indicates that the overall success rate in the identification of new imprinted genes based on sequence features alone remains low, highlighting that such approach has to be considered promising but still in need of improvements. Furthermore, such bioinformatics-based screens cannot substitute, but rather only assist, the direct experimental approach/validation.

SCREENS BASED ON GENE EXPRESSION ANALYSIS

Screens dedicated to unravelling parental allelic imbalance in gene expression appear especially suitable to identify imprinted loci. However, the design of such expression screens needs to fulfil several important requirements: (i) ideally the cDNA source should allow the discrimination of the maternal and paternal alleles; (ii) the cDNA source should represent both protein-coding and non-coding genes (non-coding RNAs) as the expression of both can be regulated by imprinting; (iii) only genes that are imprinted in that tissue and/or at the developmental stage analysed will be identified; (iv) genes identified by these screens need to be distinguished from genes which show random mono-allelic expression [28–30].

The first successful expression screens were based on subtractive hybridization and differential display techniques performed using cDNA from uniparental (i.e. gynogenetic and androgenetic) embryos or embryonic fibroblast cell lines [31–33]. The subsequent development of high throughput microarray techniques allowed screening of the expression of thousands of genes at the same time. Specifically, a large-scale microarray comprising 27 663 full-length mouse cDNAs was utilized to identify imprinted genes by comparing gene expression levels in whole 9.5 dpc gynogenetic and androgenetic mouse embryos. This analysis identified more than 2100 imprinted candidate transcripts (1403 maternally- and 698 paternally-expressed, respectively), including 56 non-coding RNAs [9]. In a more recent study, gene expression levels in 9.5 dpc-old gynogenetic and control embryos were compared using the Affymetrix GeneChip probe array, which contains more than 45 000 genes and ESTs. 'Only'

39 candidate transcripts (including 18 genes already known to be imprinted), which were paternally expressed, were identified [34]. However, when these results were experimentally validated, only very few of the candidate transcripts were successfully confirmed as imprinted while the majority showed no imprinted expression [8, 32, 34, 35]. Combined, these observations are consistent with the idea that the disruption of imprinted gene expression in parthenogenetic and andrenogenetic embryos (and the ensuing developmental defects) might affect also the expression of many non-imprinted genes, which would be detected as false imprinted genes when using uniparental material in expression screens [32, 34, 35]. In addition, uniparental mouse embryos are arrested early in post-implantation development, thus making not possible the identification of imprinted genes that are expressed at later stages during development and/or in tissues which are not yet present or are not functional in such embryos at the moment of death.

Mouse strains that carry specific UPDs or duplications of 'imprinted' chromosomal regions [18, 36] can, at least partially, overcome these limitations. Some of these strains are indeed viable until birth, so that differential gene expression profiles can be investigated in different tissues and compared with the profiles obtained in wild type animals using the microarray technology [37]. Such an approach has successfully identified imprinted genes, of which some were imprinted in a tissue-specific manner [38–41]. For instance, Schulz *et al*. [41] conducted microarray experiments using cDNA derived from different tissues of UPD mice and successfully identified three new maternally expressed genes in the placenta and four novel, brain-specific, paternally expressed transcripts. However, the limitations of expression screens based on UPD mice are obvious. Indeed, the investigation for new imprinted candidates is restricted to phenotypically-defined 'imprinted' chromosomal regions [18, 36]. In addition, as in uniparental embryos, the imprinted defects observed in these embryos may distort the expression pattern of non-imprinted genes, generating false-positive imprinted genes.

An alternative approach is the use of informative single nucleotide polymorphic (SNP) variants that enable not only to establish the parental origin of expression of a given gene, but also to screen physiologically normal 'material'. These can be done in the mouse by reciprocal crossing of different mouse strains. In humans, in order to decipher the origin of complex diseases, haplotype map databases have been built by the international HapMAp consortium (http://hapmap.ncbi.nlm.nih.gov/) and the '1000 genomes' project (http://browser.1000genomes.org/index.html). The identified informative human SNPs, which are available at the dsSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/), can be used in transcription screens to determine the allelic expression of a specific gene or set of genes. Several studies used a SNP-specific microarray approach to investigate allelic-specific expression in human tissues [42] and cell lines [28, 29, 43]. While not necessarily dedicated to the prediction of new imprinted genes, these studies did confirm the differential allelic expression of known imprinted genes and identified several 'high-confidence' new imprinted transcripts. However, the main outcome of these studies is that allelic expression bias, irrespectively of the parental origin of the alleles, is a widespread phenomenon in the human genome [28], possibly affecting up to 20% of genes [30]. In order to identify new imprinted genes, Pollard *et al*. [44] designed an approach that allowed discriminating 'true' candidate imprinted genes from those that present random mono-allelic expression. In this study allelic expression was investigated using SNP-specific microarrays in lymphoblastoid cell lines derived from 67 unrelated individuals. Among the genes, which showed differential allelic expression, 'true' candidate imprinted genes were identified through the over- or under-expression of the SNP-associated-allele in comparison to the other allele in different heterozygous individuals. Accordingly, out of the 2625 human genes analysed 61 candidate imprinted genes were identified. Among the 15 genes (out of the 61 identified) experimentally tested for imprinting, seven showed strong evidence of, which was however not formally proven, being imprinted. Given that this screen covered ~10% of the protein-coding genes in humans (but not the non-coding RNAs), we should expect no more than a few hundred imprinted genes in lymphoblastoid cells [44]. Although forthcoming microarrays will offer a more extensive coverage, such an approach has several limitations. The analysis is conducted on custom-selected, and necessarily limited, genomic regions and they require an a priori knowledge of SNP position and transcript sequences. In addition, microarrays analysis cannot give a reliable quantitative ratio of the expression of a given gene

at the two parental alleles, thus not facilitating the identification of imprinted genes that display an allelic parental bias (rather than an 'all-or-none') in expression.

The recent development of computer-assisted deep-sequencing approaches, the next-generation sequencing technologies, provides a powerful and promising alternative to microarray-based analyses. Particularly interesting for the identification of new imprinted genes is the RNA-sequencing (RNA-seq) approach [45] because it enables quantitative measurements of allelic bias in whole transcriptomes when applied to polymorphic cDNA sources. Wang *et al*. [46] sequenced the entire transcriptome of neonatal brain samples from mice generated from reciprocal crossing between different strains. They identified 26 candidate imprinted genes, including 12 novel genes of which only three were finally predicted to be imprinted. A similar approach was used by Babak et *al.* [47] to build a map of imprinted genes in 9.5 dpc mouse embryos. The design for the cDNA synthesis from strand-specific total RNA allowed including intronic regions and non-polyadenylated transcripts. This approach identified 80% of the already known imprinted genes and further confirmed the existence of six novel imprinted genes. In addition, this map highlighted the incomplete characterization of known imprinted loci by revealing novel imprinted transcripts in these regions (including non-coding RNAs) [47].

Besides transcriptome analysis, mono-allelic gene expression can be assessed by investigating allele-specific binding of transcription factors. In an innovative approach, Maynard *et al*. [48] investigated allele-specific binding of RNA polymerase II (RNAP) in human lung fibroblasts. They designed a ChIP-SNP approach in which regions precipitated with an anti-RNAP antibody were analysed using a SNP genotyping microarray. By this means, in addition to identifying already known imprinted genes they demonstrated that a microRNA cluster adjacent to *MEG3* and known to be imprinted in the sheep and mouse is also imprinted in humans [48].

SCREENS BASED ON EPIGENETIC FEATURES

Screens based on the detection of differences in the epigenetic marks (i.e. DNA methylation, histone modifications) that are present at the maternal and paternal allele of a given gene also provide a relevant strategy to identify novel imprinted loci. Allelic epigenetic differences are evident at ICRs, key regions for the regulation of imprinting (Figure 1). ICRs are marked by DNA methylation acquired on either the maternal or the paternal allele in the female or male germ line (constituting a germline DMR) [12]. In addition to DNA methylation, ICRs are also marked by differential histone modifications in somatic cells (see below) [49–52]. The main advantage of screens focused on the identification of germline DMRs/ICRs is that they can be carried out in any cell type because their epigenetic marks are maintained throughout development and adult life, regardless of the expression status of the imprinted genes (Figure 1). On the other hand, ICRs are discrete elements that often control entire imprinted clusters which cover hundreds to thousands of kilobases and can contain more than 10 genes (Figure 1). Therefore, such screens are more likely to identify chromosomal regions with several imprinted genes rather than individual imprinted genes. The concomitant identification of differentially methylated somatic regions (somatic DMRs) at promoters could more readily reveal individual imprinted genes. However, these features are more an exception than a generality as most of the promoters of imprinted genes identified so far do not show allelic differences in DNA methylation (Figure 1).

As in transcriptomic analyses, the last years witnessed a tremendous technological progress that facilitates genome-wide profiling of epigenetic features in an unbiased way. Methylated DNA isolated by precipitation with anti-5mC antibodies (MeDIP assay) or Methyl-CpG Binding proteins (MIRA assay) can be further analysed through microarray hybridization or deep sequencing methods [53–55]. Another promising approach is bisulphite-sequencing (BS-seq) in which methylation-dependent bisulphite conversion of DNA (to differentiate between methylated and unmethylated cytosines) is combined with high-throughput sequencing to quantitatively map DNA methylation at single-base resolution in a whole genome [56]. Similarly, genome-wide profiling of histone modifications is obtained by specific chromatin immunoprecipitation followed by tiling array hybridization (ChIP-chip) or deep sequencing (ChIP-seq) [57].

As described below, these technical breakthroughs offer powerful means to undertake a

systematic study in order to identify novel imprinted loci and should provide major insights in the coming years.

Screens based on differential methylation

The first successful screens, such as restriction landmark genomic scanning (RLGS) [58] and methylation-sensitive representational difference analysis (Me-RDA) [59], applied an approach based on the use of methylation sensitive restriction enzymes [60–62]. This strategy was then used to map methylated CpG islands in the human genome in order to identify new imprinted genes [63, 64]. Indeed, most, although not all, ICRs share also CpG island features. The rationale for this strategy is based on the observation that the majority of CpG islands present in the mouse and human genomes are normally unmethylated, thus one can expect that a relevant proportion of the methylated ones would be located at imprinted loci.

A major drawback of this approach comes from the initial digestion with a site-specific restriction enzyme that drastically reduces the coverage of the analysis.

Genome-wide maps of methylated cytosines (i.e. methylomes) obtained by unbiased approaches should enable the systematic identification of DMRs. Of particular interest is the recent demonstration that the BS-Seq approach is applicable to the complex human [65] and mouse [66] genomes. Further developments in which such approaches will be combined with SNP data would be extremely useful to efficiently identify new candidate DMRs associated with imprinted loci. Alternatively, genome-wide DNA methylation profiling can also be used to compare the normal genome to those known to present with methylation imprint defects. This method was recently successfully applied in a study [67] in which genome-wide CpG methylation analysis was carried out using blood samples from a patient with multiple imprinting defects and normal controls. Following bisulphite treatment, DNA was hybridized to a commercially available array developed to assay CpG methylation in more than 14 000 genes. In addition to confirming in this patient the hypomethylation of known 'imprinted' DMRs, this study identified new candidate DMRs. One of these regions was associated with *RB1* which was then demonstrated to be imprinted in humans [67].

Screen based on chromatin features—a 'coming of age' approach

Several locus-specific studies on histone modifications, which were further supported by genome-wide analyses, revealed the existence of an ICR-specific chromatin signature. Specifically, the DNA methylated allele is consistently associated with histone marks [i.e. tri-methylation on lysine 9 of histone H3 (H3K9me3), trimethylation on lysine 20 of histone H4 (H4K20me3)] that are characteristic of repressive chromatin. By contrast, the unmethylated allele is characterized by histone modifications [i.e. H3/H4 acetylation and di- and tri-methylation of lysine 4 of histone H3 (H3K4me2/me3)], which are typical of permissive chromatin [51, 68, 69] (Figure 2). Based on availability of Chip-seq-derived genome-wide profiling of these histone marks in human and mouse cell lines [e.g. see [69, 70] and http://www.broadinstitute.org/science/projects/epigenomics/chip-seq-data], machine-learning algorithms can be applied to these databases for the systematic identification of regions characterized by this specific chromatin signature [71, 72]. In addition, the demonstration that the chromatin features identified by ChIP-seq can be read in an allele-specific manner by using SNPs [69] allows assigning each chromatin modification to a specific parental allele, thus facilitating the identification of new candidate ICRs.

Beside the identification of putative ICRs, genome-wide chromatin signatures can also be used to identify regions that are allelically transcribed. This has been shown by Mikkelsen *et al.* [69] who, following Chip-seq combined with informative SNPs for allele discrimination, observed allelic imbalance of H3K36me3, a mark associated with transcription elongation, at several imprinted genes or microRNAs.

Further improvement of these approaches can be obtained by crossing information obtained by genome-wide profiling of different epigenetic features. By using a custom microarray that covered several mouse imprinted chromosomal regions ('mouse imprinted array') Dindot *et al.* showed that known ICRs can be identified as regions that have both a specific DNA methylation profile and overlapping H3K4me3 and H3K9me3 modifications. By this means they identified 11 new candidate ICRs [73]. By using a similar approach, Wen *et al.* [74] proposed that the overlapping presence of H3K4me2, DNA methylation and CTCF binding

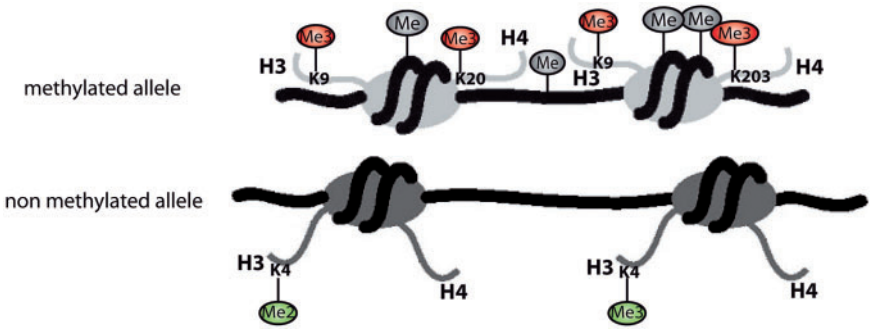

Figure 2: Epigenetic signature of ICRs. The methylated allele is consistently associated with histone marks characteristics of repressive chromatin, including H3K9me3, H4K20me3. By contrast on the opposite parental allele, absence of DNA methylation is associated with H3K4me2 and H3K4me3, hallmarks of permissive chromatin.

sites (a protein known to have regulatory function in imprinting) identify imprinted regions in human T cells and immortalized lymphoblast cell lines.

CONCLUSIONS

The characterization of more than 110 of imprinted genes during the last 20 years has greatly expanded our understanding of the regulation and function of genomic imprinting. For instance, beside a role in growth and development, their involvement in neurological processes and in behaviour also has been highlighted [11]. Nonetheless, the current list of imprinted genes, and the full biological meaning of genomic imprinting, is probably incomplete.

Although estimates of the total number of imprinted transcripts rose up to 2100 [8], combined studies suggest that their actual number should not exceed few hundreds. For instance, transcriptome approaches using mouse neonatal brain samples, one of the main 'target' tissues for imprinting with placenta [7], identified only three novel protein-coding imprinted genes [46]. In addition, another comprehensive transcriptomic analysis conducted on whole 9.5 dpc mouse embryos further supported the idea that most of the yet-uncharacterized imprinted genes are transcripts and non-coding RNAs that are localized mainly in already known imprinted domains [47].

An important issue is also to establish to which extend imprinting is present in the human genome. Although the imprinting status of some genes identified in mouse is conserved in the human orthologues, both mouse- and human-specific imprinting is documented [6, 7]. This means that although the data gathered in the mouse can be used as a frame to select which novel imprinted genes to test in humans, genome-wide screens completely focused on the human genome are also required in order to precisely identify all the imprinted genes in humans.

A systematic investigation of novel imprinted gene candidates in the mouse and human genomes is now virtually feasible. The remarkable progress made in genome-wide profiling of histone modifications and DNA methylation, combined with SNP information, provides powerful tools to screen genomic regions for chromatin signatures known to constitutively mark imprinted regions. In addition, the establishment of comprehensive epigenetic maps of mutants of key regulators of imprinting will be an extremely promising approach for a systematic identification of imprinted genes. For instance, the progeny of mouse females deficient for *DNMT3L*, in which we can expect methylation defects only in imprinted genes [14, 17], as well as human complete hydatidiforme moles or ovarian teratomas, which contain only one parental DNA set, are relevant material to establish such maps. In parallel, transcriptome sequencing offers an extremely informative and powerful approach to carry out systematic surveys on imprinted gene expression in various tissues.

A further obvious development relies on combining these different approaches. Computational-assisted crossing of databases obtained through epigenome- and transcriptome-wide analyses should offer the unique opportunity of establishing the full list of 'high-confidence' candidate imprinted genes. Nonetheless, it is important to keep in mind that any

identified candidate necessarily requires to be experimentally validated as an imprinted gene through classical approaches.

Key Points

- The current list of imprinted genes is probably incomplete, thus limiting our knowledge on imprinting biological functions.
- Numerous screens have been undertaken to identify novel imprinted transcripts, utilizing specific DNA sequence motifs, differential allelic-expression and specific epigenetic signatures.
- The comparison of data in extensive epigenome- and transcriptome-wide databases, combined with SNP information, will facilitate the identification of all imprinted genes in the genome.
- Only after exhaustive attempts to identify all imprinted genes using genome-wide methods, can we establish to which extent imprinting is present in the human genome.

Acknowledgements
We thank Michael Weber, Dave Monk and Robert Feil for advices and critical reading of the manuscript.

FUNDING

Work carried out by the authors was supported by the Association Pour la Recherche sur le Cancer (grant 'Subvention fixe' no. 4980; 2010) and la Ligue contre le Cancer (grant 'Subvention comité Herault'; 2010).

References

1. Barton SC, Surani MA, Norris ML. Role of paternal and maternal genomes in mouse development. *Nature* 1984;**311**:374–6.
2. McGrath J, Solter D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 1984;**37**:179–83.
3. Barlow DP, Stoger R, Herrmann BG, *et al*. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature* 1991;**349**:84–7.
4. Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. *Nature* 1991;**351**:153–5.
5. DeChiara TM, Robertson EJ, Efstratiadis A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 1991;**64**:849–59.
6. 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.
7. Morison IM, Ramsay JP, Spencer HG. A census of mammalian imprinting. *Trends Genet* 2005;**21**:457–65.
8. Barlow DP. Gametic imprinting in mammals. *Science* 1995;**270**:1610–3.
9. Nikaido I, Saito C, Mizuno Y, *et al*. Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome Res* 2003;**13**:1402–9.
10. Smith FM, Garfield AS, Ward A. Regulation of growth and metabolism by imprinted genes. *Cytogenet Genome Res* 2006;**113**:279–91.
11. Wilkinson LS, Davies W, Isles AR. Genomic imprinting effects on brain development and function. *Nat Rev Neurosci* 2007;**8**:832–43.
12. Arnaud P, Feil R. Epigenetic deregulation of genomic imprinting in human disorders and following assisted reproduction. *Birth Defects Res C Embryo Today* 2005;**75**:81–97.
13. Feinberg AP, Cui H, Ohlsson R. DNA methylation and genomic imprinting: insights from cancer into epigenetic mechanisms. *Semin Cancer Biol* 2002;**12**:389–98.
14. Bourc'his D, Xu GL, Lin CS, *et al*. Dnmt3L and the establishment of maternal genomic imprints. *Science* 2001;**294**:2536–9.
15. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. *Nature* 1993;**366**:362–5.
16. Kaneda M, Okano M, Hata K, *et al*. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 2004;**429**:900–3.
17. Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 2002;**129**:1983–93.
18. Cattanach BM, Kirk M. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 1985;**315**:496–8.
19. Williamson CM, Blake A, Thomas S, *et al*. *World Wide Web Site - Mouse*. MRC Harwell: Oxfordshire, 2009.
20. Greally JM. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci USA* 2002;**99**:327–32.
21. Ke X, Thomas NS, Robinson DO, Collins A. A novel approach for identifying candidate imprinted genes through sequence analysis of imprinted and control genes. *Hum Genet* 2002;**111**:511–20.
22. Ke X, Thomas NS, Robinson DO, Collins A. The distinguishing sequence characteristics of mouse imprinted genes. *Mamm Genome* 2002;**13**:639–45.
23. Wang Z, Fan H, Yang HH, *et al*. Comparative sequence analysis of imprinted genes between human and mouse to reveal imprinting signatures. *Genomics* 2004;**83**:395–401.
24. Wood AJ, Roberts RG, Monk D, *et al*. A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS Genet* 2007;**3**:e20.
25. Luedi PP, Hartemink AJ, Jirtle RL. Genome-wide prediction of imprinted murine genes. *Genome Res* 2005;**15**:875–84.
26. Luedi PP, Dietrich FS, Weidman JR, *et al*. Computational and experimental identification of novel human imprinted genes. *Genome Res* 2007;**17**:1723–30.
27. Ruf N, Bahring 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.

28. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. *Science* 2007;**318**:1136–40.
29. Pant PV, Tao H, Beilharz EJ, *et al*. Analysis of allelic differential expression in human white blood cells. *Genome Res* 2006;**16**:331–9.
30. 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.
31. Kaneko-Ishino T, Kuroiwa Y, Miyoshi N, *et al*. Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. *Nat Genet* 1995;**11**:52–9.
32. Mizuno Y, Sotomaru Y, Katsuzawa Y, *et al*. Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. *Biochem Biophys Res Commun* 2002;**290**:1499–505.
33. Piras G, El Kharroubi A, Kozlov S, *et al*. Zac1 (Lot1), a potential tumor suppressor gene, and the gene for epsilon-sarcoglycan are maternally imprinted genes: identification by a subtractive screen of novel uniparental fibroblast lines. *Mol Cell Biol* 2000;**20**:3308–15.
34. Kobayashi H, Yamada K, Morita S, *et al*. Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. *Genomics* 2009;**93**:461–72.
35. 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.
36. Cattanach BM, Beechey CV. Genomic imprinting in the mouse: possible final analysis. In: Reik W, Surani A (eds). *Genomic Imprinting*. New York: Oxford University Press Inc., 1997:118–45.
37. Choi JD, Underkoffler LA, Collins JN, *et al*. Microarray expression profiling of tissues from mice with uniparental duplications of chromosomes 7 and 11 to identify imprinted genes. *Mamm Genome* 2001;**12**:758–64.
38. Choi JD, Underkoffler LA, Wood AJ, *et al*. A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG island. *Mol Cell Biol* 2005;**25**:5514–22.
39. Menheniott TR, Woodfine K, Schulz R, *et al*. Genomic imprinting of Dopa decarboxylase in heart and reciprocal allelic expression with neighboring Grb10. *Mol Cell Biol* 2008;**28**:386–96.
40. Ogata T, Kagami M, Ferguson-Smith AC. Molecular mechanisms regulating phenotypic outcome in paternal and maternal uniparental disomy for chromosome 14. *Epigenetics* 2008;**3**:181–7.
41. Schulz R, Menheniott TR, Woodfine K, *et al*. Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. *Nucleic Acids Res* 2006;**34**:e88.
42. Lo HS, Wang Z, Hu Y, *et al*. Allelic variation in gene expression is common in the human genome. *Genome Res* 2003;**13**:1855–62.
43. Bjornsson HT, Albert TJ, Ladd-Acosta CM, *et al*. SNP-specific array-based allele-specific expression analysis. *Genome Res* 2008;**18**:771–9.
44. Pollard KS, Serre D, Wang X, *et al*. A genome-wide approach to identifying novel-imprinted genes. *Hum Genet* 2008;**122**:625–34.
45. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;**10**:57–63.
46. Wang X, Sun Q, McGrath SD, *et al*. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One* 2008;**3**:e3839.
47. Babak T, Deveale B, Armour C, *et al*. Global survey of genomic imprinting by transcriptome sequencing. *Curr Biol* 2008;**18**:1735–41.
48. Maynard ND, Chen J, Stuart RK, *et al*. Genome-wide mapping of allele-specific protein-DNA interactions in human cells. *Nat Methods* 2008;**5**:307–9.
49. Delaval K, Govin J, Cerqueira F, *et al*. Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J* 2007;**26**:720–9.
50. Fournier C, Goto Y, Ballestar E, *et al*. Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J* 2002;**21**:6560–70.
51. Henckel A, Nakabayashi K, Sanz LA, *et al*. Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals. *Hum Mol Genet* 2009;**18**:3375–83.
52. Yang Y, Li T, Vu TH, *et al*. The histone code regulating expression of the imprinted mouse Igf2r gene. *Endocrinology* 2003;**144**:5658–70.
53. Guibert S, Forne T, Weber M. Dynamic regulation of DNA methylation during mammalian development. *Epigenomics* 2009;**1**:81–9.
54. Huang YW, Huang TH, Wang LS. Profiling DNA methylomes from microarray to genome-scale sequencing. *Technol Cancer Res Treat* 2010;**9**:139–47.
55. Weber M, Davies JJ, Wittig D, *et al*. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005;**37**:853–62.
56. Cokus SJ, Feng S, Zhang X, *et al*. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 2008;**452**:215–9.
57. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009;**10**:669–80.
58. Hayashizaki Y, Shibata H, Hirotsune S, *et al*. Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. *Nat Genet* 1994;**6**:33–40.
59. Kelsey G, Bodle D, Miller HJ, *et al*. Identification of imprinted loci by methylation-sensitive representational difference analysis: application to mouse distal chromosome 2. *Genomics* 1999;**62**:129–38.
60. Maeda N, Hayashizaki Y. Genome-wide survey of imprinted genes. *Cytogenet Genome Res* 2006;**113**:144–52.
61. Peters J, Beechey C. Identification and characterisation of imprinted genes in the mouse. *Brief Funct Genomic Proteomic* 2004;**2**:320–33.
62. Smith RJ, Arnaud P, Kelsey G. Identification and properties of imprinted genes and their control elements. *Cytogenet Genome Res* 2004;**105**:335–45.
63. Strichman-Almashanu LZ, Lee RS, Onyango PO, *et al*. A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res* 2002;**12**:543–54.
64. Yamada Y, Watanabe H, Miura F, *et al*. A comprehensive analysis of allelic methylation status of

CpG islands on human chromosome 21q. *Genome Res* 2004; **14**:247–66.

65. Lister R, Pelizzola M, Dowen RH, *et al*. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009;**462**:315–22.
66. Popp C, Dean W, Feng S, *et al*. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 2010;**463**:1101–5.
67. Kanber D, Berulava T, Ammerpohl O, *et al*. The human retinoblastoma gene is imprinted. *PLoS Genet* 2009;**5**: e1000790.
68. McEwen KR, Ferguson-Smith AC. Distinguishing epigenetic marks of developmental and imprinting regulation. *Epigenetics Chromatin* 2010;**3**:2.
69. Mikkelsen TS, Ku M, Jaffe DB, *et al*. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007;**448**:553–60.
70. Barski A, Cuddapah S, Cui K, *et al*. High-resolution profiling of histone methylations in the human genome. *Cell* 2007;**129**:823–37.
71. Hon G, Ren B, Wang W. ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput Biol* 2008;**4**:e1000201.
72. Hon G, Wang W, Ren B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol* 2009;**5**:e1000566.
73. Dindot SV, Person R, Strivens M, *et al*. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res* 2009;**19**:1374–83.
74. 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.