Genome-Wide Detection of Imprinted Differentially Methylated Regions Using Nanopore Sequencing

Vahid Akbari[1,2], Jean-Michel Garant[1], Kieran O'Neill[1], Pawan Pandoh[1], Richard Moore[1], Marco A. Marra[1,2], Martin Hirst[1,3], Steven J.M. Jones[1,2*].

1- Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

2- Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

3- Department of Microbiology and Immunology, Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada

* Author for correspondence: Steven J.M. Jones (sjones@bcgsc.ca)

Abstract

Imprinting is a critical part of normal embryonic development in mammals, controlled by defined parent-of-origin (PofO) differentially methylated regions (DMRs) known as imprinting control regions. As we and others have shown, direct nanopore sequencing of DNA provides a mean to detect allelic methylation and to overcome the drawbacks of methylation array and short-read technologies. Here we leverage publicly-available nanopore sequence data for 12 standard B-lymphocyte cell lines to present the first genome-wide mapping of imprinted intervals in humans using this technology. We were able to phase 95% of the human methylome and detect 94% of the well-characterized imprinted DMRs. In addition, we found 28 novel imprinted DMRs (12 germline and 16 somatic), which we confirmed using whole-genome bisulfite sequencing (WGBS) data. Analysis of WGBS data in mouse, rhesus, and chimp suggested that 12 of these are conserved. We also detected subtle parental methylation bias spanning several kilobases at seven known imprinted clusters. These results expand the current state of knowledge of imprinting, with potential applications in the clinic. We have also demonstrated that nanopore long reads, can reveal imprinting using only parent-offspring trios, as opposed to the large multi-generational pedigrees that have previously been required.

Introduction

The addition of a methyl group to the 5-carbon of cytidine is the most prevalent and stable epigenetic modification of human DNA (Laurent et al., 2010). DNA methylation is involved in gene regulation and influences a vast array of biological mechanisms, including embryonic development and cell fate, genome imprinting, X-chromosome inactivation, and transposon silencing (Moore et al., 2013; Smith and Meissner, 2013). In mammals, there are two copies or alleles of a gene, one inherited from each parent. Most gene transcripts are expressed from both alleles. However, there is a subset of genes which are expressed from a single allele either randomly such as in X-inactivation, or based upon PofO. The latter is known as imprinting (Chess, 2013; Khamlichi and Feil, 2018).

In imprinting, mono-allelic expression of a gene or cluster of genes is controlled by a *cis-acting* imprinting control region (ICR) (Bartolomei and Ferguson-Smith, 2011). The main mechanism by which this occurs is PofO-defined differential methylation at ICRs, also known as imprinted differentially methylated regions (DMRs) (Bartolomei and Ferguson-Smith, 2011; Maupetit-Méhouas et al., 2016). ICRs are classified as germline (or primary) or somatic (or secondary), hereinafter referred to as gDMR and sDMR. Germline ICRs are established during the first wave of methylation reprogramming at germ cell development and escape the second methylation reprogramming after fertilization (Zink et al., 2018). Secondary or somatic ICRs are established *de-novo* after fertilization during somatic development, usually under the control of a nearby primary ICR (Zink et al., 2018). Imprinted clusters of genes may span up to ~4 Mb, by acting through a CCCTC-binding factor (CTCF) binding site or by allelic expression of a long non-coding RNA (Bartolomei and Ferguson-Smith, 2011; da Rocha and Gendrel, 2019). By contrast, individually-imprinted genes are typically regulated by PofO-derived differential methylation at the gene promoter (Bartolomei and Ferguson-Smith, 2011).

Imprinting is implicated in various genetic disorders, either from aberrations in imprinting itself, or from deleterious variants affecting the expressed allele at an ICR. Loss of imprinting is also widely observed in human cancers (Goovaerts et al., 2018; Jelinic and Shaw, 2007; Tomizawa and Sasaki, 2012). Thus, accurate mapping and characterization of ICRs in humans is key to the treatment and actionability of genetic disorders, and to personalized oncogenomonics.

To detect ICRs, accurate assignment of methylation data to paternal and maternal alleles is required. Achieving this with traditional bisulfite sequencing or arrays is challenging. Several studies have used samples with large karyotypic abnormalities, such as uniparental disomies (UPDs), teratomas, and hydatidiform moles, to infer regions of imprinting [14–16]. This approach relies on rare structural variants, but also on the assumption that both normal methylation and the imprinted state remain intact in spite of substantial genomic aberrations. A much larger study by Zink *et al.* leveraged a genotyped, multi-generation pedigree spanning nearly half the population of Iceland (n=150,000), in combination with whole genome oxidative bisulfite sequencing (oxBS-Seq), to phase methylation and infer parent-of-origin (Zink et al., 2018). However, despite being able to phase nearly every SNP in that cohort, they were only able to phase 84% of CpG methylation (CpGs on chromosomes 1-22) in over 200 samples due to the short length of reads. Further, that study was based on a single, genetically-isolated population, which may not be representative of the wider human population. A comprehensive mapping of ICRs using a technology more suited to phasing reads, based on individuals more representative of the human population, could greatly advance our understanding of imprinting, with direct benefits for human health.

We have previously shown that nanopore sequencing can detect allelic methylation in a single sample and accurately determine PofO using only trio data. We also developed the software NanoMethPhase for this purpose (Akbari et al., 2021). Here, we applied NanoMethPhase to public nanopore data from a diverse set of 12 lymphoblastoid cell lines (LCLs) from the 1,000 Genomes Project (1KGP) and Genome in a Bottle (GIAB) to investigate genome-wide allele-specific methylation (ASM) and detect novel DMRs

(Figure 1A) (De Coster et al., 2019; Shafin et al., 2020; Zook et al., 2019, 2016). Using trio data from 1KGP for these cell lines we phased nanopore long reads to their PofO and inferred allelic methylation (Akbari et al., 2021; Auton et al., 2015). Nanopore was able to detect haplotype and methylation status for 26.5 million autosomal CpGs (Chromosomes 1-22), which represents 95% of the autosomal CpGs in the GRCh38 (Kent et al., 2002). We further used public whole-genome bisulfite sequencing (WGBS) data to confirm the presence of the detected DMRs in other tissues and to class the novel DMRs as being germline or somatic. We captured 94% of the well-characterized DMRs (Those reported by at least two studies) and detected 28 novel DMRs (12 germline and 16 somatic). We determined that 43% of these novel DMRs show evidence of conservation in rhesus and chimp. Collectively, our results extend the set of known imprinted intervals in human and demonstrate a major contribution in our ability to characterize imprinting by ASM, brought about by the capabilities of nanopore sequencing.

Results

Assessing the Effectiveness of Nanopore Methylation Calling and Detection of Known Imprinted DMRs

We performed correlation analysis among cell lines and NA12878 nanopore-called methylation with WGBS data (ENCFF835NTC) to confirm the reliability of methylation calling (Figure 1B). We observed high correlation across cell lines (r = 0.75-0.93), as expected due to their being the same cell type. NA12878 nanopore-called methylation also showed the highest correlation ($r = 0.89$) with NA12878 WGBS (Figure 1B), as expected. Additionally, to examine performance on detection of known DMRs, we gathered the list of reported DMRs from previous studies (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). This included 383 imprinted intervals, of which 68 we assigned as "well-characterized" because they were reported by at least two genome-wide mapping studies or were previously known to be imprinted (Supplementary file1). Subsequently, we haplotyped the methylome in each cell line, performed differential methylation analysis (DMA) between alleles across cell lines. 95% (26.5M) of human autosomal CpGs could be assigned to a haplotype. We detected 172 allelic DMRs (p-value < 0.001, |methylation difference| > 0.25, and detected in at least 4 cell lines in each haplotype). See supplementary file 2 for more details. Of the 172 detected DMRs, 96 (56%) overlapped with at least one previously reported, while the remaining 76 (44%) were novel. Of the well-characterized DMRs (those detected in at least two previous whole-genome mapping studies or known from prior evidence), 64/68 (94%) were detected in our study (Figure 1C, supplementary file2). All DMRs which overlapped with previously-reported DMRs displayed consistent PofO with those studies.

We similarly examined the power of nanopore sequencing to detect allelic DMRs within a single sample, by comparing to previous studies (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). On average, 88% (M ± SD = 24.5M ± 1.7M) of the human methylome could be assigned to

a parental haplotype in each LCL. Of the well-characterized DMRs, ~71% (M ± SD = 48 ± 4.8) could be detected in a single LCL. An additional 32 DMRs (SD = 9.7) reported by only one previous study were detected in each sample (Supplementary figure S1).

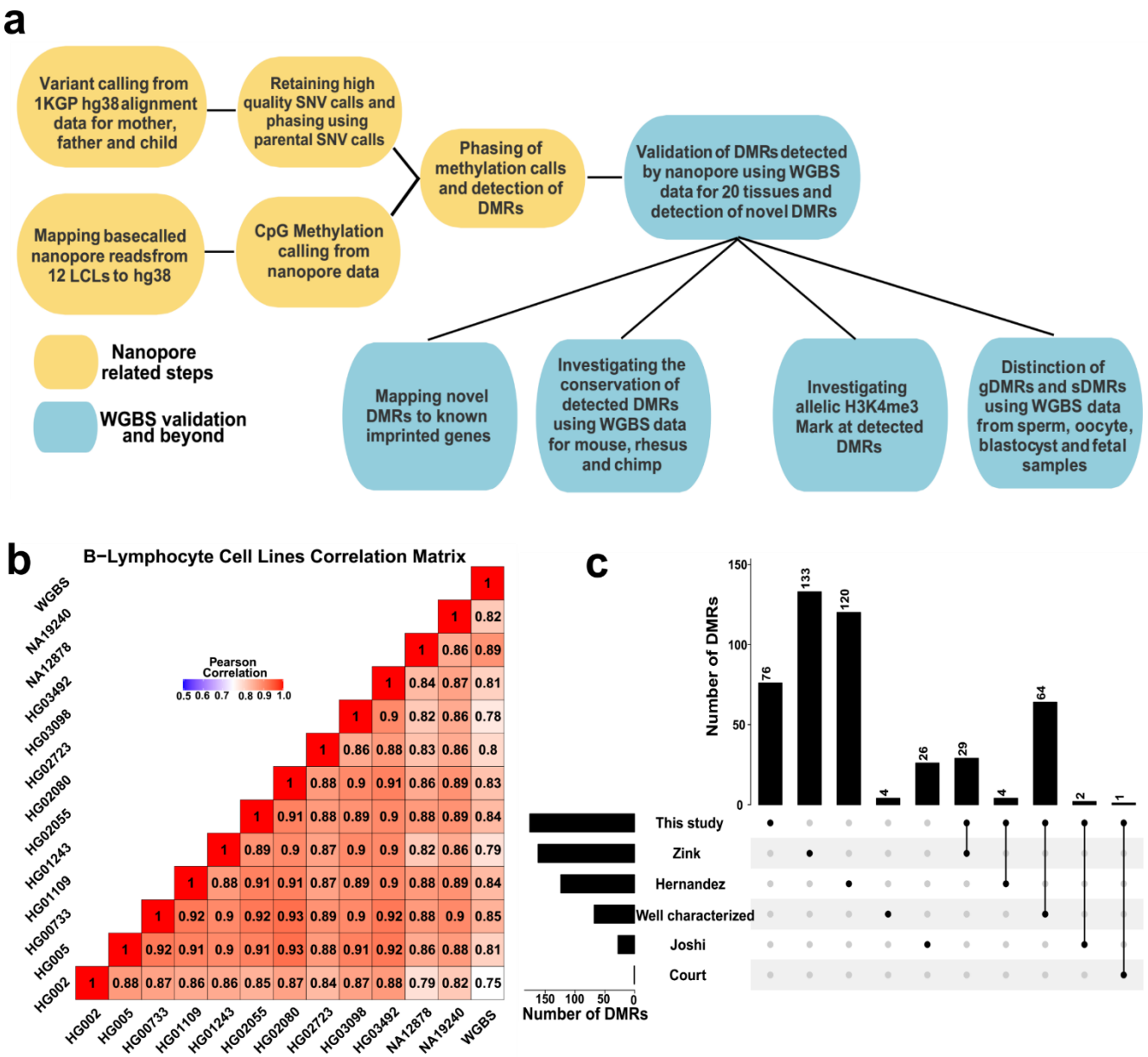

Figure 1: Detection of allelic methylation using nanopore sequencing. a) The flowchart of the study representing all the analysis steps. b) Pearson correlation matrix of the nanopore CpG methylation frequencies for the 12 LCLs and NA12878 whole-genome bisulfite sequencing. c) Upset plot of the number of DMRs detected by us and previous studies and their overlaps.

Confirmation of Novel Imprinted DMRs

As noted above, we detected 76 allelic DMRs that did not overlap with previously-reported ICRs (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). In order to determine their validity as novel DMRs, we used 24 WGBS datasets from 20 tissue samples within the Roadmap Epigenomics Project (See materials and methods) (Bernstein et al., 2010). We first examined the 96 allelic DMRs which overlapped with the reported DMRs. 79 out of 96 DMRs that overlapped with reported regions showed adjusted p-value < 0.000005 and log2 fold change > 0.15, while only 5, 6, 7, and 8 intervals were detected as significant in the control intervals including 200 randomly selected 1kb bins, CpG islands, 2kb, and 3kb bins, respectively (Figure 2A, Supplementary file 3). Applying this approach to the 76 not previously reported DMRs, the WGBS data supported 28 significant DMRs (Figure 2A and 2B, Supplementary file 3). In agreement with previous studies reporting higher number of maternally methylated intervals, 10 of the 28 novel DMRs were paternally methylated and 18 were maternally methylated (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016). Overall, 107 out of 172 DMRs were validated in tissue WGBS data from which 28 were novel and 79 were reported by the previous studies (Figure 2C, Supplementary file 2) (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018).

We also sought to examine the significance of the other 283 previously reported imprinted regions which did not overlap with our detected DMRs. We examined these 283 DMRs in WGBS data and only 139/283 DMRs (49%) were significant (adjusted p-value < 0.000005 and log2 fold change > 0.15. Supplementary file 4). We also mapped these 283 intervals to the DMRs detected in each LCL sample. 81/283 (27%) of them were detected in at least one sample with consistent reported PofO, of which 41 were in common with WGBS analysis (Supplementary file 5).

Determination of Germline vs Somatic Status of Novel Imprinted DMRs

We performed DMA between oocyte and sperm and overlapped detected DMRs to the 28 novel DMRs. 12 of the novel DMRs overlapped with DMRs from oocyte versus sperm (p-value < 0.001. More than 40% methylation in oocyte and less than 20% in sperm and vice versa) from which 11 were maternally methylated and 1 was paternally methylated (Figure 3). We then examined the methylation of somatic and germline DMRs in early human embryonic cells and fetal tissues to investigate whether the imprinting of the 12 candidate gDMRs survived the second round of de- and re-methylation and if the other 16 novel sDMRs were established during development. We used blastocyst WGBS data from early cleavage-stage embryos and fetal tissue (Bernstein et al., 2010; Okae et al., 2014). All novel candidate gDMRs showed partial methylation in the blastocyst indicating the gDMRs escaped de-methylation after fertilization (Figure 3). All novel gDMRs and sDMRs displayed partial methylated in fetal tissues indicating survival of gDMRs during somatic development and establishment of sDMRs. Overall, 12 of the novel DMRs detected to be germline while 16 detected as sDMRs (Figure 2C and Figure 3, Supplementary file 6).

148

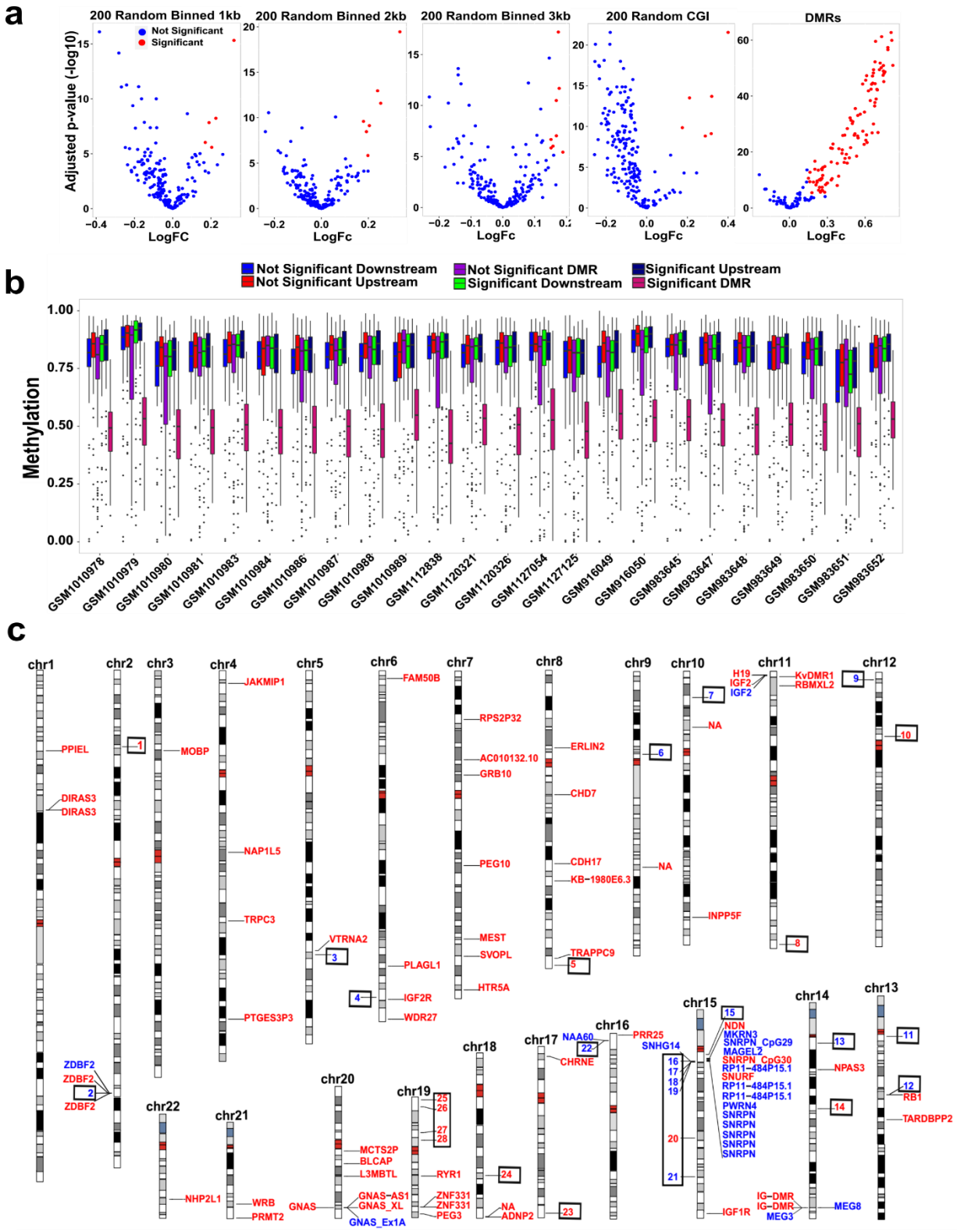

Figure 2: Validation of detected DMRs using WGBS data. a) Limma results for partial methylation (30%-70%) of DMRs detected using nanopore in 20 tissue samples WGBS data along 200 randomly selected CpG islands, 1kb, 2kb, and 3kb intervals as controls. Red dots adjusted p-value < 0.000005 and log2 fold change > 0.15. b) Box blot showing partial methylation at significant DMRs while not significant DMRs and adjacent regions (down- and upstream to significant and not significant DMRs) are not partially methylated. c) Idiogram of the 107 DMRs which validated by WGBS. On the left on each chromosome are paternally methylated DMRs and on the right are maternally methylated DMRs. Red color represents gDMRs and blue represents sDMRs. Novel DMRs are boxed and numbered.

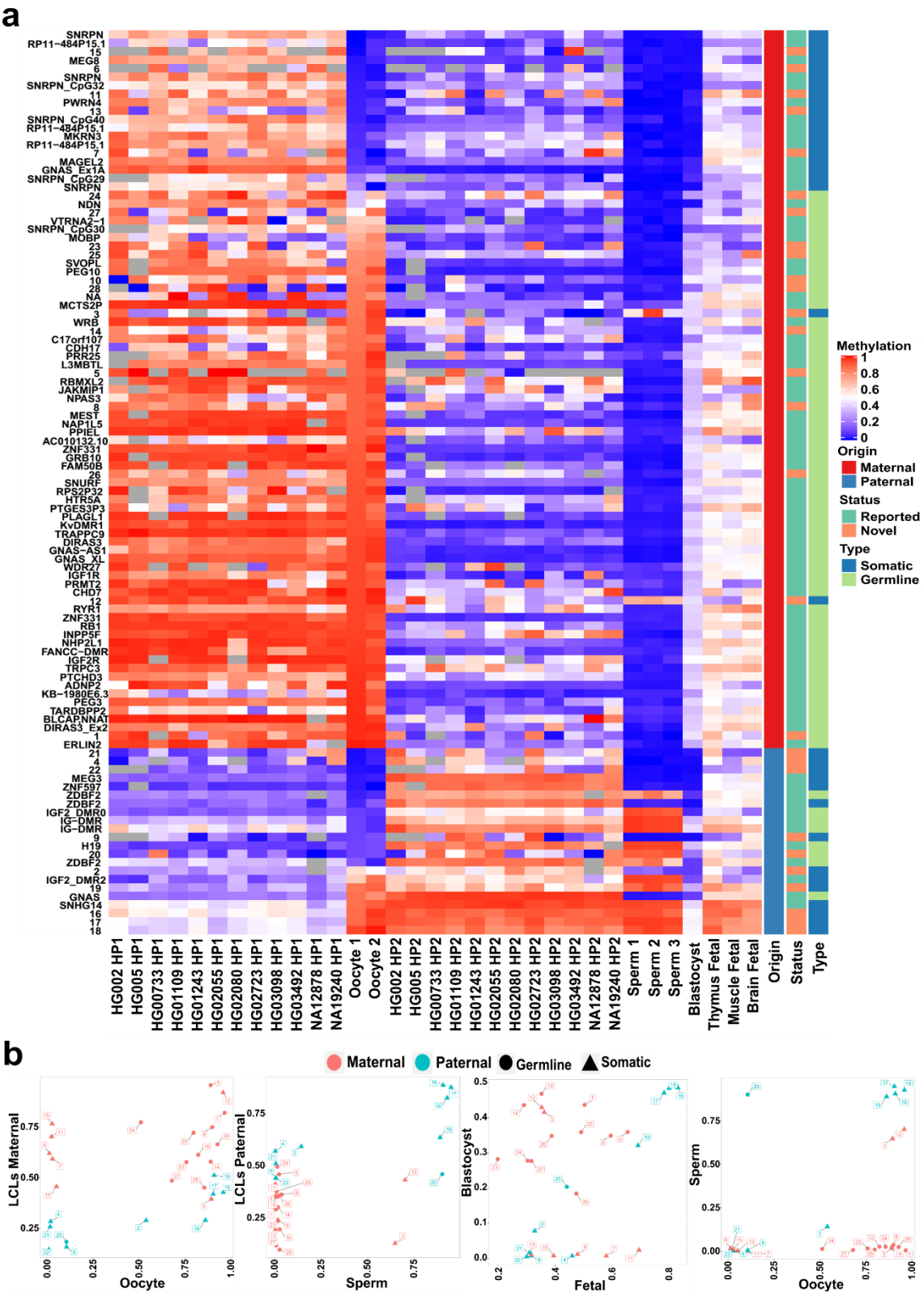

Figure 3: Detection of novel gDMRs and sDMRs. a) Heatmap displaying average methylation of the 107 DMRs validated by WGBS in nanpore LCLs and WGBS samples and their overlapping DMRs from WGBS data for sperm and oocyte. b) Dot plots representing the methylation of novel gDMRs and sDMRs in each sample in respect to other sample. Maternally methylated germline DMRs display high methylation in oocyte and very low or no methylation in sperm and are partially methylated in blastocyst and fetal samples. Paternally methylated germline DMRs display high methylation in sperm and very low or no methylation in oocyte and are partially methylated in blastocyst and fetal samples. Somatic DMRs do not display relevant methylation in sperm or oocyte and are methylated or unmethylated in both sperm and oocyte while they display parental methylation bias in LCLs and partial methylation in fetal samples.

Allelic Histone Methylation of H3K4 is Enriched at Germline DMRs

The H3K4me3 histone mark is protective to DNA methylation. At ICRs, the unmethylated allele is usually enriched for this histone modification (Court et al., 2014; John and Lefebvre, 2011). We used H3K4me3 chromatin immunoprecipitation sequencing (ChIP-seq) data for 7 LCLs and their heterozygous single-nucleotide variant (SNV) calls from 1KGP. 81/107 of the detected DMRs could be examined (See material and methods). Of these, 43 reported and 9 novel DMRs showed a significant allelic count in ChIP-seq data (Fisher's combined p-value binomial < 0.01) (Supplementary files 6 and 7). Among the 7 LCLs with ChIP-seq data, only NA12878 and NA19240 were among LCLs with nanopore data and a phased methylome. Therefore, we examined if the allelic H3K4me3 and methylation are in opposite alleles in these cell lines. 23 reported and 5 novel DMRs were significant for allelic H3K4me3 in NA12878 and/or NA19240. 21 reported and 4 novel DMRs showed opposite allelic state between H3K4me3 and methylation (Supplementary file 7).

Allelic H3K4me3 mostly overlapped with gDMRs. Overall, 77% of assessable gDMRs and 39% of sDMRs were significant for allelic H3K4me3. This is consistent with previous studies demonstrating the protective role of H3K4me3 against DNA methylation, specifically at germline ICRs in the second round of re-methylation during implantation and somatic development (Chen and Zhang, 2020; Hanna and Kelsey, 2014).

Conservation of Detected Imprinted DMRs across Mammals

To investigate the conservation of detected DMRs and determine if any of the novel DMRs are conserved in mammals we used WGBS data from house mouse (Mus musculus) , rhesus macaque (Macaca mulatta) , and chimpanzee (Pan troglodytes) (Hon et al., 2013; Tung et al., 2012). We examined whether any of the orthologous regions in these mammals display significant partial methylation (Materials and Methods). Of the 107 DMRs detected by nanopore and validated in WGBS data, 71, 105 and all 107 had orthologs

in mouse, rhesus and chimp, respectively. Orthologs of the 77/107 detected DMRs showed significant partial methylated in at least one of the three mammals (Figure 4A, Supplementary file 6). Of these, 65 were reported DMRs (56 well-characterized) and 12 novel DMRs. We detected 24 significant orthologous DMRs in mouse. 18 of these were reported to be imprinted by previous studies in mouse (Gigante et al., 2019; Xie et al., 2012). All significant DMRs in mouse, except one, were also significant in rhesus and/or chimp suggesting their existence in their common ancestor. These DMRs mapped to well-known imprinted clusters including *KCNQ1*, *H19*, *GNAS*, SNURF/*SNRPN*, *PLAGL*, *SGCE*, *BLCAP*, *PEG3*, *PEG10*, *PEG13*, *GRB10*, *BLCAP*, *NAP1L5*, *INPP5F*, and *MEG3* where their allelic PofO expression has already been reported in mouse and other mammals ("Geneimprint," 2021; Morison et al., 2001).

Sperm, oocyte and embryo WGBS data for mouse and rhesus were used to investigate if the DMRs that detected as germline or somatic in human are germline or somatic in these mammals and vice versa (Dahlet et al., 2020; Gao et al., 2017; Jung et al., 2017; Saenz-de-Juano et al., 2019). 62 of the human DMRs had significant orthologs in rhesus. Of these, 51 were germline and 11 were somatic in human and in rhesus 45 were germline and 17 were somatic (More than 40% methylation in oocyte and less than 20% in sperm and vice versa. Figure 4B). 24 human DMRs had significant orthologs in mouse. Of these, 21 were germline and 3 were somatic in human and in mouse 17 were germline and 7 were somatic (Figure 4B). Nine gDMRs in human were somatic in rhesus and/or mouse and three gDMRs from mouse or rhesus were somatic in human. This is consistent with previous studies indicating imprinting is largely conserved in mammals while ICR identity at the germline stage is not completely conserved (Cheong et al., 2015).

192

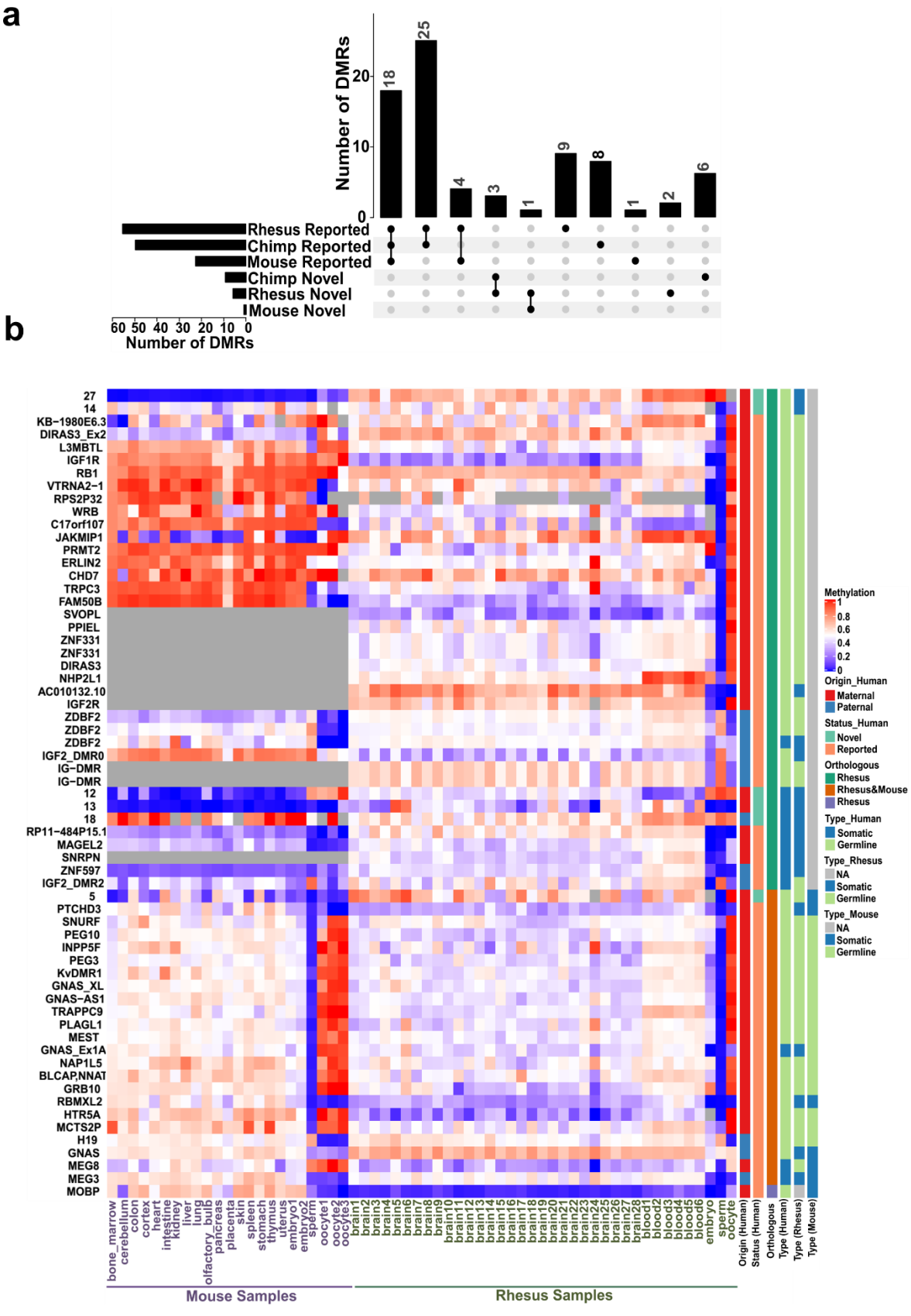

Figure 4: Conservation of detected DMRs. a) Upset plot representing the number of reported and novel DMRs with evidence of conservation (partial methylation) in each of the mammals. b) Heatmap representing the average methylation in orthologous intervals for mouse and rhesus in different tissues and also in sperm, oocyte, and embryonic samples.

Novel DMRs within Known Imprinted Gene Domains and Contiguous Blocks of Parental Methylation Bias

We gathered the list of 259 imprinted genes from previous studies (Supplementary file 8) (Babak et al., 2015; Baran et al., 2015; "Geneimprint," 2021; Jadhav et al., 2019; Morison et al., 2001; Zink et al., 2018). 14 novel DMRs (6 germline and 8 somatic) mapped close (<1.03Mb) to imprinted genes (Supplementary file 6 and 9).

Of the 8 sDMRs close imprinted genes, only one mapped to a CpG island, and that was a small (<300bp) CpG island ~13 Kb downstream of the maternally expressed *NAA60* gene (Supplementary figure S2). Four novel sDMRs (All paternally methylated) mapped in the Prader-Willi syndrome and Angelman syndrome (PWS/AS) cluster. Previous studies reported continuous subtle paternal methylation bias at the PWS/AS cluster (Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). Consistent with previous studies, the four novel sDMRs at this cluster were large (>5Kb) and seemed to constitute near-continuous paternal methylation spanning a ~200kb region. This included the *SNORD116* cluster genes and several other genes such as *PWAR1* and *6*, *PWARSN* and *IPW* (Supplementary figure S3). This paternally methylated somatic block is downstream of the maternally methylated germline *SNURF/SNRPN* ICR, which is associated with PWS and shows evidence of conservation in chimp, rhesus, and mouse. Moreover, the allele-specific expression (ASE) track from Zink *et al.* displayed strong paternal expression across this ~200kb region (Zink et al., 2018). Another three novel sDMRs mapped close *RB1/LPAR6*, *IGF2R* (Supplementary figures S4 and S5) and *GPR1-AS/ZDBF2*. The novel sDMR at *GPR1-AS/ZDBF2* were close to 2 known paternal gDMRs. Moreover, LCLs PofO methylation track at the *ZDBF2* gene body showed continuous subtle paternal bias. Together, these suggest a ~65kb paternally methylated block interrupted by unmethylated CpG island at *ZDBF2* promoter (Supplementary figure S6). In addition to blocks with novel DMRs, we sought to detect continuous block of parental methylation bias at other regions. We detected 5 other contiguous

blocks of imprinting at *ZNF331, KCNQ1OT1, GNAS, L3MBTL1* and *ZNF597/NAA60*, ranging from 35-58Kb in size (Supplementary figures S7-11).

All the six gDMRs within imprinted gene domains were maternally methylated and they all mapped to CpG islands except a DMR mapped in the *AC024940.1* (*OVOS2*) (Supplementary figures S12-16, Figure 5). Five of them mapped to known imprinted genes without previously reported DMR or a DMR with a much greater distance from the gene compared to our DMRs including *AC024940.1*, *ZNF714, DDA1*, *ADAMTSL5,* and *NAPRT* (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). A novel gDMR mapped to the promoter of *ZNF714* which is reported to be paternally expressed (Jadhav et al., 2019; Zink et al., 2018). Thus suggesting this DMR could be the potential ICR and directly suppress maternal allele by blocking its promoter (Figure 5). *AC024940.1* reported to be paternally expressed (Zink et al., 2018). A novel germline maternal DMR mapped near the end of the *AC024940.1* gene (encompassing the intron 38 to the start of exon 40) adjacent to a CTCF binding site (Supplementary figure S12). *DDA1* and *ADAMTSL5* have been previously reported to be maternally expressed and *NAPRT* has an isoform dependent expression origin (Babak et al., 2015; Zink et al., 2018). A gDMR mapped to the end and downstream of *DDA1* gene (Supplementary figure S13). For *ADAMTSL5* and *NAPRT*, gDMRs mapped close to these genes (<150Kb) (Supplementary figures S14 and S15).

234

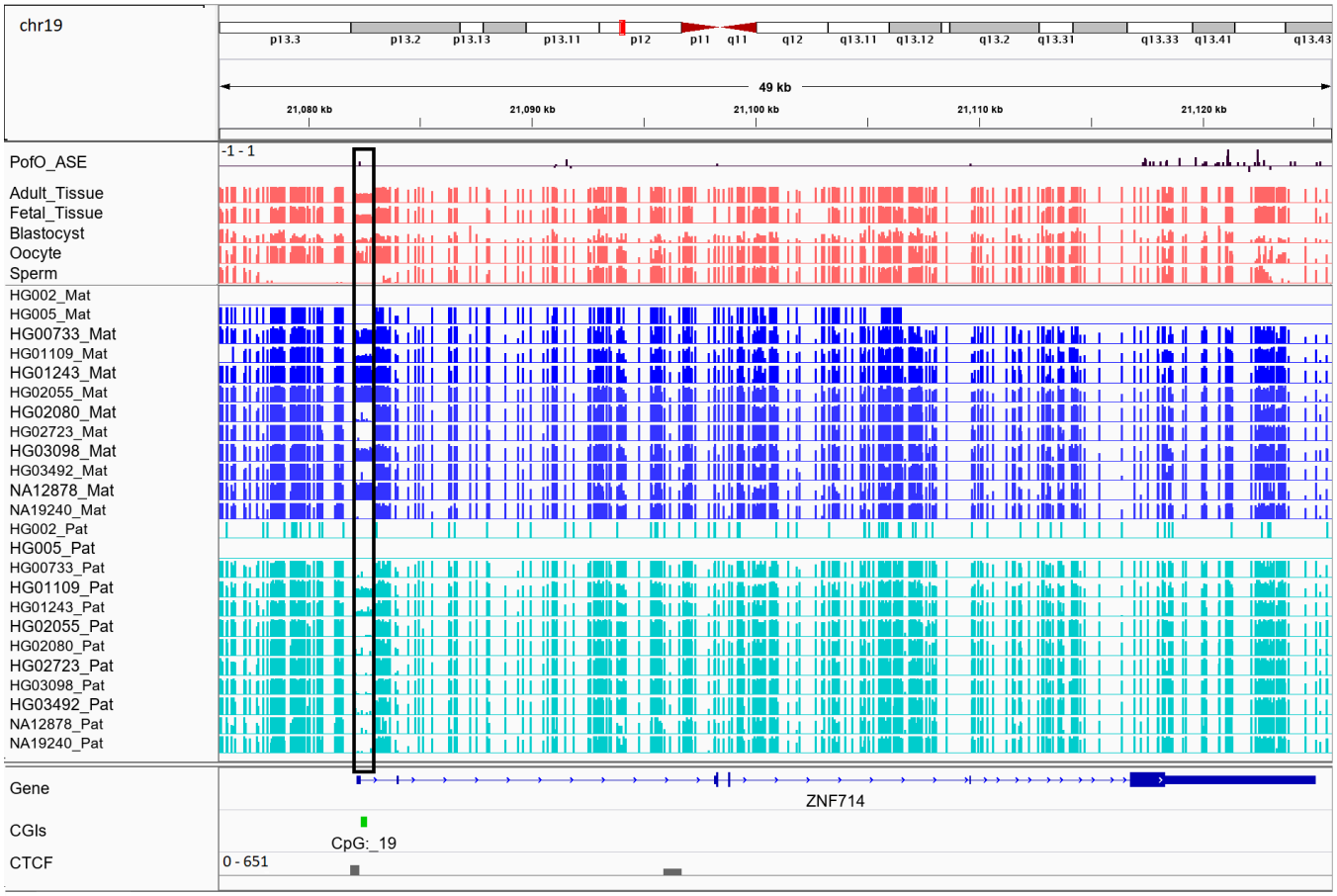

Figure 5: IGV screenshot of the novel maternally methylated germline DMR at the promoter of the paternally expressed ZNF714 gene. Black box region represents the DMR. PofO_ASE represents allele specific expression track from Zink *et al.* without any filtering for *P value* and positive vertical bars (upward) represents more paternal expression and negative bars (downward) maternal expression. The range for all methylation tracks is 0-1.

Discussion

Here we described the first genome-wide map of human allele-specific methylation using nanopore sequencing. Leveraging long reads and parental SNVs allowed us to phase methylation for ~26.5 million autosomal CpGs representing 95% of the CpGs in the human autosomal genome (GRCh38) across 12 LCLs (De Coster et al., 2019; Kent et al., 2002; Shafin et al., 2020; Zook et al., 2016). This represents a much higher resolution than previous studies aimed to capture allelic methylation (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). For example, Zink *et al.* determined the PofO of almost all genotypes using nearly half the population of Iceland (n=150,000) and used over 200 whole-genome OxBS-seq samples to detect imprinting (Zink et al., 2018). They could define PofO methylation for ~23.5 million autosomal CpGs (84%). We noticed three of our novel DMRs did not have any CpG representation from Zink *et al*. Moreover, in a further three other novel DMRs, fewer than 60% of the CpGs were captured in Zink *et al.* (Supplementary file 6). EPIC methylation arrays detect over 850k CpGs and covers almost all CpGs detectable by 450k and 27k methylation arrays. Seven of our novel DMRs did not have any CpGs covered by the EPIC array and 9 other novel DMRs had only 1 or 2 probes on this platform (Supplementary file 6). This highlights the breadth of nanopore sequencing for the purposes of ICR calling.

Even though we detected methylation for all the CpGs in the human genome (GRCh38), we were not able to phase 5% of the human methylome (Kent et al., 2002). To phase nanopore reads, we used SNVs detected from short-reads data in the 1KGP (Auton et al., 2015). Short-reads are challenging to map to complex repetitive regions which results in lack of SNVs and subsequent inability to phase reads in these regions. 75% of the unphased CpGs mapped to the ENCODE blacklist (Amemiya et al., 2019). We previously demonstrated that using SNVs detected from nanopore to phase reads results in reliable methylation phasing and detection of a few more reported DMRs (Akbari et al., 2021). Improvement in

basecalling and variant calling from nanopore reads could enable the phasing of a complete genome-wide methylome using nanopore detected SNVs.

Using nanopore sequencing we could capture 94% of the well-characterized DMRs and 35 of the DMRs reported by only one study (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). However, we were unable to detect a further 283 DMRs, mostly reported by one previous study (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). In further analyses, 180 of these DMRs were detected in at least one nanopore-sequenced LCL sample and/or validated in the Roadmap multi-tissue WGBS data we used (Supplementary files 4 and 5). We should note that nanopore data comes from a small number of B-lymphocyte cell-line samples, yet considerably diverse in ethnicity. Imprinted DMRs can be tissue-specific and polymorphic across individuals, which may explain this discrepancy (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Romanelli et al., 2014; Silver et al., 2015; Zink et al., 2018). Characterization of imprinted DMRs across a wider range of tissues and populations represents a clear path forward for the field. The ability of nanopore sequencing to characterize imprinting using only parent-offspring trios represents a relatively low-cost avenue by which this might be achieved.

We detected 107 DMRs using nanopore which were further confirmed in multi-tissue WGBS data. Twelve of these were novel gDMRs and sixteen were novel sDMRs not reported in previous studies (Court et al., 2014; Hernandez Mora et al., 2018; Joshi et al., 2016; Zink et al., 2018). These novel DMRs were supported by several lines of evidence in our analyses. 1) They displayed significant PofO methylation bias in nanopore LCLs. 2) They were significantly partially methylated in WGBS data from 20 human tissues. 3) gDMRs demonstrated escape from the second de-methylation step. 4) They were partially methylated in three fetal tissue samples. 5) 43% of those for which H3K4me3 ChIP-seq data could be phased showed significant allelic H3K4me3. 6) 43% showed evidence of conservation in at least one of the three mammals including chimp, rhesus, and mouse. 7) 71% mapped to at least one regulatory region

including CpG island, CTCF binding site and enhancer. These novel DMRs represent a substantial and well-validated expansion of known regions of imprinting, which may aid future research and diagnosis in the fields of genetic medicine and oncology.

Of the 107 DMRs, 20 mapped to the PWS/AS cluster. Previous studies demonstrated two paradigms of imprinting at this cluster, either PofO methylation confined to particular regulatory regions such as CpG islands or subtle paternal bias across this cluster with spikes of maternal methylation (Court et al., 2014; Joshi et al., 2016; Sharp et al., 2010; Zink et al., 2018). Although we did not observe paternal methylation bias across the whole PWS/AS cluster, we did detect a paternal methylation block spanning ~200Kb, immediately downstream of the known maternally methylated PWS *SNURF/SNRPN* ICR. This block encompasses the *SNORD116* cluster and other adjacent genes with strong paternal expression (Supplementary figure S3). Probes with paternal methylation bias at the *SNORD116* cluster have been reported which span about 95Kb region and paternal deletion of this cluster results in PWS phenotypes (Hernandez Mora et al., 2018; Joshi et al., 2016; Matsubara et al., 2019). Slight hypomethylation of *SNORD116* cluster in cases with PWS phenotype and hypermethylation in the cases with AS phenotype have been reported (Matsubara et al., 2019). Our analysis extends and more clearly delineates this paternally biased block.

Beyond the PWS/AS cluster, we detected another six blocks of allelic methylation bias (Supplementary figure S6-S11). All of the blocks represented several common features. 1) They were detected in imprinted genes that appeared in cluster. 2) All of them were accompanied by a strong PofO expression bias from the subtle hypermethylated allele. 3) There was at least one well-characterized and conserved gDMR in each block (except *ZNF597/NAA60* block with a conserved sDMR). 4) The well-characterized DMRs in these blocks displayed significant allelic H3K4me3 (except DMR in *L3MBTL1* block which could not be examined due to the lack of SNV). 5) Well-characterized DMRs in these blocks overlapped to the promoter of genes with subtle PofO methylation bias at the gene body and DMR itself displayed opposite

306 PofO methylation (except for *ZDBF2/GPR1-AS* block that DMR did not mapped to the promoter and had
307 the same PofO with the gene body). This represents a novel facet of imprinting biology. To explain this,
308 we can consider that CpG methylation at gene bodies is positively (but weakly) correlated with gene
309 expression (Ball et al., 2009; Yang et al., 2014). Within these blocks, we saw parental methylation bias at
310 the parentally expressed or active allele. This may suggest that subtle parental methylation is linked to
311 parental ASE. However, ASE is observed in many other imprinted genes whose gene bodies do not show
312 parental methylation bias. One possible explanation could be that the subtle parental methylation bias is
313 used by cells to express important genes (genes which can regulate other genes in the cluster or have
314 regulatory roles) in an imprinted cluster with higher fidelity through its gene body methylation on active
315 allele. For example, at the *KCNQ1OT1* and *GNAS* clusters the methylation blocks overlap *KCNQ1OT1*
316 and *GNAS-AS1* genes both of which encode antisense RNA transcripts that regulate other genes in the
317 imprinted cluster (Chiesa et al., 2012; Turan and Bastepe, 2013). However, further studies are needed to
318 reveal the mechanism producing these contiguous slight parental methylation bias blocks and their
319 functional role.

320 Orthologous regions of ~72% of the detected DMRs were demonstrated significant partial methylation in
321 at least one of the chimp, rhesus, and mouse. There were a considerably higher number of orthologous
322 sites and significant orthologous DMRs in chimp and rhesus in agreement with more similarities and less
323 distance to these primates compare to mouse in the human evolution. Ortologs of the 12 novel DMRs
324 were mostly displayed significant partial methylation in rhesus and/or chimp while the other 16 novel
325 DMRs were not significant in any of the examined mammals (Figure 4). This suggests that the novel
326 DMRs (except one which had significant orthologous in mouse) are established after divergence of
327 primates' common ancestor from mouse and majority of them established after the divergence of human
328 common ancestor from chimp. Court *et al.* detected 14 novel DMRs, at the time of their study, and did
329 not detect any imprinted orthologs of their novel DMRs in mouse (Court et al., 2014). All 14 also

overlapped with our detected DMRs and six of them had orthologous regions in mm10 using the UCSC liftover file (Kent et al., 2002). Two of the orthologs displayed partial methylation in mouse, one in *Rian* gene which did not examined in Court *et al*. and the other in *Htr5a* gene which reported not to be conserved in mouse by Court *et al.* (Court et al., 2014). When looking into their analysis, it seems that they examined different orthologous region (Supplementary figure S19). For *Htr5a*, they examined the CpG island (CpG:_102) ~50 kb away from the gene while we examined the region spanning the first or second exon (two transcripts) of *Htr5a* which was partially methylated while CpG:_102 was also unmethylated in our study.

Using reported imprinted genes, 50% of the novel DMRs mapped close to known imprinted genes (Babak et al., 2015; Baran et al., 2015; "Geneimprint," 2021; Jadhav et al., 2019; Morison et al., 2001; Zink et al., 2018). Five of our novel gDMRs could be potential ICRs for reported imprinted genes without reported ICR. Specifically, maternal methylation of CpG island overlapping promoter of *ZNF714* as it can directly repress maternal allele and results in the reported paternal expression (Figure 5) (Jadhav et al., 2019; Zink et al., 2018). *ZNF714* is a member of the zinc finger family proteins which have several imprinted genes with developmental roles (Babak et al., 2015; Baran et al., 2015; Camargo et al., 2012; Jadhav et al., 2019; Zink et al., 2018). ZNF714 has been reported to be associated with non-syndromic cleft lip (Camargo et al., 2012). Thus, this new imprinted DMR could be of potential clinical value. In contrast to imprinting which is established in the germline and usually consistent across tissues, allelic expression is only present if the imprinted gene is expressed in the tissue. Moreover, studies have used short read sequencing to detect ASE which is confounded with several limitations (Aird et al., 2011; Steijger et al., 2013). Therefore, a comprehensive ASE analysis using long-read technologies capturing various tissues might explain ASE around the novel DMRs without evidence of a close imprinted gene. Paternal expression bias of *PTCHD3* and maternal expression bias for *FANCC* are detected in Zink *et al*. while they could not detect any associated DMR (Zink et al., 2018). Hernandez *et al.* detected 3 and 1 maternally methylated

probes at the promoter of *PTCHD3* and intron one of *FANCC*, respectively, but were not able to examine the parental expression (Hernandez Mora et al., 2018). We also detected two maternally methylated gDMRs overlapping the promoter of *PCTHD3* and intron one of *FANCC* (Supplementary figures S17 and S18). There were no phased CpG for these DMRs in Zink *et al.* study (Supplementary file 6). Orthologous regions for the *PTCHD3* DMR were also detected to be partially methylated in all three mammals but the *FANCC* DMR was only partially methylated in chimp. These gDMRs could potentially explain the missing ICR for ASE of these genes. The gDMR at the *PTCHD3* promoter can directly suppress maternal allele. *FANCC* gDMR overlaps to a CpG island and CTCF binding site. CTCF is a methylation sensitive DNA-binding protein and CpG methylation can inhibit CTCF binding (Hashimoto et al., 2017; Renda et al., 2007). Moreover, CTCF binding to the first intron of major immediate-early gene of the human cytomegalovirus (HCMV) in HCMV-infected cells resulted in repression of this gene (Puerta et al., 2014). Therefore, the maternally methylated DMR in intron 1 of maternally expressed *FANCC* suggests a mechanism through which paternal allele is suppressed by CTCF binding at DMR while DNA methylation inhibits CTCF binding at maternal allele.

Overall, our study represents a near-complete genome-wide map of human allele-specific methylation by leveraging long-read nanopore technology. This allowed us to expand the set of reported imprinted DMRs using just 12 LCLs with parental SNPs and explain novel DMRs as potential ICRs for several imprinted genes with unknown ICR. 43% of the novel DMRs demonstrated partial methylation in other mammals suggesting their conservation. We detected seven large PofO bias methylation blocks spanning multiple kilobasesd and displaying several common features. We have suggested two avenues of further investigation: 1) Looking for tissue and individual polymorphism in imprinting, and 2) determining the mechanism and function of the subtle parental bias blocks. We have also shown that nanopore sequencing is a cheap and easy way to call ICRs and can open the way to answering those questions in future. This study provides a blueprint for further surveys using nanopore sequence data and demonstrates the potential

of this approach to study personalized allelic methylation in disease such as cancer with wide spread allelic methylation aberrations.

Materials and Methods

Nanopore Sequencing Data and Detection of Allele-Specific Methylation

We used publicly available nanopore sequencing data for 12 LCLs including HG002, HG005, HG00733, HG01109, HG01243, HG02055, HG02080, HG02723, HG03098, HG03492, NA12878, and NA19240 (ERR3219853 & ERR3219854) (De Coster et al., 2019; Shafin et al., 2020; Zook et al., 2016). All cell lines had trio information available from 1KGP or GIAB (Auton et al., 2015; Zook et al., 2019). Raw nanopore fast5 files along with basecalled fastq files for 12 LCLs were obtained and basecalled reads mapped to GRCh38 using Minimap2 with the setting *minimap2 –ax map-ont* (Kent et al., 2002; Li, 2018). Subsequently, CpG methylations were called using nanopolish with default parameters (Simpson et al., 2017). Methylation calls for each sample preprocessed using NanoMethPhase *methyl_call_processor* module for downstream analysis (Akbari et al., 2021). To detect allelic methylation we first called variants using Strelka2 and default parameters from alignment files of each LCL and its parents obtained from 1KGP GRCh38 (Auton et al., 2015; Kim et al., 2018). For HG002 and HG003 variant call data were obtained from GIAB (Zook et al., 2019). For each LCL a mock phased vcf file with defined parent of origin of each high-quality heterozygous SNV was created using an in-house bash script (https://github.com/vahidAK/NanoMethPhase/tree/master/scripts: Trio_To_PhaseVCF_4FemaleChild.sh & Trio_To_PhaseVCF_4MaleChild.sh). Subsequently, we detected haplotype methylome in each sample using NanoMethPhase with the setting *nanomethphase phase –mbq 0*. Finally, DMRs between haplotypes were called using NanoMethPhase *dma* module. To avoid the confounding effects of X-inactivation, and because previous studies demonstrated no evidence of imprinting at sex chromosomes, we only examined autosomal chromosomes (Court et al., 2014; Joshi et al., 2016; Zink et al., 2018).

WGBS Data and Detection of Novel DMRs

To validate allelic methylation in other tissues and also detect potential novel ICRs we used 24 public WGBS (GSM1010978, GSM1010979, GSM1010980, GSM1010981, GSM1010983, GSM1010984, GSM1010986, GSM1010987, GSM1010988, GSM1010989, GSM1112838, GSM1120321, GSM1120326, GSM1127054, GSM1127125, GSM916049, GSM916050, GSM983645, GSM983647, GSM983648, GSM983649, GSM983650, GSM983651, GSM983652) for 20 tissue samples from Epigenomics Roadmap including adipose, adrenal gland, liver, aorta, brain hippocampus, breast luminal epithelial, breast myoepithelial, esophagus, gastric, left ventricle, lung, ovary, pancreas, psoas muscle, right atrium, right ventricle, sigmoid colon, small intestine, spleen, and thymus (Bernstein et al., 2010). Wig files which include fractional methylation data were obtained and converted to bed format using UCSC tools and lifted over to hg38 coordinates using CrossMap and UCSC lift over chain file (Kent et al., 2002; Zhao et al., 2014). All, bed format files were then merged to keep CpGs that are common in at least 10 samples. At imprinting control regions only one allele is methylated and we expect to observe partial methylation (~50%) at such regions. However, the adjacent sites which are not imprinted display ~0% or ~100% methylation. Therefore, we used a comparison between detected DMRs with their adjacent sites in WGBS data. For each DMR we determined the number of CpG sites with methylation rates between 30-70% (partial methylation) and normalized it by dividing the numbers to all CpGs in the interval. We also determined this ratio for the adjacent sites (>=20kb away and not been reported as imprinted gene or ICR). We then used limma's linear model to perform statistical analysis of the ratios at each DMR and adjacent sites (Codes are available on https://github.com/vahidAK/NanoMethPhase/tree/master/scripts: PartialCpGMethylationAtDMRandAdjacent.py and ComparePartialMethylationAtDMRsToAdjacentUsingLimma.R). As controls and because ICRs are usually overlapped with CpG islands, we examined 200 randomly selected CpG islands and 200 randomly selected 1kb, 2kb, and 3kb intervals with more than 15 CpGs.

Detection of Germline and Somatic DMRs

If a DMR is germline, it is established during germ cell development and survived the pre-implantation methylation reprograming. Therefore, gDMRs will overlap with DMR detected from oocyte vs sperm with consistent methylation direction, i.e. maternally methylated DMRs display high methylation in oocyte and very low or no methylation in sperm and vice versa. Moreover, gDMRs need to display partial methylation after fertilization and early development.

In order to discriminate gDMRs from somatic, we used public WGBS data for sperm, oocyte, blastocyst, and fetal tissues (GSM1172595 thymus, GSM1172596 muscle, GSM941747 brain) (Bernstein et al., 2010; Okae et al., 2014). Read counts for methylated and unmethylated CpG sites were obtained for sperm and oocyte samples and DMA was performed using NanoMethPhase *dma* module. To detect potential candidate gDMRs, we overlapped detected DMRs from oocyte vs sperm DMA to detected imprinted DMRs from nanopore. We further used blastocyst and fetal tissues to investigate if potential gDMRs escaped the second round of methylation reprograming and if sDMRs are stablished during somatic development.

Allelic H3K4me3 Analysis

H3K4me3 ChIP-seq fastq files were obtained for NA12878, NA12891, NA12892, NA19238, NA19239, NA19240, and NA18507 and aligned to the GRCh38 reference genome using bwa-mem default setting (Adoue et al., 2014; Kent et al., 2002; Li and Durbin, 2009). SNVs were called for these samples from 1KGP GRCh38 alignment files using strelka2 (Auton et al., 2015; Kim et al., 2018). We then counted the number of reads with minimum mapping quality of 20 and base quality of 10 at each heterozygous SNV and kept those with more than 5 mapped reads for binomial test. The reference allelic counts and total counts at each heterozygous SNV (or maternal allelic counts and total counts in case for trios) were used

448 to detect significant allelic bias using a two-sided binomial test under the default probability of $P = 0.5$ in
449 python SciPy package (Virtanen et al., 2020).

450 **Mammalian Conservation of DMRs**

451 To test whether any of the detected novel DMRs are conserved in other mammals we used WGBS data
452 for mouse (GSE42836), Macaque (GSE34128 and GSE151768), and Chimp (GSE151768) to examine
453 partial methylation in orthologous intervals (Hon et al., 2013; Tung et al., 2012). Mouse, Macaque, and
454 Chimp coordinates lifted over to mm10, RheMac8, and PanTro5 coordinates using CrossMap and
455 appropriate liftover file from UCSC, if they were not already in this coordinates. The list of detected
456 human DMRs were also converted to the orthologous regions for each mammal using CrossMap and the
457 appropriate UCSC liftover file (Kent et al., 2002; Zhao et al., 2014). Since many coordinates in human
458 splinted to several orthologous in other mammals, we merged orthologous intervals which were <=200bp
459 apart. Finally, we used our approach explained in aforementioned section (WGBS Data and Detection of
460 Novel DMRs) to detect ortologs with significant partial methylation.

461 To examine the somatic and germline orthologous DMRs, we used WGBS data from mouse embryo
462 (GSM3752614, GSM4558210) , sperm (GSE79226) , oocyte (GSM3681773, GSM3681774,
463 GSM3681775) and Rhesus embryo (GSM1466814), sperm (GSM1466810), and oocyte samples
464 (GSM1466811) (Dahlet et al., 2020; Gao et al., 2017; Jung et al., 2017; Saenz-de-Juano et al., 2019).

Acknowledgements: SJMJ and MM acknowledge funding from the Canada Research Chairs program and the Canadian Foundation for Innovation. VA acknowledge funding from the University of British Columbia Four Year Doctoral Fellowship**.**

Competing interests: The authors declare that there is no competing interests.

References

Adoue V, Schiavi A, Light N, Almlöf JC, Lundmark P, Ge B, Kwan T, Caron M, Rönnblom L, Wang C, Chen S-H, Goodall AH, Cambien F, Deloukas P, Ouwehand WH, Syvänen A-C, Pastinen T. 2014. Allelic expression mapping across cellular lineages to establish impact of non-coding SNPs. *Mol Syst Biol* **10**:754.

Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* **12**:1–14.

Akbari V, Garant J-M, O'Neill K, Pandoh P, Moore R, Marra MA, Hirst M, Jones SJM. 2021. Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase. *Genome Biol* **22**:68.

Amemiya HM, Kundaje A, Boyle AP. 2019. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**:9354.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, Lee S, Muzny D, Reid JG, Zhu Y, Wang J, Chang Y, Feng Q, Fang X, Guo X, Jian M, Jiang H, Jin X, Lan T, Li G, Li J, Li Y, Liu S, Liu X, Lu Y, Ma X, Tang M, Wang B, Wang G, Wu H, Wu R, Xu X, Yin Y, Zhang D, Zhang W, Zhao J, Zhao M, Zheng X, Lander ES, Altshuler DM, Gabriel SB, Gupta N, Gharani N, Toji LH, Gerry NP, Resch AM, Flicek P, Barker J, Clarke L, Gil L, Hunt SE, Kelman G, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R, Roa A, Smirnov D, Smith RE, Streeter I, Thormann A, Toneva I, Vaughan B, Zheng-Bradley X, Bentley DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach H, Sudbrak R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann B, Yaspo M-L, Mardis ER, Wilson RK, Fulton L, Fulton R, Sherry ST, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R, Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P, O'Sullivan C, Ostapchuk Y, Phan L, Ponomarov S, Schneider V, Shekhtman E, Sirotkin K, Slotta D, Zhang H, McVean GA, Durbin RM, Balasubramaniam S, Burton J, Danecek P, Keane TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Schmidt JP, Davies CJ, Gollub J, Webster T, Wong B, Zhan Y, Auton A, Campbell CL, Kong Y, Marcketta A, Gibbs RA, Yu F, Antunes L, Bainbridge M, Muzny D, Sabo A, Huang Z, Wang J, Coin LJM, Fang L, Guo X, Jin X, Li G, Li Q, Li Y, Li Z, Lin H, Liu B, Luo R, Shao H, Xie Y, Ye C, Yu C, Zhang F, Zheng H, Zhu H, Alkan C, Dal E, Kahveci F, Marth GT, Garrison EP, Kural D, Lee W-P, Fung Leong W, Stromberg M, Ward AN, Wu J, Zhang M, Daly MJ, DePristo MA, Handsaker RE, Altshuler DM, Banks E, Bhatia G, del Angel G, Gabriel SB, Genovese G, Gupta N, Li H, Kashin S, Lander ES, McCarroll SA, Nemesh JC, Poplin RE, Yoon SC, Lihm J, Makarov V, Clark AG, Gottipati S, Keinan A, Rodriguez-Flores JL, Korbel JO, Rausch T, Fritz MH, Stütz AM, Flicek P, Beal K, Clarke L, Datta A, Herrero J, McLaren WM, Ritchie GRS, Smith RE, Zerbino D, Zheng-Bradley X, Sabeti PC, Shlyakhter I, Schaffner SF, Vitti J, Cooper DN, Ball E V, Stenson PD, Bentley DR, Barnes B, Bauer M, Keira Cheetham R, Cox A, Eberle M, Humphray S, Kahn S, Murray L, Peden J, Shaw R, Kenny EE, Batzer MA, Konkel MK, Walker JA, MacArthur DG, Lek M, Sudbrak R, Amstislavskiy VS, Herwig R, Mardis ER, Ding L, Koboldt DC, Larson D, Ye K, Gravel S, Consortium T 1000 GP, authors C, committee S, group P,

Medicine BC of, BGI-Shenzhen, Harvard BI of MIT and, Research CI for M, European Molecular Biology Laboratory EBI, Illumina, Genetics MPI for M, University MGI at W, Health USNI of, Oxford U of, Institute WTS, group A, Affymetrix, Medicine AEC of, University B, College B, Laboratory CSH, University C, Laboratory EMB, University H, Database HGM, Sinai IS of M at M, University LS, Hospital MG, University M, National Eye Institute NIH. 2015. A global reference for human genetic variation. *Nature* **526**:68–74.

Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, Kukurba KR, Zhang R, Li JB, van der Kooy D. 2015. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat Genet* **47**:544–549.

Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, Church GM. 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**:361–368.

Baran Y, Subramaniam M, Biton A, Tukiainen T, Tsang EK, Rivas MA, Pirinen M, Gutierrez-Arcelus M, Smith KS, Kukurba KR. 2015. The landscape of genomic imprinting across diverse adult human tissues. *Genome Res* **25**:927–936.

Bartolomei MS, Ferguson-Smith AC. 2011. Mammalian Genomic Imprinting. *Cold Spring Harb Perspect Biol* **3**.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR, Farnham PJ, Hirst M, Lander ES, Mikkelsen TS, Thomson JA. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**:1045–1048.

Camargo M, Rivera D, Moreno L, Lidral AC, Harper U, Jones M, Solomon BD, Roessler E, Vélez JI, Martinez AF, Chandrasekharappa SC, Arcos-Burgos M. 2012. GWAS reveals new recessive loci associated with non-syndromic facial clefting. *Eur J Med Genet* **55**:510–514.

Chen Z, Zhang Y. 2020. Maternal H3K27me3-dependent autosomal and X chromosome imprinting. *Nat Rev Genet* **21**:555–571.

Cheong CY, Chng K, Ng S, Chew SB, Chan L, Ferguson-Smith AC. 2015. Germline and somatic imprinting in the nonhuman primate highlights species differences in oocyte methylation. *Genome Res* **25**:611–623.

Chess A. 2013. Random and non-random monoallelic expression. *Neuropsychopharmacology* **38**:55–61.

Chiesa N, De Crescenzo A, Mishra K, Perone L, Carella M, Palumbo O, Mussa A, Sparago A, Cerrato F, Russo S, Lapi E, Cubellis MV, Kanduri C, Cirillo Silengo M, Riccio A, Ferrero GB. 2012. The KCNQ1OT1 imprinting control region and non-coding RNA: new properties derived from the study of Beckwith-Wiedemann syndrome and Silver-Russell syndrome cases. *Hum Mol Genet* **21**:10–25.

Court F, Tayama C, Romanelli V, Martin-Trujillo A, Iglesias-Platas I, Okamura K, Sugahara N, Simón C, Moore H, Harness J V., Keirstead H, Sanchez-Mut JV, Kaneki E, Lapunzina P, Soejima H, Wake N, Esteller M, Ogata T, Hata K, Nakabayashi K, Monk D. 2014. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Res* **24**:554–569.

da Rocha ST, Gendrel A-V. 2019. The influence of DNA methylation on monoallelic expression. *Essays Biochem* **63**:663–676.

Dahlet T, Argüeso Lleida A, Al Adhami H, Dumas M, Bender A, Ngondo RP, Tanguy M, Vallet J, Auclair G, Bardet AF, Weber M. 2020. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nat Commun* **11**:3153.

De Coster W, De Rijk P, De Roeck A, De Pooter T, D'Hert S, Strazisar M, Sleegers K, Van Broeckhoven C. 2019. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res* **29**:1178–1187.

Gao F, Niu Y, Sun YE, Lu H, Chen Y, Li S, Kang Y, Luo Y, Si C, Yu J, Li C, Sun N, Si W, Wang H, Ji W, Tan T. 2017. De novo DNA methylation during monkey pre-implantation embryogenesis. *Cell Res* **27**:526–539.

Geneimprint. 2021. http://www.geneimprint.com

Gigante S, Gouil Q, Lucattini A, Keniry A, Beck T, Tinning M, Gordon L, Woodruff C, Speed TP, Blewitt ME, Ritchie ME. 2019. Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Res* **47**:e46–e46.

Goovaerts T, Steyaert S, Vandenbussche CA, Galle J, Thas O, Van Criekinge W, De Meyer T. 2018. A comprehensive overview of genomic imprinting in breast and its deregulation in cancer. *Nat Commun* **9**:1–14.

Hanna CW, Kelsey G. 2014. The specification of imprints in mammals. *Heredity (Edinb)* **113**:176–183.

Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X. 2017. Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol Cell* **66**:711-720.e3.

Hernandez Mora JR, Tayama C, Sánchez-Delgado M, Monteagudo-Sánchez A, Hata K, Ogata T, Medrano J, Poo-Llanillo ME, Simón C, Moran S, Esteller M, Tenorio J, Lapunzina P, Kagami M, Monk D, Nakabayashi K. 2018. Characterization of parent-of-origin methylation using the Illumina Infinium MethylationEPIC array platform. *Epigenomics* **10**:941–954.

Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B. 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* **45**:1198–1206.

Jadhav B, Monajemi R, Gagalova KK, Ho D, Draisma HHM, van de Wiel MA, Franke L, Heijmans BT, van Meurs J, Jansen R, Consortium G, Consortium B, 't Hoen PAC, Sharp AJ, Kiełbasa SM. 2019. RNA-Seq in 296 phased trios provides a high-resolution map of genomic imprinting. *BMC Biol* **17**:50.

Jelinic P, Shaw P. 2007. Loss of imprinting and cancer. *J Pathol* **211**:261–268.

John RM, Lefebvre L. 2011. Developmental regulation of somatic imprints. *Differentiation* **81**:270–280.

Joshi RS, Garg P, Zaitlen N, Lappalainen T, Watson CT, Azam N, Ho D, Li X, Antonarakis SE, Brunner HG, Buiting K, Cheung SW, Coffee B, Eggermann T, Francis D, Geraedts JP, Gimelli G, Jacobson SG, Le Caignec C, de Leeuw N, Liehr T, Mackay DJ, Montgomery SB, Pagnamenta AT, Papenhausen P, Robinson DO, Ruivenkamp C, Schwartz C, Steiner B, Stevenson DA, Surti U, Wassink T, Sharp AJ. 2016. DNA Methylation Profiling of Uniparental Disomy Subjects Provides a Map of Parental Epigenetic Bias in the Human Genome. *Am J Hum Genet* **99**:555–566.

Jung YH, Sauria MEG, Lyu X, Cheema MS, Ausio J, Taylor J, Corces VG. 2017. Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* **18**:1366–

1382.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Res* **12**:996–1006.

Khamlichi AA, Feil R. 2018. Parallels between mammalian mechanisms of monoallelic gene expression. *Trends Genet* **34**:954–971.

Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, Chen X, Kim Y, Beyter D, Krusche P, Saunders CT. 2018. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**:591–594.

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, Wei C-L. 2010. Dynamic changes in the human methylome during differentiation. *Genome Res* **20**:320–331.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**:3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**:1754–1760.

Matsubara K, Itoh M, Shimizu K, Saito S, Enomoto K, Nakabayashi K, Hata K, Kurosawa K, Ogata T, Fukami M, Kagami M. 2019. Exploring the unique function of imprinting control centers in the PWS/AS-responsible region: finding from array-based methylation analysis in cases with variously sized microdeletions. *Clin Epigenetics* **11**:36.

Maupetit-Méhouas S, Montibus B, Nury D, Tayama C, Wassef M, Kota SK, Fogli A, Cerqueira Campos F, Hata K, Feil R, Margueron R, Nakabayashi K, Court F, Arnaud P. 2016. Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. *Nucleic Acids Res* **44**:621–635.

Moore LD, Le T, Fan G. 2013. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **38**:23–38.

Morison IM, Paton CJ, Cleverley SD. 2001. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* **29**:275–276.

Okae H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, Kikuchi H, Yoshida H, Tanaka A, Suyama M, Arima T. 2014. Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet* **10**:e1004868–e1004868.

Puerta MF, Ruth C, Fang L, Robert P, Zhong D, A. R-MY, S. BM, M. LP, Qiyi T, L. H-F. 2014. CTCF Binding to the First Intron of the Major Immediate Early (MIE) Gene of Human Cytomegalovirus (HCMV) Negatively Regulates MIE Gene Expression and HCMV Replication. *J Virol* **88**:7389–7401.

Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone P V. 2007. Critical DNA Binding Interactions of the Insulator Protein CTCF: A SMALL NUMBER OF ZINC FINGERS MEDIATE STRONG BINDING, AND A SINGLE FINGER-DNA INTERACTION CONTROLS BINDING AT IMPRINTED LOCI *. *J Biol Chem* **282**:33336–33345.

Romanelli V, Nakabayashi K, Vizoso M, Moran S, Iglesias-Platas I, Sugahara N, Simón C, Hata K, Esteller M, Court F, Monk D. 2014. Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in

cancer. *Epigenetics* **9**:783–790.

Saenz-de-Juano MD, Ivanova E, Billooye K, Herta A-C, Smitz J, Kelsey G, Anckaert E. 2019. Genome-wide assessment of DNA methylation in mouse oocytes reveals effects associated with in vitro growth, superovulation, and sexual maturity. *Clin Epigenetics* **11**:197.

Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, Armstrong J, Tigyi K, Maurer N, Koren S, Sedlazeck FJ, Marschall T, Mayes S, Costa V, Zook JM, Liu KJ, Kilburn D, Sorensen M, Munson KM, Vollger MR, Monlong J, Garrison E, Eichler EE, Salama S, Haussler D, Green RE, Akeson M, Phillippy A, Miga KH, Carnevali P, Jain M, Paten B. 2020. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol*.

Sharp AJ, Migliavacca E, Dupre Y, Stathaki E, Sailani MR, Baumer A, Schinzel A, Mackay DJ, Robinson DO, Cobellis G, Cobellis L, Brunner HG, Steiner B, Antonarakis SE. 2010. Methylation profiling in individuals with uniparental disomy identifies novel differentially methylated regions on chromosome 15. *Genome Res* **20**:1271–1278.

Silver MJ, Kessler NJ, Hennig BJ, Dominguez-Salas P, Laritsky E, Baker MS, Coarfa C, Hernandez-Vargas H, Castelino JM, Routledge MN, Gong YY, Herceg Z, Lee YS, Lee K, Moore SE, Fulford AJ, Prentice AM, Waterland RA. 2015. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome Biol* **16**:118.

Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* **14**:407.

Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**:204–220.

Steijger T, Abril JF, Engström PG, Kokocinski F, Hubbard TJ, Guigó R, Harrow J, Bertone P. 2013. Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods* **10**:1177–1184.

Tomizawa S, Sasaki H. 2012. Genomic imprinting and its relevance to congenital disease, infertility, molar pregnancy and induced pluripotent stem cell. *J Hum Genet* **57**:84–91.

Tung J, Barreiro LB, Johnson ZP, Hansen KD, Michopoulos V, Toufexis D, Michelini K, Wilson ME, Gilad Y. 2012. Social environment is associated with gene regulatory variation in the rhesus macaque immune system. *Proc Natl Acad Sci U S A* **109**:6490–6495.

Turan S, Bastepe M. 2013. The GNAS complex locus and human diseases associated with loss-of-function mutations or epimutations within this imprinted gene. *Horm Res Paediatr* **80**:229–241.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli A Pietro, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik D V, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold G-L, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N,

Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y, Contributors S 1. . 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**:261–272.

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B. 2012. Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. *Cell* **148**:816–831.

Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26**:577–590.

Zhao H, Sun Z, Wang J, Huang H, Kocher J-P, Wang L. 2014. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**:1006–1007.

Zink F, Magnusdottir DN, Magnusson OT, Walker NJ, Morris TJ, Sigurdsson A, Halldorsson GH, Gudjonsson SA, Melsted P, Ingimundardottir H, Kristmundsdottir S, Alexandersson KF, Helgadottir A, Gudmundsson J, Rafnar T, Jonsdottir I, Holm H, Eyjolfsson GI, Sigurdardottir O, Olafsson I, Masson G, Gudbjartsson DF, Thorsteinsdottir U, Halldorsson B V, Stacey SN, Stefansson K. 2018. Insights into imprinting from parent-of-origin phased methylomes and transcriptomes. *Nat Genet* **50**:1542–1552.

Zook JM, Catoe D, McDaniel J, Vang L, Spies N, Sidow A, Weng Z, Liu Y, Mason CE, Alexander N, Henaff E, McIntyre ABR, Chandramohan D, Chen F, Jaeger E, Moshrefi A, Pham K, Stedman W, Liang T, Saghbini M, Dzakula Z, Hastie A, Cao H, Deikus G, Schadt E, Sebra R, Bashir A, Truty RM, Chang CC, Gulbahce N, Zhao K, Ghosh S, Hyland F, Fu Y, Chaisson M, Xiao C, Trow J, Sherry ST, Zaranek AW, Ball M, Bobe J, Estep P, Church GM, Marks P, Kyriazopoulou-Panagiotopoulou S, Zheng GXY, Schnall-Levin M, Ordonez HS, Mudivarti PA, Giorda K, Sheng Y, Rypdal KB, Salit M. 2016. Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci Data* **3**:160025.

Zook JM, McDaniel J, Olson ND, Wagner J, Parikh H, Heaton H, Irvine SA, Trigg L, Truty R, McLean CY, De La Vega FM, Xiao C, Sherry S, Salit M. 2019. An open resource for accurately benchmarking small variant and reference calls. *Nat Biotechnol* **37**:561–566.