A Survey of Imprinted Gene Expression in Mouse Trophoblast Stem Cells

J. Mauro Calabrese,*† Megan D. Schertzer,*† Della Yee,†§ and Terry Magnuson†‡,1
*Department of Pharmacology, †Lineberger Comprehensive Cancer Center, ‡Department of Genetics, and §Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599

ABSTRACT  Several hundred mammalian genes are expressed preferentially from one parental allele as the result of a process called genomic imprinting. Genomic imprinting is prevalent in extra-embryonic tissue, where it plays an essential role during development. Here, we profiled imprinted gene expression via RNA-Seq in a panel of six mouse trophoblast stem lines, which are ex vivo derivatives of a progenitor population that gives rise to the placental tissue of the mouse. We found evidence of imprinted expression for 48 genes, 31 of which had been described previously as imprinted and 17 of which we suggest as candidate imprinted genes. An equal number of maternally and paternally biased genes were detected. On average, candidate imprinted genes were more lowly expressed and had weaker parent-of-origin biases than known imprinted genes. Several known and candidate imprinted genes showed variability in parent-of-origin expression bias between the six trophoblast stem cell lines. Sixteen of the 48 known and candidate imprinted genes were previously or newly annotated noncoding RNAs and six encoded for a total of 60 annotated microRNAs. Pyrosequencing across our panel of trophoblast stem cell lines returned levels of imprinted expression that were concordant with RNA-Seq measurements for all eight genes examined. Our results solidify trophoblast stem cells as a cell culture-based experimental model to study genomic imprinting, and provide a quantitative foundation upon which to delineate mechanisms by which the process is maintained in the mouse.

Genomic imprinting is an epigenetic process initiated during mammalian gametogenesis, which results in preferential expression of genes from one parentally inherited allele over the other. More than one hundred fifty imprinted genes have been identified in mammals (Williamson et al. 2013). As a class, these genes play important roles in development, growth, metabolism, and social adaptation (Ferguson-Smith 2011; Garfield et al. 2011). Defects in imprinting can cause cancer, in the form of Wilm’s tumor, and other human diseases, including Angelman, Prader-Willi, Beckwith-Wiedemann, and Silver-Russell syndromes (Butler 2009). Faithful maintenance of imprinting also plays an important role in reprogramming and maintenance of stem cell identity (Zacharek et al. 2011; Stadtfeld et al. 2012).

Genomic imprinting is critical and prevalent in the placenta, consistent with its essential role in development. Parthenogenetic embryos with no contribution from the paternal genome die early in gestation with severe defects in the trophectoderm, the pool of cells that give rise to the placenta (Surani and Barton 1983). Genomic imprinting is required in placental tissue as early as, or before, embryonic gestational day 6.5 (Clarke et al. 1988). Moreover, much of the tissue-specific imprinting that is known to occur is found in the placenta (Wang et al. 2011b; Prickett and Oakey 2012; Wang et al. 2013b; Court et al. 2014), again underscoring its importance for extraembryonic development.

The study of genomic imprinting has helped to define paradigms of epigenetic regulation and long noncoding RNA (lncRNA) function...
in mammals (Barlow 2011; Ferguson-Smith 2011). For example, parent-of-origin—specific DNA methylation deposited at imprinted control regions during gametogenesis is a master regulator of imprinted states. Accordingly, genomic imprinting has served as an important model to understand the deposition, propagation, and biological function of DNA methylation in development and organismal homeostasis (Kelsey and Feil 2013). In addition to DNA methylation, several imprinted genes also require IncRNAs to propagate their allelic epigenetic states, or, are themselves IncRNAs (Lee and Bartolomei 2013). Indeed, some of the earliest IncRNAs identified, H19 and Kcnqot1, are imprinted and were discovered because of their strong associations with human disease (Lee and Bartolomei 2013). The study of imprinted IncRNAs will continue to provide important paradigms as newly described IncRNAs emerge as essential regulators of diverse physiological processes.

Considering the importance of genomic imprinting in health and disease, its prevalence in placental tissue, and its paradigmatic role in defining mechanisms of epigenetic regulation and IncRNA function, it remains a high priority to develop appropriate experimental models to study the process in extra-embryonic tissues. Mouse trophoblast stem cells (TSCs) offer one such experimental framework. TSCs are ex vivo derivatives of the tropho-dermal stem cell population that mediates implantation and gives rise to the placenta, and they provide a renewable, extra-embryonic—derived cell population free of maternal tissue contamination (Quinn et al. 2006). Furthermore, they are easily propagated in culture. As a result, TSCs are amenable to large-scale genomic and biochemical studies, and their transcriptional outputs can be modified via overexpression, knockdown, or precision genome-editing approaches.

Here, we profiled allele-specific gene expression via RNA-Seq in a panel of six F1-hybrid mouse TSC lines. We detected parent-of-origin (PO) biased expression of 48 genes, an equal number of which were expressed with maternal and paternal biases, respectively. Thirty-one of these were known imprinted genes, whereas 17 had not been previously reported to exhibit PO expression bias and could be considered candidate imprinted genes. Sixteen of the 48 PO-biased genes were known or putative IncRNAs. Further, six of the known imprinted genes expressed in TSCs encode for a total of 60 known microRNAs. PO biases in gene expression detected via RNA-Seq were concordant with those detected via pyrosequencing for eight genes examined across six profiled TSC lines. Our results provide a quantitative foundation upon which to dissect mechanisms that underpin PO biased gene expression in mouse TSCs.

**MATERIALS AND METHODS**

**TSC derivation and culture**

TSCs were derived and propagated in an undifferentiated state using protocols described in (Quinn et al. 2006).

**RNA-Seq**

Before RNA extraction with Trizol, TSC lines were passaged twice off of irradiated feeder cells. For these two passages, TSCs were cultured in 70% feeder-conditioned media plus growth factors, as described in (Quinn et al. 2006). At each passage after trypsinization, TSC suspensions were pre-plated for 30 min to deplete irradiated feeder populations. CB.1 and BC.1 RNA-Seq data were collected in Calabrese et al. (2012). cDNA libraries for CB.2, CB.3, BC.2, and BC.3 TSC lines were prepared in this work, from 4 μg of total TSC RNA using Kapa Biosystem’s Stranded mRNA-Seq kit, which maintains strand information and enriches for poly-adenylated transcripts via oligo dT bead purification. cDNA libraries prepared from CB.2, CB.3, BC.2, and BC.3 TSC lines were sequenced once on Illumina’s HiSeq and once on Illumina’s NextSeq500 instruments, respectively, and data were pooled per cell line to derive final allele-specific expression ratios. Total read counts obtained per TSC line were as follows: [CB.1, 69,788,067]; [CB.2, 123,636,335]; [CB.3, 135,130,738]; [BC.1, 60,678,397]; [BC.2, 123,714,166]; [BC.3, 150,370,343]. Sequence data collected as part of the present study were deposited in Gene Expression Omnibus (GEO) database under accession number GSE63968.

**Allele-specific read counts**

Allele-specific read counts per gene were determined as in Calabrese et al. (2012). In brief, Cast single-nucleotide polymorphisms (SNPs) from Keane et al. (2011) were substituted into their corresponding mm9 genomic positions to create an in silico Cast genome, and SNP-overlapping reads that uniquely aligned to either the B6 (mm9) or Cast genomes were retained (Kent et al. 2002). A nonredundant list of mouse genes was annotated from the set of UCSC Known Genes as in Calabrese et al. (2012) via the use of the longest exemplar per gene to count allele-specific expression. In addition to The University of California Santa Cruz (UCSC) Known Genes, uniquely aligning RNA-Seq reads from CB.1 and BC.1 TSCs that did not match in strand with, or were not located within ±5 kb of, any UCSC Known Gene, were selected for clustering to approximate newly annotated transcriptional units. Units reported represent strand-matched reads falling within ±5 kb of each other. Allele-specific counts represent the total number of SNP-overlapping reads that uniquely mapped between the start and end of each gene or transcriptional unit, including intronic regions.

**Calculation of significance of PO bias**

Allelic read counts for all autosomal genes were imported into edgeR and normalized using edgeR’s counts per million (CPM) metric. Only genes whose normalized allele-specific counts summed to more than 1 CPM in each of the six profiled TSC lines were tested for differential allelic expression. Differential expression between Cast and B6 alleles was tested separately in the CB.x and BC.x TSC lines, such that each group of F1-hybrid TSC lines was represented by three biological replicates: CB.1, 0.2, and 0.3 for the CB.x lines, and BC.1, 0.2, and 0.3 for the BC.x lines. Differential expression between Cast and B6 alleles within each F1-hybrid group was tested via edgeR’s generalized linear model likelihood ratio test, and P-values from both tests were adjusted to false discovery rates using the Benjamini-Hochberg method (Robinson et al. 2010). Genes exhibiting PO biases with false discovery rates scores of ≤ 0.05 in both CB.x and BC.x cell lines were considered to be significantly biased.

**Calculation of total gene expression levels**

Total (i.e., allele- nonspecific) gene expression levels were approximated for all UCSC Known Genes in each TSC line using the Tophat and Cufflinks algorithms and are reported using the reads per kilobase per million aligned reads (RPKM) metric; for newly annotated transcription units (e.g., the Tsci transcripts), exonic coordinates were not clearly apparent, and thus reads that matched in strand and fell between the start and end of the unit were used to calculate RPKM via custom scripts.

**Pyrosequencing**

PCR primers for individual pyrosequencing assays were designed to amplify a less than 200 base pair exonic region surrounding a known SNP for each of eight genes (Table 1). Sequencing primers were either
directly adjacent to the SNP or one base pair removed. To perform pyrosequencing assays, 5 μg of RNA from each TSC line was reverse transcribed with Superscript III (Invitrogen). PCR amplification from cDNA was performed with Apex Taq DNA Polymerase (Genesee Scientific) and the cycle number shown in Table 1 using the following PCR conditions: 95° for 30 sec, 56° for 30 sec, and 72° for 30 sec. The PyroMark Q96 MD Pyrosequencer (Biotaq, AB), PyroMark Gold Q96 CDT Reagents (QIAGEN), and Streptavidin Sepharose beads (GE Healthcare) were used for pyrosequencing. Quantification of allele-specific expression was performed using PyroMark Q96 MD software. Box and whisker diagrams were generated using matplotlib version 1.4.2.

RESULTS AND DISCUSSION
Detection and validation of autosomal PO expression bias in TSCs
In a previous study, we generated a panel of F1-hybrid mouse TSC lines that were used to measure molecular properties of the inactive X-chromosome (Calabrese et al. 2012). These TSCs were derived from reciprocal crosses between two diverse, inbred mouse strains, CAST/EiJ (Cast) and C57BL/6J (B6). Using a high confidence, validated set of ~18 million informative SNPs from (Keane et al. 2011), allele-specific gene expression can be measured accurately in these cells by the counting of SNPs contained within uniquely mapping high-throughput sequencing reads (Calabrese et al. 2012).
To measure allelic biases in TSC gene expression, we analyzed strand-specific RNA-Seq data collected from six of these reciprocally derived F1-hybrid TSC lines: three lines were derived from a Cast mother and B6 father (referred to as “CB.x” lines) and three lines from a B6 mother and Cast father (referred to as “BC.x” lines). F1-hybrid TSC lines were grouped on the basis of their respective parentage, and
Table 2 Known imprinted genes expressed with significant PO bias in TSCs

| Gene ID | Chr | CB.1 | CB.2 | CB.3 | BC.1 | BC.2 | BC.3 | CBp | BCp | B | SNPs | RPKM |
|---------|-----|------|------|------|------|------|------|-----|-----|----|------|------|
| AK076687a | chr2 | 5.3 (10, 180) | 4.5 (26, 557) | 0.4 (2, 551) | 1.1 (2, 185) | 1.8 (5, 269) | 0.7 (3, 439) | 5E-37 | 6E-62 | P | 226 | 1.03 |
| Sfmbt2 | chr2 | 2.3 (144, 6233) | 8.4 (790, 8599) | 1.6 (290, 17004) | 1.0 (79, 7694) | 1.5 (299, 19439) | 1.0 (217, 22508) | 8E-31 | 6E-62 | P | 1092 | 78.8 |
| Nespas | chr2 | 0.0 (0, 251) | 0.0 (0, 251) | 2.2 (5, 223) | 0.0 (0, 88) | 4.0 (8, 191) | 0.6 (1, 161) | 4E-51 | 2E-30 | P | 34 | 2.58 |
| Phe17 | chr3 | 32.4 (665, 1387) | 30.3 (2099, 4819) | 46.8 (1568, 1786) | 13.8 (130, 809) | 24.7 (1483, 4515) | 15.5 (820, 4457) | 2E-03 | 2E-18 | P | 321 | 48.2 |
| Sgce | chr6 | 4.5 (11, 231) | 8.8 (78, 809) | 6.6 (47, 666) | 2.9 (4, 135) | 6.3 (60, 890) | 3.7 (33, 870) | 1E-68 | 1E-52 | P | 319 | 19.9 |
| Peg10 | chr6 | 0.1 (18, 12098) | 0.2 (226, 149701) | 0.3 (310, 123681) | 0.1 (4, 3763) | 0.5 (507, 150895) | 6.0 (4951, 78040) | 0E+00 | 4E-10 | P | 29 | 793 |
| Mest | chr7 | 0.0 (0, 16) | 1.7 (14, 829) | 0.3 (11, 3209) | 0.4 (2, 447) | 0.5 (13, 2780) | 1.1 (71, 6602) | 3E-117 | 4E-105 | P | 42 | 65.1 |
| Peg3 | chr7 | 0.2 (7, 4413) | 0.1 (21, 24916) | 0.2 (44, 19674) | 0.5 (17, 3529) | 19.6 (4032, 16588) | 41.4 (10215, 14486) | 1E-302 | 2E-02 | P | 107 | 75.2 |
| Nespasa | chr7 | 32.4 (665, 1387) | 30.3 (2099, 4819) | 46.8 (1568, 1786) | 13.8 (130, 809) | 24.7 (1483, 4515) | 15.5 (820, 4457) | 2E-03 | 2E-18 | P | 321 | 48.2 |
| Phlda2 | chr7 | 100.0 (120, 0) | 100.0 (4921, 1) | 99.3 (3279, 24) | 97.6 (244, 92) | 98.7 (1211, 16) | 98.7 (155, 72) | 2E-55 | 1E-06 | M | 290 | 13.6 |
| Gabt1 | chr7 | 15.3 (342, 1896) | 18.3 (1745, 7809) | 14.8 (1624, 9329) | 12.9 (1275, 7163) | 16.8 (1557, 7804) | 1E-14 | 2E-04 | M | 9 | 3.30 |
| Slc22a18 | chr7 | 88.5 (46, 6) | 89.8 (264, 30) | 88.7 (235, 30) | 92.9 (13, 1) | 76.5 (436, 141) | 69.6 (385, 168) | 1E-40 | 1E-04 | M | 141 | 6.75 |
| Pde10a | chr17 | 90.9 (251, 25) | 88.3 (878, 116) | 85.6 (600, 101) | 97.6 (290, 7) | 98.7 (3325, 5) | 98.9 (11062, 8) | 2E-89 | 4E-129 | M | 109 | 85.3 |
| Mirga | chr17 | 100.0 (159, 0) | 100.0 (10, 0) | 93.1 (54, 4) | 100.0 (772, 0) | 100.0 (716, 0) | 99.8 (2007, 4) | 1E-24 | 4E-150 | M | 119 | 11.4 |

Gene ID and chromosomal location are listed. For each CB.x and BC.x cell line, the percent of allele-specific reads mapping to the maternally inherited chromosome is shown. For each CB.x and BC.x cell line, the percent of allele-specific reads mapping to the maternally inherited chromosome is shown. In that same column, the absolute number of maternal and paternal SNP-overlapping read counts detected in each TSC line, respectively, are shown in parentheses. P-values for corresponding PO biases in CB.x and BC.x cell lines were corrected for multiple testing via Benjamini Hochberg, and are listed under “Cbp” and “Bcp” columns, respectively. B, parental direction of PO bias: P, paternally biased; M, maternally biased; NS, not significantly biased. PO, parent-of-origin; TSC, trophoblast stem cell; SNPs, single-nucleotide polymorphisms—the number of distinct SNPs covered by RNA-Seq reads for the respective transcript. RPKM, average allele-specific RPKM (reads per kilobase of transcript per million aligned reads) of each respective transcript across the 6 profiled TSC lines; lncRNA, long noncoding RNA.

a Known or putative lncRNAs.
significant allelic biases in gene expression in each group were detected using edgeR and corrected for multiple testing via the Benjamini-Hochberg method (Robinson et al. 2010) (Figure 1). To increase robustness of our significance calls, genes were only considered for allelic analysis if the sum of their allele-specific read counts was >1 CPM in all six profiled TSC lines. Under this CPM cutoff, 13,593 genes were eligible for allelic expression analysis (Supporting Information, Table S1).

In total, we detected 48 genes expressed with significant PO bias in TSCs. Twenty-four genes were expressed with a maternal bias and 24 with a paternal bias. This equal representation of PO bias between TSCs was annotated previously as imprinted in the mouse placenta, where paternally biased genes predominate (Wang et al. 2013b). PO expression bias for the remaining 17 genes has not been previously described (Table 3). Notable known imprinted genes with significant PO expression biases in TSCs included the Kcnq1ot1 imprinted IncRNA and many of its nearby target genes (Cdkn1c, Cd81, Phlda2, Scc2a18, Tsc4), the Airn imprinted IncRNA and two of its nearby target genes (Igf2r and Sfmbt2), H19, Grb10, Meg3, Mrg, Gab1 (Okae et al. 2012), and Slc22a18 and its antisense noncoding transcript, AK076687 (Wang et al. 2011a). Another PO-biased gene, annotated as D7ertd715e in the mouse, is syntenic to a complex series of IncRNA transcripts that originate from the imprinted Prader-Willi locus on human chromosome 15, and may be a mouse homolog, or it may be an 3' extension of the neighboring imprinted gene, Surprn. The D7ertd715e transcript was recently reported to be imprinted in trophoblast cells derived from horse/donkey F1-hybrids (Wang et al. 2013b).

An additional 41 known imprinted genes present in MRC Harwell’s Imprinting Resource were expressed in TSCs with enough allelic coverage to pass our threshold for analysis but were not detected as significantly PO biased (Williamson et al. 2013) (Table S1). Many of these genes, such as Wt1, Ube3a, Rasgrf1, and Zdbf2, were neutrally biallelic across the profiled TSC lines, and at least four, Pon2, Klf14, Atp10a, and Art5, were expressed with significant strain-of-origin bias (as opposed to a PO bias), underscoring the tissue-specificity with which imprinted gene expression is known to occur (Prickett and Oakey 2012).

To gain a sense of the accuracy with which our RNA-Seq analysis pipeline detected PO biases in TSCs, we re-measured PO expression for eight genes using QIAGEN’s PyroMark pyrosequencing assay. These eight genes included three known imprinted genes expressed with significant PO bias in TSCs (Igf2r, H19, and Gab1), one known imprinted gene whose PO expression bias in TSCs was not called as significant in our analysis (Igf2), and four PO biased genes that to our knowledge have not been previously reported as imprinted. Pyrosequencing primers were designed around single informative SNPs contained within each gene and assays were performed per gene in each of the six TSC lines profiled for RNA-Seq. In all 48 cases (eight genes in six TSC lines), allelic biases determined via RNA-Seq and pyrosequencing were concordant (Figure 2). This high level of concordance mirrors that observed in our previous analysis of X-linked gene expression in TSCs, where allelic biases determined via RNA-Seq and an alternate method were concordant in 18 of 18 assays (nine genes in two TSC lines) (Calabrese et al. 2012). Considering the data shown in Figure 2 and in our previous work, we conclude that the majority of allelic measurements reported by our RNA-Seq analysis are accurate approximations of steady state gene expression levels present in TSCs.
General characteristics of known and candidate imprinted TSC genes

Many of the 48 PO biased genes displayed variability in their PO expression bias across our panel of TSC lines, including *Sfmbt2*, *Phf17*, *Peg3*, *Usp29*, *Cd81*, and *Peg2* (Table 2). However, this variability was most notable for *H19* and *Igf2*, two neighboring genes whose imprinting status is conserved between human and mouse. *H19* is maternally expressed, and *Igf2* paternally expressed, in both species (Fedoriw et al. 2012a). *H19* was expressed at >95% from the maternal allele in all but one profiled TSC line, BC.3, where its maternal-to-paternal expression ratio was 51-to-49 (Table 2). The variation in PO expression bias of the neighboring paternally biased *Igf2* was even more pronounced than that of *H19*, to the extent that significant PO expression bias was not detected for *Igf2* in our panel of TSCs. Paternal expression of *Igf2* ranged from a low of 24% in BC.3 (meaning it was maternally biased in that TSC line), to a high of 92% in BC.3 (meaning it was maternally biased in that TSC line), to a high of 92% in BC.3 (Table 2).

**Figure 2** Detection of allele-specific expression by pyrosequencing. Individual subplots show maternal expression in six reciprocally derived F1-hybrid TSC lines for eight genes assayed. CB.1, CB.2, and CB.3 were derived from a Cast mother and B6 father, whereas BC.1, BC.2, and BC.3 were derived from a B6 mother and Cast father. Box and whisker plots represent data collected from six technical replicates, except in the H19 subplot, where data derived from four technical replicates. For comparison, numerical values above/below each box and whisker plot show the corresponding maternal expression percentages determined from RNA-Seq in each TSC line. TSC, trophoblast stem cell.

**Figure 3** Genomic environment surrounding Tsci CIG transcripts. (A–F) Name and genomic coordinates for each Tsci transcript relative to UCSC Genome Build mm9. Shown are wiggle density profiles of TSC RNA-Seq data pooled from all six profiled lines, partitioned by matching genomic strand (RNA "+" or "-"), as well as CB.1 DNaseI Hypersensitivity density from (Calabrese et al. 2012). The names of the University of California Santa Cruz Known Genes and Tsci transcripts are indicated above or below their genomic locations. Transcript names in pink and blue signify genes expressed with significant maternal and paternal biases in TSCs, respectively. Arrowheads indicate direction of transcription detected via RNA-Seq. Box-and-wishbone structures indicate splice-forms annotated in the UCSC database or those detected via Cufflinks analysis of pooled TSC RNA-Seq data. RNA-Seq read count density has been log-10 transformed, DNaseI I read count density has not. In (C), the yellow portion of the box-and-wishbone structure of Tsci3 corresponds to the location of the Ensembl noncoding Gene, ENSMUSG00000061469. CIG, candidate imprinted genes. TSC, trophoblast stem cell.
Notably, the TSC line that displayed the near equal maternal-to-paternal expression ratio for H19, TSC line BC3, was the same line that displayed a maternal expression bias for Igf2 (Table 2). Although no other known imprinted gene displayed a variation in PO bias as dramatic as Igf2, a handful of known imprinted genes, such as Osbp5, Impact, Dkc1, and Dkk2, were expressed with a mild PO bias in only five of six profiled TSC lines and as a result were not detected as significantly biased by edgeR (Table S1). Interindividual variation in imprinted gene expression was observed recently in mule and hinny trophoblast cells (Wang et al. 2013b), and its documentation here in mouse TSCs lends support to the idea that such variation may be a conserved feature of genomic imprinting across mammals. Although it remains to be tested in future studies, we speculate that at least some of the variability in imprinted gene expression that we observed in TSCs is due to stochastic variation in levels of DNA methylation or other epigenetic marks at imprinted control regions; this putative variation could have been acquired in cell culture, or may have been naturally present in individual trophoblast cells at the time of TSC derivation. In either case, the observed variation in imprinted gene expression supports the recently proposed notion that genomic imprinting may have evolved as a means to confer robustness to developing embryos during changes in fetal environmental conditions (Radford et al. 2011). Further, the presence of such variation supports our strategy to detect consistently imprinted transcripts by profiling PO biased expression across multiple TSC lines.

We detected significant PO expression biases for 17 genes not previously reported to be subject to genomic imprinting, referred to hereafter as candidate imprinted genes (CIGs; Table 3). In general, total expression levels of the CIGs were lower than those of the known imprinted genes. Average and median expression levels for the CIGs were 7.7 and 0.3 RPKM, respectively, compared with 190 and 13.6 RPKM for known imprinted genes (Tables 2 and Table 3). Like that observed for the known imprinted genes, some CIGs, such as Tsci1 and Tsci6, showed strong PO bias in all six TSC lines profiled, whereas others exhibited more mild levels of PO bias (Table 3). For example, Qk had an average maternal bias in the CB.x lines of 55%, and an average maternal bias in the BC.x lines of 66%, and all of these values were confirmed via pyrosequencing (Figure 2). We speculate that certain genes with mild PO expression biases in TSCs, such as the CIG Qk, may not be imprinted in the canonical sense, but may exhibit imprinted expression due to their proximity to strongly imprinted controlling elements that are able to impose a PO expression bias on nearby susceptible genes.

We detected six CIGs that are not currently annotated as transcribed regions or genes in the mm9 or mm10 UCSC builds of the mouse genome (Kent et al. 2002). We have provisionally named these transcripts TSC Imprinted (i.e., Tsci) 1 through 6 (Table 3 and Figure 3). A minority fraction of one of these transcripts, Tsci3, overlapped with a nonprotein coding gene annotated in the Ensembl genome database, ENSMUSG00000081469 (Cunningham et al. 2014) (yellow box-and-wishbone in Figure 3C). The remaining five Tsci transcripts had no corresponding annotations in UCSC, Ensembl, or GENCODE builds of the mouse transcriptome (Kent et al. 2002; Harrow et al. 2012; Cunningham et al. 2014).

We assessed the coding potential of the six Tsci transcripts using two prediction algorithms, CPAT and CPC, and found that the Tsci3 transcript has potential to encode for a 225 amino acid hypothetical protein that does not appear to be conserved in human (Kong et al. 2007; Wang et al. 2013a). Notably, Tsci3 was transcribed divergently from another hypothetical protein, AK017220, which itself was detected as a CIG in TSCs (Table 3). We suggest the remaining five Tsci transcripts are putative lncRNAs given their length of greater than 200 nucleotides and their lack of coding potential. One of these transcripts, Tsci2, has a transcribed counterpart in its syntenic human region, the GENCODE lncRNA RP4-724E13.2 (Harrow et al. 2012); the other five do not.

Our identification of the Tsci transcripts supports efforts put forth by the ENCODE consortium and others to perform RNA-Seq across large panels of cell and tissues (Yue et al. 2014). lncRNAs are known to be expressed with a tissue specificity equal to that of protein coding genes (Cabili et al. 2011; Derrien et al. 2012), but at present their locations in genomes cannot be predicted, and thus their existence must be determined via empirical measurement. On the basis of our discovery of the Tsci transcripts in this work, it seems likely that continued RNA-Seq profiling in rare or understudied cell populations might also uncover new IncRNAs. By this same logic, similar profiling efforts performed in F1-hybrid backgrounds may uncover additional tissue-specific imprinted transcripts.

Consistent with the tendency of imprinted genes to be localized within clusters in the genome, 13 of 17 CIGs were located near known imprinted genes or other CIGs (Figure 3 and not shown). These include the R74862 IncRNA, located within the cluster of maternally biased genes surrounding the Kcnq1ot1 IncRNA, Gisp1 and Tsci1, which are divergent transcript pairs, Tsci2, located adjacent to Grb10 and Grb10as, AK017220 and Tsci3, also divergent transcript pairs, Tsci4, located between the known imprinted genes Dlk1 and Meg3, and six CIGs surrounding the known imprinted genes Pde10a and Airn (1700010I14Rik, Qk, Pacrg, Park2, Mas1, and Dact2).

We next examined the 24 maternally and paternally biased genes, respectively, for significant enrichment in functional ontologies using

| Table 4 Imprinted genes expressed in TSCs that encode miRNAs |
|-------------------------------------------------------------|
| Host Gene          | miRNAs                  |
|--------------------|-------------------------|
| Sfmbt2             | miR669a-3, miR467d, miR669a-2.5, miR297b, miR669a-2.2, miR669e, miR669a-2.9, miR467a-1.9, miR466, miR466g, miR467a-1.5, miR4660, miR669a-2.8, miR467a-1.10, miR297a-3, miR669a-2.7, miR467a-1.1, miR669h, miR669a-2.1, miR669g, miR669a-1, miR669p-1.1, miR669c, miR467a-1.6, miR669m-1, miR669a-2.4, miR669a-2.10, miR669p-1.2, miR667a-1.2, miR669k, miR467a-1.8, miR669a-2.6, miR467a-3, miR669a-2.3, miR669i, miR467a-1.7, miR669b, miR467b, miR669j, miR467a-1.3, miR467a-1.4 |
| Mest               | miR335                  |
| Usp29              | miR3099                 |
| H19                | miR675                  |
| Meg3               | miR1906-1, miR770       |
| Mirtg              | miR377, miR134, miR496, miR154, miR412, miR485, miR382, miR410, miR668, miR3072, miR453, miR409, miR541, miR369 |

TSCs, trophoblast stem cells, miRNA, microRNA.
including et al. 2013). TSCs also expressed high levels of several imprinted single gene assays (Lewis observed in prior studies examining TSC imprinted expression via silenced by the imprinted lncRNAs molecular etiology of these diseases, and, more broadly, may shed light on imprinted gene expression is established and maintained may there-

Bartolomei 2013). The focused study of the mechanisms by which Barlow, D. P., 2011 Genomic imprinting: a mammalian epigenetic discovery. J. Assist. Reprod. Genet. 26: 477–486. Cabili, M. N., C. Trapnell, L. Goff, M. Koziol, B. Tazon-Vega et al., 2011 Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25: 1915–1927. Calabrese, J. M., W. Sun, L. Song, J. W. Mugford, L. Williams et al., 2012 Site-specific silencing of regulatory elements as a mechanism of X inactivation. Cell 151: 951–963. Clarke, H. J., S. Varnum, V. R. Prideaux, and J. Rossant, 1988 The development potential of parthenogenetically derived cells in chimeric mouse embryos: implications for action of imprinted genes. Development 104: 175–182. Court, F., C. Tayama, V. Romanelli, A. Martin Trujillo, I. Iglesias-Platas et al., 2014 Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of the human imprintome and suggests a germine methylation independent establishment of imprinting. Genome Res. 24: 554–569. Cunningham, F., M. R. Amode, D. Barrett, K. Beal, K. Billis et al., 2015 Ensembl 2015. Nucleic Acids Res. 43: D662–D669. Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali et al., 2012 The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 22: 1775–1789. Fedorow, A., J. Mugford, and T. Magnuson, 2012a Genomic imprinting and epigenetic control of development. Cold Spring Harb. Perspect. Biol. 4: a008136. Fedorow, A. M., J. M. Calabrese, W. Mu, D. Yee, and T. Magnuson, 2012b Differentiation-driven nuclear association of the mouse imprinted Kcnq1 locus, G3 (Bethesda) 2: 1521–1528. Ferguson-Smith, A. C., 2011 Genomic imprinting: the emergence of an epigenetic paradigm. Nat. Rev. Genet. 12: 565–575. Garfield, A. S., M. Cowley, F. M. Smith, K. Moorwood, J. E. Stewart-Cox et al., 2011 Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature 469: 534–538. Harrow, J., A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans et al., 2012 GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22: 1760–1774. Huang da, W., B. T. Sherman, and R. A. Lempicki, 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics re-
sources. Nat. Protoc. 4: 44–57. Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong et al., 2011 Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477: 289–294. Kelsey, G., and R. Feil, 2013 New insights into establishment and maintenance of DNA methylation imprints in mammals. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368: 20110336. Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle et al., 2002 The human genome browser at UCSC. Genome Res. 12: 996–1006. Kong, L., Y. Zhang, Z. Q. Ye, X. Q. Liu, S. Q. Zhao et al., 2007 CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35: W345–W349. Lee, J. T., and M. S. Bartolomei, 2013 X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell 152: 1308–1322. Lewis, A., K. Green, C. Dawson, L. Redrup, K. D. Huyhn et al., 2006 Epigenetic dynamics of the Kcnq1 imprint domain in the early embryo. Development 133: 4203–4210. Miri, K., K. Latham, B. Panning, Z. Zhong, A. Andersen et al., 2013 The imprinted polycomb group gene Smbt2 is required for trophoblast maintenance and placenta development. Development 140: 4840–4849. Okae, H., H. Hiura, Y. Nishida, R. Funayama, S. Tanaka et al., 2012 Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinte expression. Hum. Mol. Genet. 21: 548–558. OMIM 2013 Online Mendelian Inheritance in Man, edited by J. H. U. McKusick-Nathans Institute of Genetic Medicine, Baltimore, MD. Available at: http://omim.org/. Accessed: March 3, 2015. Prickett, A. R., and R. J. Oakley, 2012 A survey of tissue-specific genomic imprinting in mammals. Mol. Genet. Genomics 287: 621–630.
Quinn, J., T. Kunath, and J. Rossant, 2006 Mouse trophoblast stem cells. Methods Mol. Med. 121: 125–148.

Radford, E. J., S. R. Ferron, and A. C. Ferguson-Smith, 2011 Genomic imprinting as an adaptive model of developmental plasticity. FEBS Lett. 585: 2059–2066.

Rebhan, M., V. Chalifa-Caspi, J. Prilusky, and D. Lancet, 1998 GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics 14: 656–664.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth, 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139–140.

Stadtfeld, M., E. Apostolou, F. Ferrari, J. Choi, R. M. Walsh et al., 2012 Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. Nat. Genet. 44: 398–405, S1–2.

Surañi, M. A., and S. C. Barton, 1983 Development of gynogenetic eggs in the mouse: implications for parthenogenetic embryos. Science 222: 1034–1036.

Wang, Q., J. Chow, J. Hong, A. F. Smith, C. Moreno et al., 2011a Recent acquisition of imprinting at the rodent Sfmbt2 locus correlates with insertion of a large block of miRNAs. BMC Genomics 12: 204.

Wang, X., P. D. Soloway, and A. G. Clark, 2011b A survey for novel imprinted genes in the mouse placenta by mRNA-seq. Genetics 189: 109–122.

Wang, X., D. C. Miller, R. Harman, D. F. Antczak, and A. G. Clark, 2013b Paternally expressed genes predominate in the placenta. Proc. Natl. Acad. Sci. USA 110: 10705–10710.

Williamson, C. M., A. Blake, S. Thomas, C. V. Beechey, J. Hancock et al., 2013 Mouse Imprinting Data and References. MRC Harwell, Oxfordshire, UK.

Yue, F., Y. Cheng, A. Breschi, J. Vierstra, W. Wu et al., 2014 A comparative encyclopedia of DNA elements in the mouse genome. Nature 515: 355–364.

Zacharek, S. J., C. M. Fillmore, A. N. Lau, D. W. Gludish, A. Chou et al., 2011 Lung stem cell self-renewal relies on BMI1-dependent control of expression at imprinted loci. Cell Stem Cell 9: 272–281.

Communicating editor: B. J. Andrews