

RESEARCH ARTICLE

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

Sahar V. Mozaffari^{1,2*}, Michelle M. Stein², Kevin M. Magnaye², Dan L. Nicolae^{1,2,3}, Carole Ober^{1,2}

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

* smozaffari@uchicago.edu



OPEN ACCESS

Citation: Mozaffari SV, Stein MM, Magnaye KM, Nicolae DL, Ober C (2018) Parent of origin gene expression in a founder population identifies two new candidate imprinted genes at known imprinted regions. *PLoS ONE* 13(9): e0203906. <https://doi.org/10.1371/journal.pone.0203906>

Editor: Keith William Brown, University of Bristol, UNITED KINGDOM

Received: July 4, 2018

Accepted: August 29, 2018

Published: September 11, 2018

Copyright: © 2018 Mozaffari et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The accession number for the Hutterite data reported in this paper is dbGAP:phs000185.

Funding: This work was supported by National Institutes of Health (NIH) (<http://www.nih.gov>) grants HL085197 (C.O.) and HD21244 (C.O.); 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 1S100D018495-01. S.V.M. has been supported by NIH Grant T32

Abstract

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, *PXDC1* and *PWAR6*, 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.

Introduction

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 development and growth [1–3]. 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].

GM007197 and the Ruth L. Kirschstein NRSA Award F31HL134315. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

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 (Table 1). 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

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)

<https://doi.org/10.1371/journal.pone.0203906.t001>

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). See [S2 Table](#) for all SNPs used to assign parental origin.

Transcripts in 15,889 genes were detected as expressed in 306 individuals (see [Methods](#)). 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](#)). See [S2 Table](#) for all SNPs used to assign parental origin for genes in [S1 Table](#) and [Table 2](#).

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 (*CDKN1C*, *NDN*, *SNRPN*) paternally expressed and one previously predicted to be imprinted (*IFITM1*) maternally expressed [17].

We expect some imprinted genes to have 'leaky' expression, such that there is some expression from the parental chromosome that is normally silenced. Slight bias in allelic expression that results in leaky imprinting could be due to detection of a biallelically expressed isoform or to a minority of cells that are biallelically expressed. 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 genes with prior evidence of imprinted expression: *ZDBF2*, *PEG10*, *SNHG14*, *NHP2L1*, *L3MBTL1*, *ZNF331*, *LPAR6*, *FAM50B*, *KCNQ1*, *NAP1L5*, and *IGF1R*. Parent of origin expression for *ZDBF2* is shown in [Fig 1A](#). We identified two additional genes that showed asymmetry, with expression from mostly one parent (*PXDC1*, *PWAR6*), which we consider potentially new candidate 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).

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. It is possible that *PWAR6* is a processed exon of the *UBE3A-ATS/SNHG14* non-coding transcript. The plots of parental expression of *PWAR6* and *PXDC1* are shown in [Fig 1B and 1C](#). The few individuals that have more expression from the parental allele that should be silenced ($n = 2$ in *ZDBF2*, [Fig 1A](#)) could be due to errors in genotyping, errors in the assignment of parentally inherited alleles, or from errors in RNA-seq.

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 *DAAMI*, which is involved in cytoskeleton, specifically filopodia formation [25,26], and has a suggested role for cytoskeleton organization during Mammalian testis

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) Candidate Imprinted Genes, (D) Genes with Asymmetrical Parent of Origin Expression but Not Likely Imprinted. 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. Genes with prior evidence of imprinted expression				
<i>ZDBF2</i>	1.59e-41	2	148	geneimprint.com, Baran et al.[14], and Babak et al.[8]
<i>PEG10</i>	5.51e-38	2	136	geneimprint.com, Baran et al.[14], and Babak et al.[8]
<i>SNHG14</i>	1.64e-36	2	131	Baran et al. [14]
<i>NHP2L1</i>	1.24e-33	23	189	Babak et al. [8] and Docherty et al.[18]
<i>L3MBTL1</i>	6.72e-31	2	107	geneimprint.com and Li et al.[19]
<i>ZNF331</i>	4.05e-25	36	184	Daelemans et al. [20]and Baran et al. [14]
<i>LPAR6</i>	2.65e-23	0	76	Baran et al. [14]
<i>FAM50B</i>	5.29e-23	0	75	geneimprint.com, Baran et al. [14]
<i>KCNQ1</i>	1.34e-22	79	1	geneimprint.com, Baran et al. [14]
<i>NAPIL5</i>	3.76e09	0	29	geneimprint.com
<i>IGF1R</i>	1.11e-05	14	49	Geneimprint.com, Sun et al. [21,22], Boucher et al. [23], Al Adhami et al. [24]
B. Conflicting Evidence for Imprinting Status in the literature				
<i>PRIM2</i>	5.53e-05	30	71	geneimprint.com, Santoni et al. [15]
C. New Candidate Imprinted Genes				
<i>PXDC1</i>	9.83e-14	12	81	-
<i>PWAR6</i>	2.27e-13	0	43	-
D. Genes with Asymmetrical Parent of Origin Expression but Not Likely Imprinted				
<i>SNHG17</i>	6.2e-08	113	45	-
<i>ZNF813</i>	8.7e-07	63	132	-
<i>DAAM1</i>	1.78e-05	66	126	-
<i>RP11-379H18.1</i>	2.09e-05	52	106	-
<i>HMGNI38</i>	2.43e-05	32	6	-
<i>MTX2</i>	3.05e-05	0	16	-
<i>ZNF714</i>	4.61e-05	35	79	-
<i>MAF1</i>	4.45e-05	17	51	-
<i>IL16</i>	5.71e-05	61	115	-
<i>CPNE1</i>	5.56e-05	111	58	-
<i>ATP6V0D1</i>	7.03e-05	32	7	-
<i>FAHD1</i>	9.34e-05	68	29	-
<i>CNN2</i>	1.18e-04	127	72	-
<i>HSP90AB3P</i>	1.16e-04	7	31	-

<https://doi.org/10.1371/journal.pone.0203906.t002>

morphogenesis and gamete progression [27]; *RP11-379H18.1*, a noncoding RNA gene; *HMGNI38* [28]; *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 Fig).

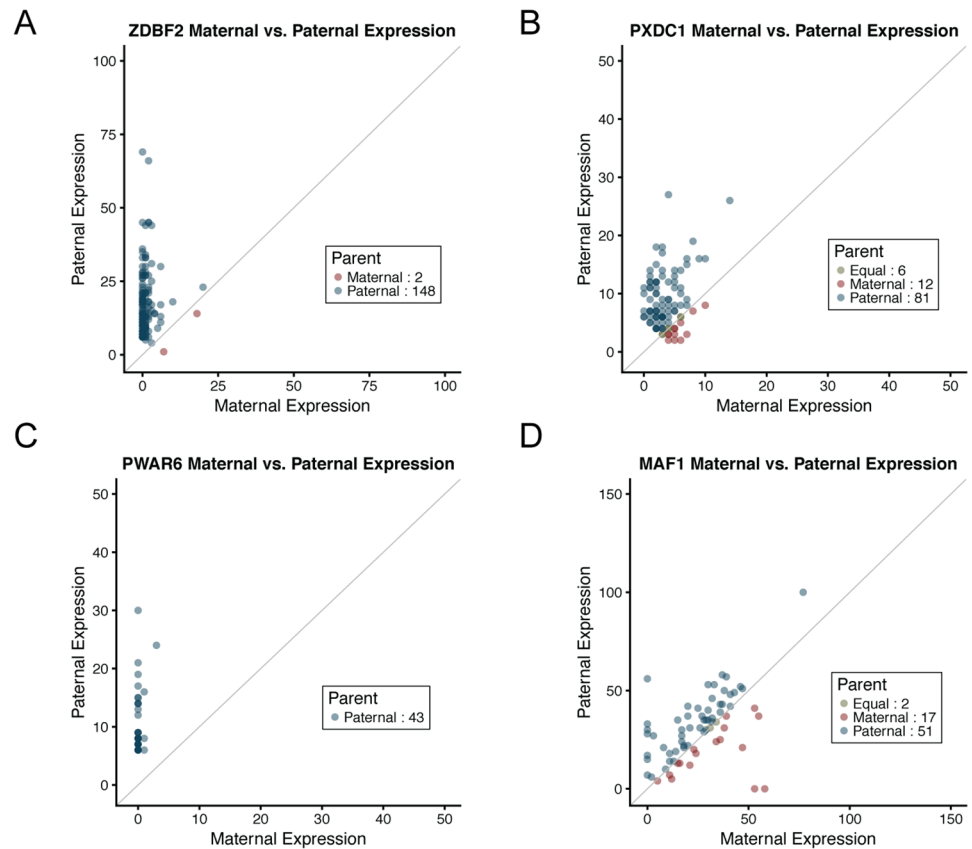


Fig 1. Plots of maternal (x-axis) and paternal (y-axis) gene expression counts for four genes that were significant in the asymmetry test, reflecting more expression from one parental haplotype than the other. Each point represents an individual. The units of x- and y-axis are gene count measures. (A) Maternally imprinted gene *ZDBF2* (paternally expressed). Most individuals have expression solely from the paternal haplotype (shown in blue); some individuals have low expression from the maternal haplotype. Two individuals have more maternal expression than paternal expression (shown in red). (B) Maternally imprinted candidate gene *PXDC1* (paternally expressed). Most individuals ($n = 81$) have more paternal expression than maternal expression; a few individuals have expression solely from the paternal haplotype (maternal expression is 0) and a few ($n = 12$) have more maternal than paternal expression. (C) Maternally imprinted candidate gene *PWAR6* (paternally expressed). All 43 individuals have more paternal expression; four also have very low amounts of maternal expression. (D) Asymmetric parental expression *MAFI*. This gene is likely not imprinted. The numbers of individuals with equal maternal and paternal expression, more maternal expression, or more paternal expression are shown in the legend.

<https://doi.org/10.1371/journal.pone.0203906.g001>

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

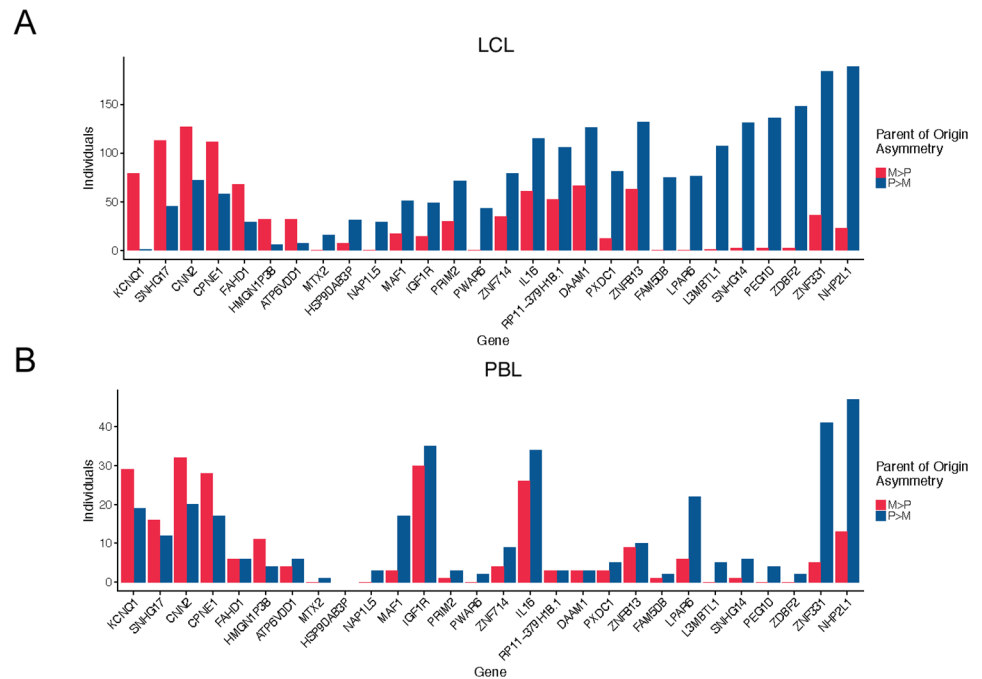


Fig 2. Histogram showing the number of individuals with more maternal expression (M>P) or more paternal expression (P>M) for the 28 genes showing parent of origin asymmetry in (A) LCLs and (B) the same 20 genes with expression from PBLs. Genes are ordered by the magnitude of the difference in the number of individuals with more maternal expression than paternal expression in LCLs.

<https://doi.org/10.1371/journal.pone.0203906.g002>

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. [29,30].

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 partial methylation range, further suggesting that these two genes are indeed imprinted (Fig 3). To provide proof of principle, we also looked at methylation near a known imprinted gene that was identified with our test, *PEG10* (Fig 3A). *PXDC1* does not have a methylated promoter, suggesting long-range regulation by the *FAM50B* DMR.

Discussion

Dysregulation of imprinted genes can have a large impact on mammalian development and has been associated with significant diseases in humans [1,2]. Studies aimed at identifying imprinted genes at genome-wide levels have used allele specific expression and imbalance to infer parent of origin [14,15]. 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. Surprisingly, *CDKN1C* showed patterns opposite of what has been reported [31,32]. This could be due to the small sample (only three individuals showed

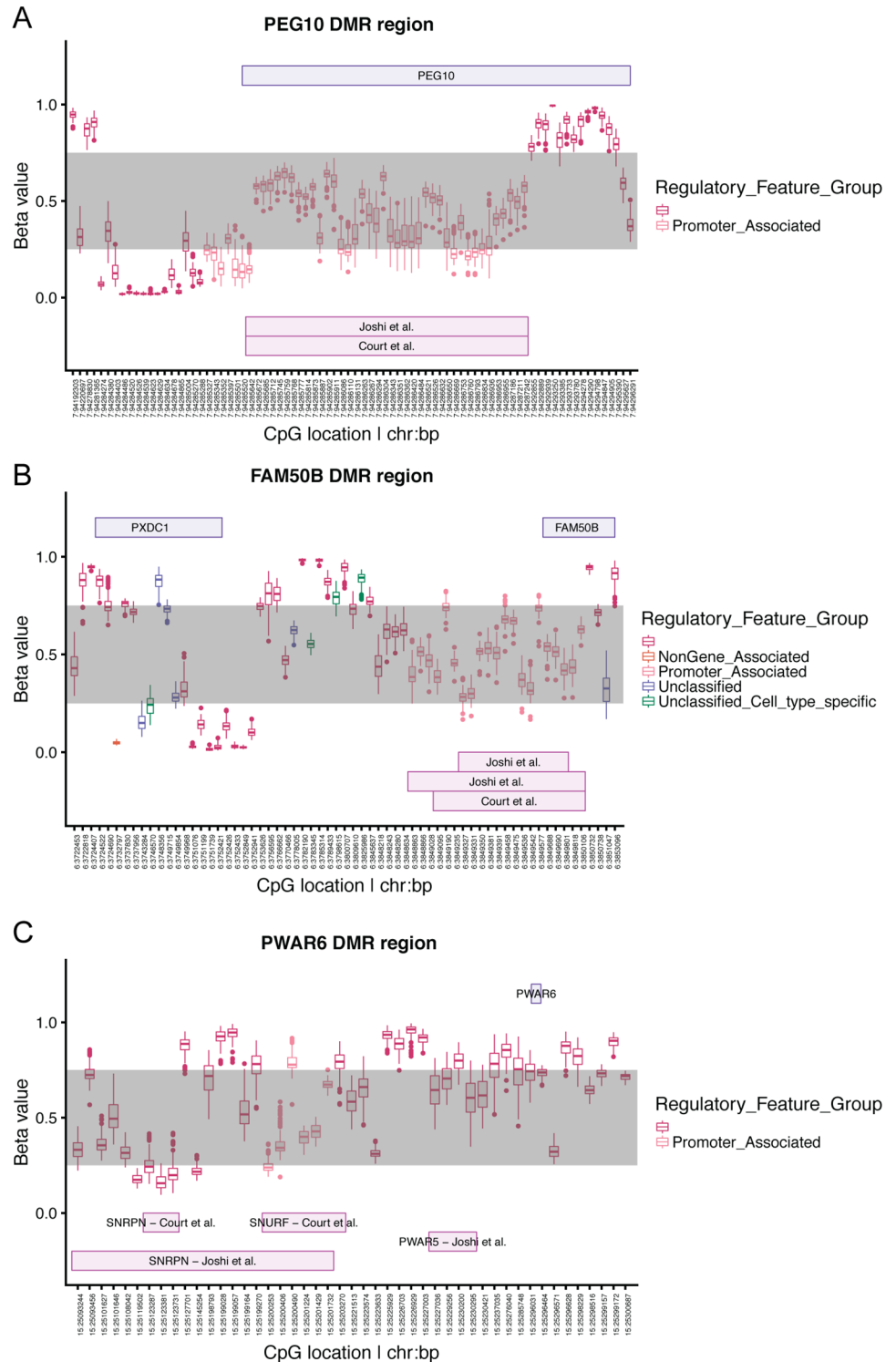


Fig 3. DNA methylation levels near known and novel candidate imprinted genes previously defined by Joshi et al. and Court et al. (A) *PEG10*, (B) *PXDC1* and *FAM50B*, (C) *PWAR6*. The x-axis shows chromosome and basepair location of CpGs. The y-axis is the beta value of methylation between 0 and 1. The grey interval marks beta values of methylation between 0.25 and 0.75. The location of genes are in purple boxes above methylation values and previously defined DMRs are in pink boxes below methylation values. Each boxplot represents beta values for all individuals at that CpG. The boxplot itself represents 50 percent of the data with box edges defining the 25th and 75th percentiles of the data. The line in the box represents the median of the methylation data.

<https://doi.org/10.1371/journal.pone.0203906.g003>

expression from one parent) or to the different cell types used here (LCLs) and in previous studies (developing brain and embryonal tumors for *CDKN1C*).

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. One of the previously characterized genes, *IGF1R*, has been reported to be both imprinted and not imprinted in humans [21–23,33]. Both of the 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 [29,30], we estimated regions for defining partial 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 [1,2,14,34]. 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. It is also possible that short-read RNA-seq cannot discriminate isoforms with alternate transcription start sites so some genes with both biallelic and imprinted transcripts will be missed in our study. Lastly, although we characterized the gene expression and methylation patterns for two candidate 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 candidate 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 potentially novel imprinted genes in this study. Our method and study population allowed us to map reads to parental haplotypes and uncovered *PWAR6* and *PXDC1* as candidate 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. The accession number for the Hutterite data reported in this paper is dbGAP:phs000185.

RNA-seq in lymphoblastoid cell lines (LCLs)

RNA-seq was performed in LCLs as previously described [35]. 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 [36,37]. To remove mapping bias, reads were processed and duplicate reads removed using WASP [38]. 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 [39]. 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. Gene counts were not normalized or standardized since we are using gene count data and comparing maternal and paternal gene expression from the same individual.

RNA-seq in peripheral blood leukocytes (PBLs)

RNA-seq was performed in whole blood as previously described [40]. 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 to detect genes that are more highly expressed and potentially imprinted. We used individuals and genes that had a parental gene count of at least 5. 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 All-Prep 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 [41],

normalized using SWAN (subset within-array normalization [42]) and quantile normalized similar to previous methylation studies [43]. 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 [44]. Principal components analysis (PCA) was used to identify significant technical covariates, and the ComBat function [45] within the R package sva [46] 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 [47].

Supporting information

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

(XLSX)

S2 Table. Parent of origin SNPs. SNPs used to assign parental gene expression for genes in Table 1 and in S1 Table.

(XLSX)

S1 Fig. Expression of genes with parent of origin asymmetry. Plots of maternal (x-axis) and paternal (y-axis) gene expression for remaining genes with parent of origin asymmetry.

(PDF)

S2 Fig. Differentially methylated regions. Density plot for Differentially Methylated Regions (DMRs) for imprinted genes from Joshi et al and Court et al. with beta values on the x-axis.

(TIF)

S3 Fig. Allelic expression of *PWAR6* across all individuals. Number of reads used to assign reads to a parent across all the SNPs in gene *PWAR6*. SNPs are along the x-axis with chromosome and base pair location. The y-axis is gene expression in gene counts across all individuals. The grey is paternal expression and is negative on the y-axis. The red is maternal expression and is positive on the y-axis.

(PDF)

S4 Fig. Allelic expression of *PXDC1* across all individuals. Number of reads used to assign reads to a parent across all the SNPs in gene *PXDC1*. SNPs are along the x-axis with chromosome and base pair location. The y-axis is gene expression in gene counts across all individuals. The grey is paternal expression and is negative on the y-axis. The red is maternal expression and is positive on the y-axis.

(PDF)

Acknowledgments

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.

Author Contributions

Conceptualization: Sahar V. Mozaffari, Dan L. Nicolae, Carole Ober.

Data curation: Sahar V. Mozaffari, Michelle M. Stein, Kevin M. Magnaye.

Formal analysis: Sahar V. Mozaffari.

Funding acquisition: Sahar V. Mozaffari, Carole Ober.

Investigation: Sahar V. Mozaffari.

Methodology: Sahar V. Mozaffari, Dan L. Nicolae.

Project administration: Sahar V. Mozaffari, Dan L. Nicolae, Carole Ober.

Resources: Carole Ober.

Software: Sahar V. Mozaffari.

Supervision: Dan L. Nicolae, Carole Ober.

Validation: Sahar V. Mozaffari, Michelle M. Stein.

Visualization: Sahar V. Mozaffari.

Writing – original draft: Sahar V. Mozaffari, Dan L. Nicolae, Carole Ober.

Writing – review & editing: Sahar V. Mozaffari, Michelle M. Stein, Kevin M. Magnaye, Dan L. Nicolae, Carole Ober.

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. [https://doi.org/10.1016/S0002-9440\(10\)65309-6](https://doi.org/10.1016/S0002-9440(10)65309-6) PMID: 10079240
2. Peters J. The role of genomic imprinting in biology and disease: an expanding view. *Nat Rev Genet*. 2014; 15: 517–530. <https://doi.org/10.1038/nrg3766> PMID: 24958438
3. Kalish JM, Jiang C, Bartolomei MS. Epigenetics and imprinting in human disease. *Int J Dev Biol*. 2014; 58: 291–298. <https://doi.org/10.1387/ijdb.140077mb> PMID: 25023695
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. <https://doi.org/10.1038/ncomms13490> PMID: 27848971
5. Zoledziwska 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. <https://doi.org/10.1038/ng.3403> PMID: 26366551
6. Barlow DP, Bartolomei MS. Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*. 2014; 6: a018382–a018382. <https://doi.org/10.1101/cshperspect.a018382> PMID: 24492710
7. Abramowitz LK, Bartolomei MS. Genomic imprinting: recognition and marking of imprinted loci. *Curr Opin Genet Dev*. 2012; 22: 72–78. <https://doi.org/10.1016/j.gde.2011.12.001> PMID: 22195775
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. <https://doi.org/10.1038/ng.3274> PMID: 25848752
9. Cattanach BM, Kirk M. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature*. Nature Publishing Group; 1985; 315: 496–498. <https://doi.org/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. [https://doi.org/10.1016/0168-9525\(89\)90138-8](https://doi.org/10.1016/0168-9525(89)90138-8)
12. Kong A, Steinthorsdottir 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. <https://doi.org/10.1038/nature08625> PMID: 20016592
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. <https://doi.org/10.1371/journal.pone.0041695> PMID: 22912676
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. <https://doi.org/10.1101/gr.192278.115> PMID: 25953952

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. <https://doi.org/10.1016/j.ajhg.2017.01.028> PMID: 28190458
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. <https://doi.org/10.1371/journal.pcbi.1004139> PMID: 25735005
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. <https://doi.org/10.1101/gr.6584707> PMID: 18055845
18. 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. <https://doi.org/10.1136/jmedgenet-2013-102116> PMID: 24501229
19. 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. <https://doi.org/10.1073/pnas.0308195101> PMID: 15123827
20. Daelemans C, Ritchie ME, Smits G, Abu-Amro 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. <https://doi.org/10.1186/1471-2156-11-25> PMID: 20403199
21. 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. <https://doi.org/10.1093/nar/gku549> PMID: 25092925
22. 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. <https://doi.org/10.1016/j.ejca.2014.10.031> PMID: 25465188
23. 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. <https://doi.org/10.1073/pnas.1415475111> PMID: 25246545
24. Adhami AI 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. <https://doi.org/10.1101/gr.175919.114> PMID: 25614607
25. 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. <https://doi.org/10.1096/fj.13-247049> PMID: 24696301
26. 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. <https://doi.org/10.1371/journal.pone.0163915> PMID: 27760153
27. 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. <https://doi.org/10.1002/jcp.25330> PMID: 26831620
28. 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. <https://doi.org/10.1101/gr.893803> PMID: 12727900
29. 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. <https://doi.org/10.1101/gr.164913.113> PMID: 24402520
30. 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. <https://doi.org/10.1016/j.ajhg.2016.06.032> PMID: 27569549
31. Hatada I, Mukai T. Genomic imprinting of p57KIP2, a cyclin-dependent kinase inhibitor, in mouse. *Nat Genet. Nature Publishing Group;* 1995; 11: 204–206. <https://doi.org/10.1038/ng1095-204> PMID: 7550351
32. 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.

33. Kalscheuer VM, Mariman EC, Schepens MT, Rehder H, Ropers HH. The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat Genet.* 1993; 5: 74–78. <https://doi.org/10.1038/ng0993-74> PMID: 8220428
34. Plasschaert RN, Bartolomei MS. Genomic imprinting in development, growth, behavior and stem cells. *Development.* 2014; 141: 1805–1813. <https://doi.org/10.1242/dev.101428> PMID: 24757003
35. 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. <https://doi.org/10.1093/hmg/ddw061> PMID: 26931462
36. 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. <https://doi.org/10.1002/0471250953.bi1114s51> PMID: 26334920
37. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* 2011; 17: 10. <https://doi.org/10.14806/ej.17.1.200>
38. 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. <https://doi.org/10.1038/nmeth.3582> PMID: 26366987
39. 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. <https://doi.org/10.1016/j.ajhg.2012.09.004> PMID: 23103226
40. 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. <https://doi.org/10.1056/NEJMoa1508749> PMID: 27518660
41. 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. <https://doi.org/10.1093/bioinformatics/btu049> PMID: 24478339
42. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol.* 2012; 13: R44. <https://doi.org/10.1186/gb-2012-13-6-r44> PMID: 22703947
43. 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. <https://doi.org/10.1172/jci.insight.90151> PMID: 27942592
44. 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. <https://doi.org/10.1371/journal.pgen.1004663> PMID: 25233095
45. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007; 8: 118–127. <https://doi.org/10.1093/biostatistics/kxj037> PMID: 16632515
46. 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. <https://doi.org/10.1093/bioinformatics/bts034> PMID: 22257669
47. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 2008; 24: 1547–1548. <https://doi.org/10.1093/bioinformatics/btn224> PMID: 18467348