Imprinted genes in a founder population

Parent of origin gene expression in a founder population identifies two new imprinted genes at known imprinted regions.

Sahar V. Mozaffari¹,²*, Michelle M. Stein², Kevin M. Magnaye², Dan L. Nicolae¹,²,³, Carole Ober¹,²

¹Committee on Genetics, Genomics & Systems Biology, University of Chicago, Chicago, Illinois, United States of America
²Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America
³Department of Statistics, University of Chicago, Chicago, Illinois, United States of America

* Corresponding author

E-mail: smozaffari@uchicago.edu (SVM)
Genomic imprinting is the phenomena that leads to silencing of one copy of a gene inherited from a specific parent. Mutations in imprinted regions have been involved in diseases showing parent of origin effects. Identifying genes with evidence of parent of origin expression patterns in family studies allows the detection of more subtle imprinting. Here, we use allele specific expression in lymphoblastoid cell lines from 306 Hutterites related in a single pedigree to provide formal evidence for parent of origin effects. We take advantage of phased genotype data to assign parent of origin to RNA-seq reads in individuals with gene expression data. Our approach identified known imprinted genes, two putative novel imprinted genes, and 14 genes with asymmetrical parent of origin gene expression. We used gene expression in peripheral blood leukocytes (PBL) to validate our findings, and then confirmed imprinting control regions (ICRs) using DNA methylation levels in the PBLs.

Large scale gene expression studies have identified known and novel imprinted genes through allele specific expression without knowing the parental origins of each allele. Here, we take advantage of phased genotype data to assign parent of origin to RNA-seq reads in 306 individuals with gene expression data. We identified known imprinted genes as well as two novel imprinted genes in lymphoblastoid cell line gene expression. We used gene expression in PBLs to validate our findings, and DNA methylation levels in PBLs to confirm previously characterized imprinting control regions that could regulate these imprinted genes.

Imprinted genes have one allele silenced in a parent of origin specific manner. In humans, approximately 105 imprinted loci have been identified, many of which play important roles in...
Dysregulation of imprinted genes or regions can cause diseases that show parent of origin effects, such as Prader-Willi or Angelman syndrome, among others [2]. Imprinted regions have also been associated with complex traits, such as height and age of menarche [4,5], as well as common diseases such as obesity and some cancers [2]. More than 80% of imprinted genes in humans are clustered in genomic regions that contain both maternally and paternally expressed genes, as well as genes that encode non-coding RNAs[2,6]. Parent-specific expression of the genes within a cluster are maintained by complex epigenetic mechanisms at cis-acting imprinting control regions (ICRs) [3], which show parent of origin specific DNA methylation patterns and chromatin modifications[7].

Using RNA-seq and allele specific expression (ASE) we can map genes to parental haplotypes and identify those that are expressed when inherited from only the father or only from the mother, a hallmark feature of imprinted loci. Parent of origin effects and imprinted genes have been most elegantly studied in mice, where two inbred strains are bred reciprocally to identify parent of origin effects on gene expression in progeny that have the same genotypes but different patterns of inheritance [8]. Additionally, uniparental inheritance of imprinted regions in mice were associated with abnormal developmental phenotypes [9] before it was shown that imprinting defects are associated with human disease[10,11]. One approach to identifying imprinted loci in humans has been to test for parent of origin effects on gene expression and phenotypes in pedigrees [4,12]. For example, Garg et al. used gene expression in LCLs from HapMap trios to identify 30 imprinting eQTLs with parent of origin specific effects on expression [13]. A study from the GTEx Consortium used RNA-seq data and allele specific expression to identify allelic imbalance in 45 different tissues. By considering genes with monoallelic expression that was evenly distributed to both the reference and alternate alleles
across individuals as evidence for imprinting, they identified 42 imprinted genes, both known and novel, and used family studies to confirm imprinting of 5 novel imprinted genes [14]. Santoni et al. identified nine novel imprinted genes using single-cell allele-specific gene expression and identifying genes with mono-allelic expression in fibroblasts from 3 unrelated individuals and probands of 2 family trios, and then used the trios to confirm parent of origin of the alleles [15].

Here, we perform a parent of origin ASE study in a large pedigree to characterize parent of origin specific gene expression in the Hutterites, a founder population of European descent, for which we have phased genotype data [16]. We use RNA-seq from lymphoblastoid cell lines (LCLs) to map transcripts to parental haplotypes and identify known and two not previously reported imprinted genes. We validated the two putative imprinted genes by showing the same patterns of parent of origin expression PBLs from different Hutterite individuals, and show DNA methylation signatures of imprinting in the PBLs at these regions.

Results

Mapping transcripts to parental haplotypes

For each of 306 individuals, the total number of transcripts at each gene was assigned as maternally inherited, paternally inherited, or unknown parent of origin. The last group included transcripts without heterozygote SNPs or transcripts with SNPs without parent of origin information. Transcripts were assigned to the parentally inherited categories using SNPs in the reads and matching alleles to either the known maternally or paternally inherited alleles. All the genes analyzed had some transcripts of unknown origin (average 97.8%, range 8.3-100%). For each gene we assigned parental origin to an average of 1.8% of transcripts (range: 0-34.7%), and
for each individual we assigned parental origin to an average of 1.4% of transcripts (range: 0-1.7%). On average, about 40 SNPs per gene were used to assign the transcripts of a gene to parent (range 1-1839 SNPs).

**Table 1. Summary Statistics for Parental Origin of Transcripts.**

|                                | Mean | Standard Deviation | Range       |
|--------------------------------|------|--------------------|-------------|
| Proportion of transcripts from each gene assigned to transcripts of unknown origin | 0.978 | 0.031              | (0.083, 1)  |
| Proportion of transcripts from each gene assigned to parental origin              | 0.018 | 0.019              | (0, 0.347)  |
| Proportion of transcripts for each individual assigned to parental origin         | 0.014 | 0.0015             | (0, 0.017)  |

After quality control (see Methods), transcripts in 15,889 genes were detected as expressed in 306 individuals. Some transcripts for 14,791 of those genes could be assigned to a parent. Of these, 75 genes were only expressed on the paternally-inherited allele in at least one individual and not on the maternally inherited allele in any individuals. Similarly, 64 genes were only expressed on the maternally-inherited allele in at least one individual and not on the paternally inherited allele in any individuals (S1 Table).

**Imprinted Genes in Lymphoblastoid Cell Lines (LCLs)**

Among the 139 genes with only paternally inherited expression or only maternally inherited expression, there are three known imprinted genes (*CDKNIC, NDN, SNRPN*) and one previously predicted to be imprinted (*IFITM1*) [17]. *CDKNIC* showed patterns opposite of what has been reported [18,19], which could be due to the small sample (only three individuals showed expression from one parent) or to the different cell types used here (LCLs) and in previous studies (developing brain and embryonal tumors for *CDKNIC*).
We expect some imprinted genes to have ‘leaky’ expression, such that there is some expression from the parental chromosome that is mostly silenced. To detect these genes, we used a binomial test to find patterns of gene expression asymmetry by parental transcript levels. This analysis identified 28 genes with an FDR <5% (Table 2). The 11 genes that showed the most asymmetry are known imprinted genes: ZDBF2, PEG10, SNHG14, NHP2L1, L3MBTL1, ZNF331, LPAR6, FAM50B, KCNQ1, NAP1L5, and IGF1R. Parent of origin expression for ZDBF2 and KCNQ1 are shown in Fig 1A and 1B, respectively. We identified two additional genes that showed asymmetry in parental expression from mostly one parent (PXDC1, PWAR6), which we consider potentially new imprinted genes. The remaining fourteen genes showed significant patterns of asymmetry but had expression from both maternal and paternal chromosomes. These genes are likely not imprinted but could have asymmetry in expression due to an expression quantitative trait loci (eQTL).

**Table 2. Results for Gene with Parent of Origin Expression Asymmetry.** Genes listed by category of imprinting status: (A) Known Imprinted, (B) Conflicting Evidence for Imprinted Status, (C) New Imprinted Genes, (D) Genes with Asymmetrical Parent of Origin Expression. Genes are ordered by significance within each category.

| Gene   | p-value         | Number of individuals with more maternal expression than paternal expression | Number of individuals with more paternal expression than maternal expression | References                              |
|--------|-----------------|------------------------------------------------------------------------------|-------------------------------------------------------------------------------|-----------------------------------------|
| A. Known Imprinted |
| ZDBF2 | 1.59e-41        | 2                                                                            | 148                                                                           | geneimprint.com, Baran et al.[14], and Babak et al.[8] |
| PEG10 | 5.51e-38        | 2                                                                            | 136                                                                           | geneimprint.com, Baran et al.[14], and Babak et al.[8] |
| Gene       | Expression | Chromosome | Chromosome Location | Reference                           |
|------------|------------|------------|---------------------|-------------------------------------|
| SNHG14     | 1.64e-36   | 2          | 131                 | Baran et al. [14]                   |
| NHP2L1     | 1.24e-33   | 23         | 189                 | Babak et al. [8] and Docherty et al.[20] |
| L3MBTL1    | 6.72e-31   | 2          | 107                 | geneimprint.com and Li et al.[21]   |
| ZNF331     | 4.05e-25   | 36         | 184                 | Daelemans et al. [22] and Baran et al. [14] |
| LPAR6      | 2.65e-23   | 0          | 76                  | Baran et al. [14]                   |
| FAM50B     | 5.29e-23   | 0          | 75                  | geneimprint.com, Baran et al. [14]  |
| KCNQ1      | 1.34e-22   | 79         | 1                   | geneimprint.com, Baran et al. [14]  |
| NAP1L5     | 3.76e09    | 0          | 29                  | geneimprint.com                     |
| IGF1R      | 1.11e-05   | 14         | 49                  | Geneimprint.com, Sun et al. [23,24], Boucher et al. [25], Al Adhami et al. [26] |

B. Conflicting Evidence for Imprinting Status in the literature

| Gene       | Expression | Chromosome | Chromosome Location | Reference                           |
|------------|------------|------------|---------------------|-------------------------------------|
| PRIM2      | 5.53e-05   | 30         | 71                  | geneimprint.com, Santoni et al. [15] |

C. New Imprinted Genes

| Gene       | Expression | Chromosome | Chromosome Location | Reference |
|------------|------------|------------|---------------------|-----------|
| PXDC1      | 9.83e-14   | 12         | 81                  | -         |
| PWAR6      | 2.27e-13   | 0          | 43                  | -         |

D. Genes with Asymmetrical Parent of Origin Expression

| Gene       | Expression | Chromosome | Chromosome Location | Reference |
|------------|------------|------------|---------------------|-----------|
| SNHG17     | 6.2e-08    | 113        | 45                  | -         |
| ZNF813     | 8.7e-07    | 63         | 132                 | -         |
| DAAM1      | 1.78e-05   | 66         | 126                 | -         |
| RP11-379H18.1 | 2.09e-05 | 52         | 106                 | -         |
| HMGN1P38   | 2.43e-05   | 32         | 6                   | -         |
| MTX2       | 3.05e-05   | 0          | 16                  | -         |
| ZNF714     | 4.61e-05   | 35         | 79                  | -         |
| MAF1       | 4.45e-05   | 17         | 51                  | -         |
| IL16       | 5.71e-05   | 61         | 115                 | -         |
| CPNE1      | 5.56e-05   | 111        | 58                  | -         |
Two genes showed gene expression signatures consistent with imprinting but have not previously been recognized as imprinted genes. The first potentially new imprinted gene is *PXDC1*, which is in the same region and next to (<100kb) a known imprinted gene, *FAM50B*. The second potentially novel imprinted gene is *PWAR6*, or Prader Willi Angelman Region RNA6, a gene encoding a regulatory class of RNA. Although this gene is located within the intron of a known imprinted gene, *SNHG14*, this noncoding RNA has not previously been recognized as having parent of origin specific expression (Fig 1C).

The remaining fourteen genes show significant asymmetry using the binomial test but do not have expression from mostly one parental chromosome. One of these genes, *SNHG17*, is a noncoding RNA. Another gene with parent of origin asymmetry, *ZNF813*, is next to a known imprinted gene, *ZNF331*. The remaining genes with asymmetrical parent origin expression have expression from both parental chromosomes, unlike imprinted genes. These genes include *DAAM1*, which is involved in cytoskeleton, specifically filopodia formation [27,28], and has a suggested role for cytoskeleton organization during Mammalian testis morphogenesis and gamete progression [29]; *RP11-379H18.1*, a noncoding RNA gene; *HMGNIP38* [30]; *MTX2*, a nuclear gene that interacts with mitochondrial membrane protein metaxin 1 and is involved in mitochondrial protein import and metabolism of proteins in mice; *MAF1*, a negative regulator of RNA polymerase 2; *ZNF714*, *CPNE1*, *IL16*, *ATP6V0D1*, *FAHD1*, *HSP90AB3P*, and *CNN2* are
the remaining genes that show parent of origin asymmetry but not with a pattern consistent with imprinting (S1 Figure).

Validation of Imprinted Genes in PBLs

Using the same methods described above, we assigned parent of origin to transcripts in PBLs from 99 Hutterite individuals not included in the LCL studies. Maternal and paternal expression in PBLs for all 28 genes identified in LCLs showed similar trends of asymmetry as in LCLs (Fig 2).

Methylation at Imprinting Control Regions.

One of the mechanisms underlying parent of origin effects on expression at imprinted loci is differential methylation at cis-acting imprinting control regions (ICRs). DNA methylation from the Illumina HumanMethylation 450K array was available in PBLs from the same individuals included in the validation study described above. To determine the expected patterns of methylation at known imprinted loci, we first looked at previously characterized methylated regions at known imprinted regions from Court et al. and Joshi et al. [31,32].

The methylation patterns at the two potentially novel imprinted genes identified in this study, PXDC1 and PWAR6, lie in or near known imprinted regions that contain previously characterized ICRs. These previously characterized ICRs show about 50% methylation (beta value of between 0.25 and 0.75) in our DNA methylation data, which likely reflect methylation at only one parental chromosome in all the cells in the sample. Methylation patterns in PBLs at these two ICRs fall within this hemi-methylation range, further suggesting that these two genes are indeed imprinted (Fig 3).
Discussion

Dysregulation of imprinted genes can have a large impact on mammalian development and has been associated with significant diseases in humans. Studies aimed at identifying imprinted genes at genome-wide levels have used allele specific expression and imbalance to infer parent of origin. Here we used a large pedigree with assigned parent of origin alleles to map transcripts to chromosomes with known parent of origin and identify imprinted genes.

Using this approach, we found genes with expression primarily from either the maternal or paternal haplotype. Because gene silencing at imprinted loci may be incomplete, we used a binomial test on parent of origin gene expression and identified 11 known imprinted genes and two potentially novel imprinted genes. Both of these novel genes, PWAR6 and PXDC1, lie in known imprinted regions but have not themselves been characterized as imprinted. The remaining genes that have significant parent of origin asymmetry in gene expression do not show clear imprinting expression patterns. To validate these findings, we mapped gene expression in PBLs from Hutterite individuals not included in the LCL study. The same genes showed similar patterns of asymmetry in these different cell sources (transformed B cells and peripheral blood leukocytes) from different individuals.

We also characterized methylation patterns near genes showing asymmetry. Using results from studies that had previously characterized ICRs in patients with uniparental disomy at many imprinted regions [31,32], we estimated regions for defining hemi-methylation near the genes identified in our study. Using this approach, we were able to provide additional supportive data for the two potentially new imprinted genes to be true imprinted genes regulated by previously characterized ICRs.
Although our study is the largest pedigree-based study to date to search genome-wide for imprinted genes, it has limitations. First, we are able to determine the parent of origin for a many transcripts in the Hutterites but we could not assign every RNA sequencing read to a parent due to lack of heterozygous sites or missing parent of origin information for alleles. Second, we conducted these studies in lymphoblastoid cell lines, and therefore could only study genes imprinted in this cell type and would miss the many imprinted genes that are tissue-specific and/or developmentally regulated[33]. Third, while we can verify previously characterized ICRs, our study is not designed to identify novel ICRs because DNA methylation values from an array cannot be assigned to parental haplotype. Lastly, although we characterized the gene expression and methylation patterns for two potentially novel imprinted genes, replication of these genes in a different population and in different tissues, and functional characterization of these genes are required to confirm their status as imprinted genes. Similarly, some of the other genes with parent of origin asymmetry in the blood cells examined in this study may show more clear-cut evidence for imprinting in other tissues or at specific periods of development.

In summary, we have identified two new imprinted genes using gene expression from a founder population. The genes with asymmetrical parental expression had similar patterns of asymmetry in a different source of blood cells and in different individuals, and we were able to replicate the methylation patterns in known ICRs near the known and novel imprinted genes in this study. Our method and study population allowed us to map reads to parental haplotypes and uncovered \textit{PW\textit{AR6}} and \textit{PXDC1} as new imprinted genes that could potentially impact disease risk and development.
Methods

Genotypes

Hutterite individuals (n=1,653) were genotyped using one of three Affymetrix genotype arrays, as previously described [16], of which 121 underwent whole genome sequencing by Complete Genomics, Inc (CGI) (n=98) or Illumina whole genome sequencing (n=27). A total of 10,235,233 variants present in the sequenced individuals were imputed and phased to the remaining 1,532 genotyped individuals using PRIMAL [16]. Parent of origin was assigned to 89.85% of the alleles with call rate 81.6842% after QC. For this study, we included individuals with genotyped parents in the primary analyses in LCLs. Written consents for these studies were obtained from the adult participants and parents of children under 18; written assents were obtained from all children. This study was approved by the University of Chicago Institutional Review Board.

RNA-seq in Lymphoblastoid Cell Lines (LCLs).

RNA-seq was performed in LCLs as previously described [34]. For this study, sequencing reads were reprocessed as follows. Reads were trimmed for adaptors using Cutadapt (reads less than 5 bp discarded) then remapped to hg19 using STAR indexed with gencode version 19 gene annotations [35,36]. To remove mapping bias, reads were processed and duplicate reads removed using WASP [37]. We used a custom script modified from WASP to separate reads that overlap maternal alleles or paternal alleles. Reads without informative SNPs (homozygous, or no parent of origin information) were categorized as unknown where the unknown, maternal, and paternal make up the total gene expression. Gene counts were quantified using STAR for each category. VerifyBamID was used to identify sample swaps [38]. Genes
mapping to the X and Y chromosome were removed; genes with a CPM log transformed value less than 1 in less than 20 individuals were also removed.

**RNA-seq in Peripheral Blood Leukocytes (PBLs)**

RNA-seq was performed in whole blood as previously described [39]. For this study, sequencing reads were reprocessed as described above for the studies in LCLs. For these analyses, we excluded 32 individuals who were also in the LCL study.

**Identifying Imprinted Genes**

We used a binomial test to detect asymmetry in parent of origin gene expression. Using the paternally and maternally assigned reads, we generated a binomial Z-score for each individual for each gene ($Z_i$) and excluded those where $Z_i=0$. For each gene, the number of subjects with $Z_i>0$ can be modeled by a Binomial distribution with probability $\frac{1}{2}$, under the null hypotheses of symmetric expression. For imprinted genes that show patterns of asymmetry, we expect a distribution of Z-scores that are skewed to one direction corresponding to asymmetric expression. Because we are only asking whether there are more individuals with more maternal expression or more paternal expression and not gene expression measures there is no need to model over-dispersion.

**DNA methylation profiling and processing in PBLs**

One milliliter of whole blood from 145 Hutterites was drawn into TruCulture (Myriad RBM; Austin, Texas) tubes containing proprietary TruCulture media. DNA was extracted using AllPrep DNA/RNA Mini Kits (Qiagen). DNA samples were bisulfite converted and hybridized to the Illumina HumanMethylation 450K array at the University of Chicago Functional Genomics Center. Samples were processed using default parameters using the R package minfi [40], normalized using SWAN (subset within-array normalization [41]) and quantile normalized...
similar to previous methylation studies [42]. Probes were removed if: (1) mapped non-uniquely to a bisulfite-converted genome; (2) mapped to sex chromosomes; (3) had a probe detection p-value $>0.01$ in at least 25% of samples; and (4) contained common SNPs within the probe sequence, as previously described [43]. Principal components analysis (PCA) was used to identify significant technical covariates, and the ComBat function [44] within the R package sva [45] was used to correct for chip effect. Analyses of DNA methylation levels were conducted using beta values, which were converted from M-values using the lumi R package [46].

**Acknowledgements**

We thank members of the Ober lab for useful discussions, Joe Urbanski and Lorenzo Pesce for assistance using Beagle, the many members of our field trip teams, and the Hutterites for their continued support of our studies.

**Funding Disclosure on Submission System**

This work was supported by NIH grants HL085197 and HD21244; and in part by NIH through resources provided by the Computation Institute and the Biological Sciences Division of the University of Chicago and Argonne National Laboratory, under grant 1S10OD018495-01. S.V.M has been supported by NIH Grant T32 GM007197 and the Ruth L. Kirschstein NRSA Award F31HL134315.
References

1. Falls JG, Pulford DJ, Wylie AA, Jirtle RL. Genomic Imprinting: Implications for Human Disease. The American Journal of Pathology. 1999;154: 635–647. doi:10.1016/S0002-9440(10)65309-6

2. Peters J. The role of genomic imprinting in biology and disease: an expanding view. Nat Rev Genet. 2014;15: 517–530. doi:10.1038/nrg3766

3. Kalish JM, Jiang C, Bartolomei MS. Epigenetics and imprinting in human disease. Int J Dev Biol. 2014;58: 291–298. doi:10.1387/ijdb.140077mb

4. Benonisdottir S, Oddsson A, Helgason A, Kristjansson RP, Sveinbjornsson G, Oskarsdottir A, et al. Epigenetic and genetic components of height regulation. Nat Comms. 2016;7: 13490. doi:10.1038/ncomms13490

5. Zoledziewska M, Sidore C, Chiang CWK, Sanna S, Mulas A, Steri M, et al. Height-reducing variants and selection for short stature in Sardinia. Nat Genet. 2015;47: 1352–1356. doi:10.1038/ng.3403

6. Barlow DP, Bartolomei MS. Genomic Imprinting in Mammals. Cold Spring Harbor Perspectives in Biology. 2014;6: a018382–a018382. doi:10.1101/cshperspect.a018382

7. Abramowitz LK, Bartolomei MS. Genomic imprinting: recognition and marking of imprinted loci. Curr Opin Genet Dev. 2012;22: 72–78. doi:10.1016/j.gde.2011.12.001

8. Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, et al. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat Genet. 2015;47: 544–549. doi:10.1038/ng.3274

9. Cattanach BM, Kirk M. Differential activity of maternally and paternally derived chromosome regions in mice. Nature. Nature Publishing Group; 1985;315: 496–498. doi:10.1038/315496a0

10. Nicholls RD, Knoll J, Butler MG, Nature SK. Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. naturecom. 1989.

11. Reik W. Genomic imprinting and genetic disorders in man. Trends in Genetics. Elsevier Current Trends; 1989;5: 332–336. doi:10.1016/0168-9525(89)90138-8

12. Kong A, Steinhorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009;462: 868–874. doi:10.1038/nature08625

13. Garg P, Borel C, Sharp AJ. Detection of Parent-of-Origin Specific Expression Quantitative Trait Loci by Cis-Association Analysis of Gene Expression in Trios. Chadwick BP, editor. PLoS ONE. 2012;7: e41695. doi:10.1371/journal.pone.0041695
14. Baran Y, Subramaniam M, Biton A, Tukiainen T, Tsang EK, Rivas MA, et al. The landscape of genomic imprinting across diverse adult human tissues. Genome Research. 2015;25: 927–936. doi:10.1101/gr.192278.115

15. Santoni FA, Stamoulis G, Garieri M, Falconnet E, Ribaux P, Borel C, et al. Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression. Am J Hum Genet. 2017;100: 444–453. doi:10.1016/j.ajhg.2017.01.028

16. Livne OE, Han L, Alkorta-Aranburu G, Wentworth-Sheilds W, Abney M, Ober C, et al. PRIMAL: Fast and Accurate Pedigree-based Imputation from Sequence Data in a Founder Population. McHardy AC, editor. PLoS Comput Biol. 2015;11: e1004139.

doi:10.1371/journal.pcbi.1004139

17. Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. Genome Research. 2007;17: 1723–1730. doi:10.1101/gr.6584707

18. Hatada I, Mukai T. Genomic imprinting of p57KIP2, a cyclin-dependent kinase inhibitor, in mouse. Nat Genet. Nature Publishing Group; 1995;11: 204–206. doi:10.1038/ng1095-204

19. Matsuoka S, Thompson JS, Edwards MC, Bartletta JM, Grundy P, Kalikin LM, et al. Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor, p57KIP2, on chromosome 11p15. Proc Natl Acad Sci USA. National Academy of Sciences; 1996;93: 3026–3030.

20. Docherty LE, Rezwan FI, Poole RL, Jagoe H, Lake H, Lockett GA, et al. Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes. J Med Genet. 2014;51: 229–238. doi:10.1136/jmedgenet-2013-102116

21. Li J, Bench AJ, Vassiliou GS, Fourouclas N, Ferguson-Smith AC, Green AR. Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies. Proc Natl Acad Sci USA. National Acad Sciences; 2004;101: 7341–7346. doi:10.1073/pnas.0308195101

22. Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, Forrest MS, et al. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. BMC Genet. 2010;11: 25. doi:10.1186/1471-2156-11-25

23. Sun J, Li W, Sun Y, Yu D, Wen X, Wang H, et al. A novel antisense long noncoding RNA within the IGF1R gene locus is imprinted in hematopoietic malignancies. Nucleic Acids Res. 2014;42: 9588–9601. doi:10.1093/nar/gku549

24. Kang L, Sun J, Wen X, Cui J, Wang G, Hoffman AR, et al. Aberrant allele-switch imprinting of a novel IGF1R intragenic antisense non-coding RNA in breast cancers. Eur J Cancer. 2015;51: 260–270. doi:10.1016/j.ejca.2014.10.031
25. Boucher J, Charalambous M, Zarse K, Mori MA, Kleinridders A, Ristow M, et al. Insulin and insulin-like growth factor 1 receptors are required for normal expression of imprinted genes. Proc Natl Acad Sci USA. National Acad Sciences; 2014;111:14512–14517. doi:10.1073/pnas.1415475111

26. Adhami Al H, Evano B, Le Digarcher A, Gueydan C, Dubois E, Parrinello H, et al. A systems-level approach to parental genomic imprinting: the imprinted gene network includes extracellular matrix genes and regulates cell cycle exit and differentiation. Genome Research. Cold Spring Harbor Lab; 2015;25:353–367. doi:10.1101/gr.175919.114

27. Hoffmann A-K, Naj X, Linder S. Daam1 is a regulator of filopodia formation and phagocytic uptake of Borrelia burgdorferi by primary human macrophages. FASEB J. 2014;28:3075–3089. doi:10.1096/fj.13-247049

28. Luo W, Lieu ZZ, Manser E, Bershadsky AD, Sheetz MP. Formin DAAM1 Organizes Actin Filaments in the Cytoplasmic Nodal Actin Network. Aspenstrom P, editor. PLoS ONE. Public Library of Science; 2016;11:e0163915–22. doi:10.1371/journal.pone.0163915

29. Pariante P, Dotolo R, Venditti M, Ferrara D, Donizetti A, Aniello F, et al. First Evidence of DAAM1 Localization During the Post-Natal Development of Rat Testis and in Mammalian Sperm. J Cell Physiol. 2016;231:2172–2184. doi:10.1002/jcp.25330

30. Strichman-Almashanu LZ, Bustin M, Landsman D. Retroposed copies of the HMG genes: a window to genome dynamics. Genome Research. Cold Spring Harbor Lab; 2003;13:800–812. doi:10.1101/gr.893803

31. Court F, Tayama C, Romanelli V, Martin-Trujillo A, Iglesias-Platas I, Okamura K, et al. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Research. Cold Spring Harbor Lab; 2014;24:554–569. doi:10.1101/gr.164913.113

32. Joshi RS, Garg P, Zaitlen N, Lappalainen T, Watson CT, Azam N, et al. DNA Methylation Profiling of Uniparental Disomy Subjects Provides a Map of Parental Epigenetic Bias in the Human Genome. Am J Hum Genet. 2016;99:555–566. doi:10.1016/j.ajhg.2016.06.032

33. Plasschaert RN, Bartolomei MS. Genomic imprinting in development, growth, behavior and stem cells. Development. 2014;141:1805–1813. doi:10.1242/dev.101428

34. Cusanovich DA, Caliskan M, Billstrand C, Michelini K, Chavarria C, De Leon S, et al. Integrated analyses of gene expression and genetic association studies in a founder population. Human Molecular Genetics. 2016;25:2104–2112. doi:10.1093/hmg/ddw061

35. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2002. pp. 11.14.1–11.14.19. doi:10.1002/0471250953.bi1114s51
36. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 2011;17: 10. doi:10.14806/ej.17.1.200

37. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Meth. NIH Public Access; 2015;12: 1061–1063. doi:10.1038/nmeth.3582

38. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and Estimating Contamination of Human DNA Samples in Sequencing and Array-Based Genotype Data. The American Journal of Human Genetics. 2012;91: 839–848. doi:10.1016/j.ajhg.2012.09.004

39. Stein MM, Hrusch CL, Gozdz J, Igartua C, Pivniouk V, Murray SE, et al. Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. N Engl J Med. 2016;375: 411–421. doi:10.1056/NEJMoa1508749

40. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30: 1363–1369. doi:10.1093/bioinformatics/btu049

41. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. Genome Biol. 2012;13: R44. doi:10.1186/gb-2012-13-6-r44

42. Nicodemus-Johnson J, Myers RA, Sakabe NJ, Sobreira DR, Hogarth DK, Naureckas ET, et al. DNA methylation in lung cells is associated with asthma endotypes and genetic risk. JCI Insight. 2016;1: e90151. doi:10.1172/jci.insight.90151

43. Banovich NE, Lan X, McVicker G, van de Geijn B, Degner JF, Blischak JD, et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. PLoS Genet. 2014;10: e1004663. doi:10.1371/journal.pgen.1004663

44. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8: 118–127. doi:10.1093/biostatistics/kxj037

45. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28: 882–883. doi:10.1093/bioinformatics/bts034

46. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24: 1547–1548. doi:10.1093/bioinformatics/btn224
**Figure 1.** Plot of maternal (x-axis) and paternal (y-axis) gene expression for four genes. (A) maternally imprinted gene \textit{ZDBF2} (paternally expressed), (B) paternally imprinted gene \textit{KCNQ1} (maternally expressed), (C) novel maternally imprinted gene \textit{PWAR6} (paternally expressed), (D) gene with asymmetry in parental expression \textit{MAF1}. Each point represents one individual. Numbers in the legend represent the number of individuals with equal maternal and paternal expression, more maternal expression, or more paternal expression.
Figure 2. Histogram showing the number of individuals with more maternal expression (M>P) or more paternal expression (P>M) for the 28 genes showing parent of origin asymmetry in (A) LCLs and (B) PBLs. Genes are ordered by the magnitude of the difference in the number of individuals with more maternal expression than paternal expression in LCLs.
Figure 3. DNA methylation levels near known and novel imprinted genes previously defined by Joshi et al. and Court et al. (A) *PEG10*, (B) *PXDC1* and *FAM50B*, (C) *PWAR6*.
Supporting Information

S1 Figure. Plots of maternal and paternal expression for remaining genes with parent of origin asymmetry.

S2 Figure. Density plot for Differentially Methylated Regions (DMRs) for imprinted genes from Joshi et al and Court et al.

S1 Table. Genes expressed only on maternal or only on paternal haplotypes in LCLs.

...
Mean methylation at CpGs

- **ZDBF2** Court et al. 8 CpGs
- **ZDBF2** Joshi et al. 3 CpGs
- **PEG10** Court et al. 54 CpGs
- **PEG10** Joshi et al. 54 CpGs
- **L3MBTL1** Court et al. 26 CpGs
- **L3MBTL1** Joshi et al. 28 CpGs
- **L3MBTL1** Docherty et al. 24 CpGs
- **ZNF331** Docherty et al. 16 CpGs
- **ZNF331** Court et al. 12 CpGs
- **ZNF331** Court et al. 4 CpGs
- **FAM50B** Court et al. 25 CpGs
- **FAM50B** Joshi et al. 31 CpGs
- **FAM50B** Joshi et al. 20 CpGs
- **NAP1L5** Court et al. 15 CpGs
- **NAP1L5** Joshi et al. 16 CpGs
- **NHP2L1** Court et al. 8 CpGs
- **IGF1R** Joshi et al. 10 CpGs
- **IGF1R** Docherty et al. 5 CpGs
- **KCNQ1** Docherty et al. – BWS 33 CpGs
- **KCNQ1** Docherty et al. – TND 28 CpGs
**Supplementary Table 1. Genes with only Maternal/only Paternal gene expression**

| Gene       | Number of Individuals | Gene Name          | Imprinted Status | Observed Expression Pattern in LCLs | Pattern consistent with known expression? |
|------------|-----------------------|--------------------|------------------|-------------------------------------|-------------------------------------------|
| ENSG00000182636 | 22                    | NDN                | Known            | Paternal                            | Consistent                                |
| ENSG000000259261 | 11                    | IGKV4OR15-8       | Maternal         |                                     |                                           |
| ENSG00000128739 | 9                     | SNRPN              | Known            | Paternal                            | Consistent                                |
| ENSG00000185044 | 9                     | RP11-435B5.4      | Maternal         |                                     |                                           |
| ENSG00000089876 | 6                     | DHX32              | Paternal         |                                     |                                           |
| ENSG00000262333 | 5                     | HNRNPA1P16        | Paternal         |                                     |                                           |
| ENSG00000196378 | 5                     | ZNF34              | Maternal         |                                     |                                           |
| ENSG00000198744 | 5                     | RP5-857K21.11     | Maternal         |                                     |                                           |
| ENSG00000233757 | 5                     | AC092835.2        | Maternal         |                                     |                                           |
| ENSG00000272933 | 5                     | RP11-47A8.5       | Maternal         |                                     |                                           |
| ENSG00000154611 | 4                     | PSMA8              | Paternal         |                                     |                                           |
| ENSG00000228109 | 4                     | MFI2-AS1          | Paternal         |                                     |                                           |
| ENSG00000103269 | 4                     | RHBDL1             | Maternal         |                                     |                                           |
| ENSG00000129654 | 4                     | FOXJ1              | Paternal         |                                     |                                           |
| ENSG00000129757 | 3                     | CDKN1C             | Known            | Paternal                            | Inconsistent                              |
| ENSG00000137936 | 3                     | BCAR3              | Paternal         |                                     |                                           |
| ENSG00000148926 | 3                     | ADM                | Paternal         |                                     |                                           |
| ENSG00000204482 | 3                     | LST1               | Paternal         |                                     |                                           |
| ENSG00000211669 | 3                     | IGLV3-10          | Paternal         |                                     |                                           |
| ENSG00000223509 | 3                     | RP11-632K20.7     | Paternal         |                                     |                                           |
| ENSG00000260442 | 3                     | RP11-22P6.3       | Paternal         |                                     |                                           |
| ENSG00000270441 | 3                     | RP11-694I15.7     | Paternal         |                                     |                                           |
| ENSG00000160828 | 3                     | STAG3L2           | Maternal         |                                     |                                           |
| ENSG00000165886 | 3                     | UBTD1              | Maternal         |                                     |                                           |
| ENSG00000181284 | 3                     | TMEM102            | Paternal         |                                     |                                           |
| ENSG00000214269 | 3                     | LGMNP1             | Maternal         |                                     |                                           |
| ENSG00000258561 | 3                     | RP11-72M17.1      | Maternal         |                                     |                                           |
| ENSG00000075089 | 2                     | ACTR6              | Paternal         |                                     |                                           |
| ENSG00000105518 | 2                     | TMEM205            | Paternal         |                                     |                                           |
| ENSG00000115457 | 2                     | IGFBP2             | Paternal         |                                     |                                           |
| ENSG00000137821 | 2                     | LRRC49             | Paternal         |                                     |                                           |
| ENSG00000158517 | 2                     | NCF1               | Paternal         |                                     |                                           |
| ENSG00000183604 | 2                     | RP11-347C12.2     | Paternal         |                                     |                                           |
| ENSG00000211637 | 2                     | IGLV4-69          | Paternal         |                                     |                                           |
| ENSG000002111940 | 2                    | IGHV3-9           | Paternal         |                                     |                                           |
| ENSG00000233426 | 2                     | EIF3F3            | Paternal         |                                     |                                           |
| ENSG00000240041 | 2                     | IGHJ4              | Paternal         |                                     |                                           |
| ENSG00000240731 | 2                     | RP5-890O3.9       | Paternal         |                                     |                                           |
| Gene ID            | Number | Gene Symbol | Paternality |
|-------------------|--------|-------------|-------------|
| ENSG00000272145   | 2      | NFYC-AS1    | Paternal    |
| ENSG00000025156   | 2      | HSF2        | Maternal    |
| ENSG000000108298  | 2      | RPL19       | Maternal    |
| ENSG00000133216   | 2      | EPHB2       | Maternal    |
| ENSG00000133328   | 2      | HRASLS2     | Maternal    |
| ENSG00000134864   | 2      | GGACT       | Maternal    |
| ENSG00000158481   | 2      | CD1C        | Maternal    |
| ENSG00000169019   | 2      | COMMD8      | Maternal    |
| ENSG00000175701   | 2      | LINC00116   | Maternal    |
| ENSG00000196465   | 2      | MYL6B       | Maternal    |
| ENSG00000198155   | 2      | ZNF876P     | Maternal    |
| ENSG00000215030   | 2      | RPL13P12    | Maternal    |
| ENSG00000232640   | 2      | RP1-266L20.2| Maternal    |
| ENSG00000233493   | 2      | TMEM238     | Maternal    |
| ENSG00000235400   | 2      | RP4-641G12.4| Maternal    |
| ENSG00000240652   | 2      | RP11-832N8.1| Maternal    |
| ENSG00000243364   | 2      | EFNA4       | Maternal    |
| ENSG00000255135   | 2      | RP11-111M22.3| Maternal   |
| ENSG00000267152   | 2      | CTD-2528L19.6| Maternal  |
| ENSG00000033122   | 1      | LRRC7       | Paternal    |
| ENSG00000096080   | 1      | MRPS18A     | Paternal    |
| ENSG00000100442   | 1      | FKBP3       | Paternal    |
| ENSG00000100632   | 1      | ERH         | Paternal    |
| ENSG00000109083   | 1      | IFT20       | Paternal    |
| ENSG00000111875   | 1      | ASF1A       | Paternal    |
| ENSG00000116819   | 1      | TFAP2E      | Paternal    |
| ENSG00000121089   | 1      | NACA3P      | Paternal    |
| ENSG00000122218   | 1      | COPA        | Paternal    |
| ENSG00000128011   | 1      | LRFN1       | Paternal    |
| ENSG00000129673   | 1      | AANAT       | Paternal    |
| ENSG00000140459   | 1      | CYP11A1     | Paternal    |
| ENSG00000148187   | 1      | MRRF        | Paternal    |
| ENSG00000150456   | 1      | N6AMT2      | Paternal    |
| ENSG00000151366   | 1      | NDUFC2      | Paternal    |
| ENSG00000154640   | 1      | BTG3        | Paternal    |
| ENSG00000158716   | 1      | DUSP23      | Paternal    |
| ENSG00000158806   | 1      | NPM2        | Paternal    |
| ENSG00000163634   | 1      | THOC7       | Paternal    |
| ENSG00000165121   | 1      | RP11-213G2.3| Paternal    |
| ENSG00000167286   | 1      | CD3D        | Paternal    |
| ENSG00000173715   | 1      | C11orf80    | Paternal    |
| ENSG00000173762   | 1      | CD7         | Paternal    |
| ENSG00000175550   | 1      | DRAP1       | Paternal    |
| Gene ID | Name | Paternality |
|---------|------|-------------|
| ENSG00000179603 | GRM8 | Paternal |
| ENSG00000181038 | METTL23 | Paternal |
| ENSG00000181852 | RNF41 | Paternal |
| ENSG00000183506 | PI4KAP2 | Paternal |
| ENSG00000197568 | HHHLA3 | Paternal |
| ENSG00000198356 | ASNA1 | Paternal |
| ENSG00000204472 | AIF1 | Paternal |
| ENSG00000211594 | IGKJ4 | Paternal |
| ENSG00000211595 | IGKJ3 | Paternal |
| ENSG00000211965 | IGHV3-49 | Paternal |
| ENSG00000215548 | RP11-764K9.4 | Paternal |
| ENSG00000225329 | RP11-325F22.5 | Paternal |
| ENSG00000226121 | AHCTF1P1 | Paternal |
| ENSG00000233912 | AC026202.3 | Paternal |
| ENSG00000239819 | IGKV1D-8 | Paternal |
| ENSG00000239830 | RPS4XP22 | Paternal |
| ENSG00000243312 | RP11-397E7.1 | Paternal |
| ENSG00000244055 | AC007566.10 | Paternal |
| ENSG00000253998 | IGKV2-29 | Paternal |
| ENSG00000257261 | RP11-96H19.1 | Paternal |
| ENSG00000259699 | HMGB1P8 | Paternal |
| ENSG00000260219 | RP11-347C12.10 | Paternal |
| ENSG00000260655 | CTA-250D10.23 | Paternal |
| ENSG00000264473 | hsa-mir-4538 | Paternal |
| ENSG00000268568 | AC007228.9 | Paternal |
| ENSG00000106211 | HSPB1 | Maternal |
| ENSG00000118514 | ALDH8A1 | Maternal |
| ENSG00000126709 | IFI6 | Maternal |
| ENSG00000131773 | KHDRBS3 | Maternal |
| ENSG00000135914 | HTR2B | Maternal |
| ENSG00000136104 | RNASEH2B | Maternal |
| ENSG00000136463 | TACO1 | Maternal |
| ENSG00000148444 | COMMD3 | Maternal |
| ENSG00000156873 | PHKG2 | Maternal |
| ENSG00000163249 | CCNYL1 | Maternal |
| ENSG00000164794 | KCNV1 | Maternal |
| ENSG00000172586 | CHCHD1 | Maternal |
| ENSG00000174871 | CNINH2 | Maternal |
| ENSG00000178922 | HYI | Maternal |
| ENSG00000183426 | NPIPA1 | Maternal |
| ENSG00000185885 | IFITM1 | Predicted |
| ENSG00000197279 | ZNF165 | Maternal |
| ENSG00000199753 | SNORD104 | Maternal |
| Gene ID          | Chromosome | Maternal | Gene Name          | Description  |
|-----------------|------------|----------|--------------------|--------------|
| ENSG00000215302 | 1          | Maternal | CTD-3092A11.1      |              |
| ENSG00000226085 | 1          | Maternal | UQCRFS1P1          |              |
| ENSG00000227053 | 1          | Maternal | RP11-395B7.4       |              |
| ENSG00000232573 | 1          | Maternal | RPL3P4             |              |
| ENSG00000237973 | 1          | Maternal | hsa-mir-6723       |              |
| ENSG00000240356 | 1          | Maternal | RPL23AP7           |              |
| ENSG00000240449 | 1          | Maternal | RP4-584D14.5       |              |
| ENSG00000253485 | 1          | Maternal | PCDHGA5            |              |
| ENSG00000254681 | 1          | Maternal | PKD1P5             |              |
| ENSG00000254887 | 1          | Maternal | CTC-378H22.1       |              |
| ENSG00000261504 | 1          | Maternal | RP11-317P15.4      |              |
| ENSG00000262691 | 1          | Maternal | CTC-277H1.7        |              |
| ENSG00000266208 | 1          | Maternal | CTD-2267D19.3      |              |
| ENSG00000268030 | 1          | Maternal | AC005253.2         |              |
| ENSG00000272468 | 1          | Maternal | RP1-86C11.7        |              |