

RESEARCH ARTICLE

Open Access

# Discovering candidate imprinted genes and imprinting control regions in the human genome



Minou Bina

## Abstract

**Background:** Genomic imprinting is a process whereby a subset of genes is expressed in a parent-of-origin specific manner. This evolutionary novelty is restricted to mammals and controlled by genomic DNA segments known as Imprinting Control Regions (ICRs) and germline Differentially Methylated Regions (gDMRs). Previously, I showed that in the mouse genome, the fully characterized ICRs/gDMRs often includes clusters of 2 or more of a set of composite-DNA-elements known as ZFBS-morph overlaps.

**Results:** Because of the importance of the ICRs to regulating parent-of-origin specific gene expression, I developed a genome-wide strategy for predicting their positions in the human genome. My strategy consists of creating plots to display the density of ZFBS-morph overlaps along the entire chromosomal DNA sequences. In initial evaluations, I found that peaks in these plots pinpointed several of the known ICRs/gDMRs along the DNA in chromosomal bands. I deduced that in density-plots, robust peaks corresponded to actual or candidate ICRs in the DNA. By locating the genes in the vicinity of candidate ICRs, I could discover potential imprinting genes. Additionally, my assessments revealed a connection between several of the potential imprinted genes and human developmental anomalies. Examples include Leber congenital amaurosis 11, Coffin-Siris syndrome, progressive myoclonic epilepsy-10, microcephalic osteodysplastic primordial dwarfism type II, and microphthalmia, cleft lip and palate, and agenesis of the corpus callosum.

**Conclusion:** With plots displaying the density of ZFBS-morph overlaps, researchers could locate candidate ICRs and imprinted genes. Since the datafiles are available for download and display at the UCSC genome browser, it is possible to examine the plots in the context of Single nucleotide polymorphisms (SNPs) to design experiments to discover novel ICRs and imprinted genes in the human genome.

**Keywords:** Gene regulation, Genetics, Developmental, Genomic imprinting, Mammals, MLL, Syndrome, ZFP57

## Background

Imprinted genes play key roles in fetal development and in postnatal processes including behavior, sleep, feeding, maintenance of body temperature, metabolic regulation, and stem cell maintenance and renewal [1]. The imprinting mechanism is relatively complex and involves

orchestrated actions of several enzymes and proteins [2, 3]. In that process, key players include ZFP57 and a complex consisting of DNMT3A and DNMT3L. This complex methylates DNA processively on a variety of CpG-rich substrates including the promoters of human genes encompassed by CpG islands [4]. Because of the importance of the ICRs to parent-of-origin specific gene expression, it is necessary to develop strategies for their localization in mammalian genomic DNA. Towards that goal, a previous study discovered that ZFP57 interacted

Correspondence: [bina@purdue.edu](mailto:bina@purdue.edu)  
Department of Chemistry, Purdue University, 560 Oval Dr., West Lafayette, IN 47907, USA



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

with a double-stranded, CpG-methylated, hexanucleotide (TGCCGC). This interaction was essential to the recognition of ICRs by ZFP57 to maintain allele-specific gene expression [5]. Since CpG containing sequences are infrequent in animal DNA [6], the instances of finding ZFBS in mammalian DNA is by far less than those observed for AT-rich hexanucleotides. Nonetheless, TGCCGC is relatively short. Therefore, it occurs often in genomic DNA [7]. Consequently, random occurrences of the hexameric site would obscure detection of the functional ZFP57 sites within the ICRs/gDMRs dispersed along relatively long genomic DNA sections.

Previously, I addressed that problem by extending the length of the canonical ZFP57 binding site to include additional nucleotides [7]. This strategy eliminated a significant fraction of randomly occurring ZFP57 sites in mouse genomic DNA and led to the discovery of the ZFBS-Morph overlaps [7, 8]. These overlaps define composite-DNA-elements consisting of the hexameric ZFP57 binding site overlapping a subset of the MLL1 morphemes [7, 8]. These morphemes correspond to the smallest ‘words’ in DNA that selectively bind the MT-domain in MLL1 [9]. The MT domain is present in both MLL1 and MLL2. In DNA binding assays, this domain interacted selectively with nonmethylated CpG-rich sequences [10, 11]. Thus, within the ICRs, the ZFBS-Morph overlaps may play a dual and antagonistic role. Binding of MLL1 and MLL2 to ICRs protecting the DNA from methylation. Binding of ZFP57 to the modified DNA maintaining allele-specific expression [5, 7, 10–12].

Since closely-spaced ZFBS-Morph overlaps impart contextual specificity to ICRs, their localization could help with pinpointing the genomic positions of the ICRs that are currently unknown. Towards this goal, I describe a Bioinformatics strategy. This strategy involved creating plots to view the positions of clusters of 2 or more ZFBS-Morph overlaps along chromosomal DNA. In these plots, peaks appear within a sliding window consisting of 850-bases. By displaying the plots at the UCSC genome browser, I could locate the peaks corresponding to several of the well-known ICRs/gDMRs. I also could examine the human genes in the context of short clinical variants and results of genome wide association studies. Here, I give an overview of how with density-plots, I could discover potential imprinted genes. I found that several of these genes were associated with disease-states and developmental anomalies known as syndromes. Examples include association of *IMPDH1* with Leber congenital amaurosis 11, *ARID1B* with Coffin-Siris syndrome, *PRDM8* with progressive myoclonic epilepsy-10, *PCNT* with microcephalic osteodysplastic primordial dwarfism type II, *CITED2* with ventricular septal defect 2, and *VAX1* with

microphthalmia, cleft lip and palate, and agenesis of the corpus callosum. Among potential imprinted genes, of key importance could be those that impact gene regulation. Examples include genes for transcription factors and genes that affect chromatin structure. Notably, there are many examples of syndromes that arise from mutations in transcription factor genes [13]. In that context, potential imprinted genes, discovered by my approach, include *CITED2*, *ZBTB2*, and *VAX1*.

## Results

To create density-plots, I used a Perl script that counted and reported the number of ZFBS-morph overlaps along a sliding window consisting of 850 bases. The script ignored isolated occurrences in the window and hence removed background noise. I selected the window size by trial and error as described in [14]. Large windows tended to produce false peaks. Small windows tended to give peaks with spikey appearance. In preliminary assessments, I found that the density peaks covering 2 ZFBS-Morph overlaps could be false or true-positive. Therefore, to locate candidate ICRs, I primarily examined the ‘robust’ peaks encompassing 3 or more ZFBS-Morph overlaps. In evaluations, I asked whether I could locate the known ICRs/gDMRs within relatively long genomic DNA sections. Concurrently, I inspected the DNA for peaks that may reflect the genomic positions of candidate ICRs. By displaying density-plots at the UCSC browser, I could obtain enlarged views to investigate the positions of the candidate ICRs with respect to genes, transcripts, and the CpG islands (CGIs). Additional assessments included determining whether the predicted imprinted genes corresponded to candidate imprinted genes discovered by studying allele-specific expression in tissues obtained from human term placenta [15]. I also inspected several of the density peaks in the context of their positions with respect to short clinical variants producing disease-states or developmental anomalies.

### Density-plots predicted ICRs for parent-of-origin specific expression of several experimentally identified candidate imprinted genes

To investigate the robustness of my approach, initially I assessed the positions of density peaks with respect to candidate imprinted genes listed in a previous report [15]. That list was produced from high-throughput examinations of allele-specific expression in tissues from human term placenta [15]. In my studies, I examined whether any of the listed genes corresponded to potential imprinted genes predicted by my approach. Among the candidates, the UCSC genome browser did not find *MGC16597* and *MGC24665*. Among the remaining candidates, density-plots revealed peaks within or in the

vicinity of *PRDM8*, *SQSTM1*, *NM\_006031*, *TJP2*, *CDK2AP1*, *MYH7B*, and *MAN2C1* (Table 1).

In the *PRDM8* locus, I observed a robust intragenic density peak. This peak maps to a CGI in the last exon of several *PRDM8* transcripts (Fig. S1). This finding agrees with the correspondence of *PRDM8* to an actual imprinted gene. Next, I checked the positions of peaks in a DNA segment that included *SQSTM1*. This segment contains two overlapping genes (*MGAT4B* and *SQSTM1*). A relatively long CpG island encompassed the TSSs of three *SQSTM1* transcripts and the longest *MGAT4B* transcript (Fig. 1, Table 1, and Table 2). In density-plots, I observed an intragenic density peak within that island. Thus, my strategy revealed a candidate ICR regulating allele-specific expression of *SQSTM1*. Furthermore, while lending support for the correspondence of *SQSTM1* to a genuine imprinted gene, density-plots also located a potential imprinted transcript produced from *MGAT4B* (Fig. 1, Table 2). For the gene listed as *NM\_006031* in reference [15], the UCSC genome browser displayed the actual gene (*PCNT*). Within that locus, density-plots revealed two very robust peaks supporting that *PCNT* could be a genuine imprinted gene (Fig. S2). For three loci, I observed peaks covering 2 ZFBS-Morph overlaps. Therefore the corresponding candidate ICRs could be true or false positive [16].

For *FLJI0300*, the genome browser showed a sequence (*AK001162*) mapping to *WDR60*. According to OMIM, this gene is associated with Short-rib thoracic dysplasia 8 (SRTD8) with or without polydactyly [17]. Within *WDR60*, I observed a density peak covering 2 ZFBS-Morph overlaps (Fig. S3). However, in plots I noticed a very robust density peak mapping to the 5' end of *ESYT2* (Fig. S3). This peak also could be a candidate ICR for regulating parent-of-origin expression from *WDR60*. Similarly, for several of the listed candidate

imprinted genes [15], I noticed density peaks far upstream or downstream of their TSSs. Examples include *XRRRA1*, *CD151*, and *VPS11*. Thus, overall, density peaks predicted ICRs for several candidate imprinted genes discovered by experimental and computational strategies. Furthermore, my results revealed that several of these genes corresponded to potential imprinted genes predicted by my approach.

### In density-plots, robust peaks pinpointed several of the known gDMRs/ICRs and revealed candidate ICRs and imprinted genes

To further assess the validity of my strategy, initially I inspected the density-plot obtained for Chr6. This chromosome includes the known intragenic ICR in the *PLAGL1* locus [18, 19]. This ICR is ~ 1 kb and maps to a CpG Island (CpG118) that encompasses the TSSs of a noncoding RNA gene (*HYMAI*) and the *PLAGL1* transcript (*ZAC1*) expressed from one of two parental alleles [18, 20]. In closeup views of the locus, clearly apparent is a single robust intragenic density peak in CpG118. Thus, this peak correctly located the ICR that regulates parent-of-origin specific transcription of *ZAC1* and *HYMAI* (Fig. 2).

Next, I inspected the density-plot obtained for Chr11 to assess the positions of peaks with respect to two of the well-known imprinted domains. In Chr11, the telomeric imprinted domain 1 includes a noncoding RNA gene (*H19*) and genes that encode insulin-like growth factor 2 (*IGF2*) and insulin (*INS*). The imprinted domain 2 is relatively long and encompasses three noncoding RNA genes (*KCNQ1OT1*, *KCNQ1-AS1* and *KCNQ1DN*) and ~ 10 protein coding genes including *KCNQ1*, *CDKN1C*, *SLC22A18*, *PHLDA2*, and *OSBPL5* [21]. In initial evaluations of plots, I inspected several long DNA sections encompassing both the *H19* – *IGF2* and the *KCNQ1* imprinted domains. In a 1.4 Mb DNA, I observed 3 robust density peaks and several peaks covering 2 ZFBS-morph overlaps (Fig. 3). One of the robust peaks appeared as a doublet and mapped to the imprinted domain 1. The other two robust peaks are in imprinted domain 2 (Fig. 3).

In domain 1, a single gDMR/ICR regulates transcription of *H19* selectively from the maternal allele and expression of *IGF2* and *INS* from the paternal allele [22]. This ICR is upstream of *H19* TSS and often is described in the context of several unique repeats [23] and sites predicted to bind the transcription factor CTCF [24]. Previously, I performed detailed analyses to determine the accuracy of the mapped repeats and the predicted CTCF sites [25, 26]. Mostly, results of my analyses agreed with the reported repeat positions [23]. However, results of the ENCODE data did not support the existence of the predicted CTCF site 5. Additionally, based

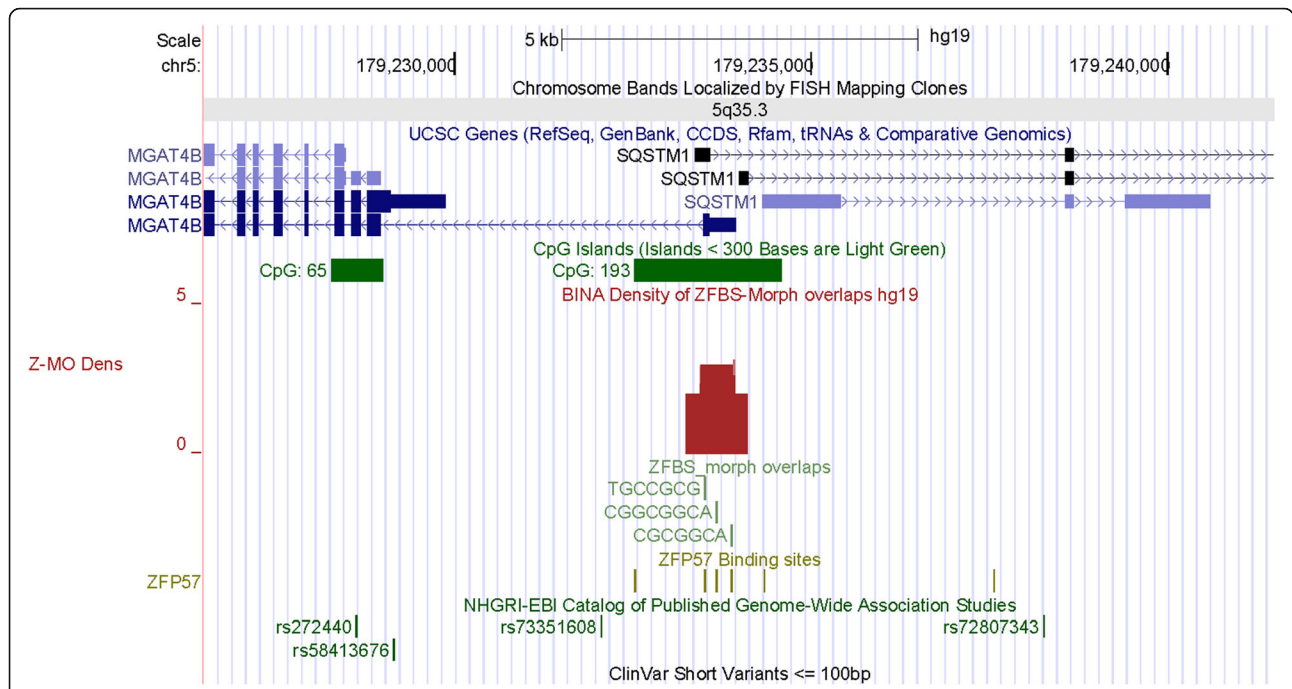
**Table 1** Predicted ICRs for candidate imprinted gene identified by experimental techniques [15]

chr4:81,122,001-81,124,000	<i>PRDM8</i>
chr5:179,233,001-179,234,500	<i>SQSTM1</i>
chr7:158,713,001-158,714,500 #	<i>WDR60</i>
chr7:158,621,001-158,622,500 &	
chr9:71,736,276-71,737,400 #	<i>TJP2</i>
chr12:123,755,871-123,756,470 #	<i>CDK2AP1</i>
chr15:75,660,396-75,661,095 #	<i>MAN2C1</i>
chr20:33,575,636-33,576,535 #	<i>MYH7B</i>
chr21:47,823,001-47,824,200	<i>PCNT</i> *
chr21:47,832,551-47,833,700	

# This density peak encompasses 2 ZFBS-morph overlaps. Therefore, it could be a true or a false positive

& This position could be a better candidate ICR For *WDR60*

\* This locus includes two robust density peaks



**Fig. 1** A candidate ICR mapping to overlapping transcripts (*MGAT4B* and *SQSTM1*). High-throughput experimental analysis has identified *SQSTM1* as a candidate imprinted gene [15]. The density-plots include a peak corresponding to a candidate ICR for imprinted expression of *SQSTM1*. Furthermore, the position of this ICR predicts parent-of-origin specific expression for the longest transcript produced from *MGAT4B*

on the ENCODE data, I discovered a new CTCF site that mapped to a chromatin boundary consisting of CTCF, RAD21, and SMC3 –for details see reference [25]. The new CTCF site (numbered 8) is in a CGI upstream of *H19* TSS. Furthermore, in closeup view of density-plots, I noticed two closely-spaced density peaks locating correctly the ICR in the imprinted domain 1 (Fig. S4).

Domain 2 is relatively long and includes the ICR (KvDMR1) that regulates expression of several imprinted genes [21, 27]. This ICR is intragenic and is located in

**Table 2** Genomic positions of candidate ICRs and imprinted genes deduced in this report

chr5:179,233,001-179,234,500	<i>MGAT4B</i>
chr6:151,711,301-151,713,300	<i>ZBTB2</i>
chr6:143,832,201-143,833,400	<i>FUCA2</i>
chr6:139,693,901-139,695,200	<i>CITED2</i>
chr6:157,098,501-157,101,000	<i>ARID1B</i> * and <i>LOC115308161</i>
chr6:150,584,501-150,586,000	<i>PPP1R14C</i> and possibly <i>IYD</i>
chr6:147,829,201-147,830,300	<i>SAMD5</i>
chr7:158,621,001-158,622,500	<i>ESYT2</i>
chr7:130,418,401-130,419,500	<i>KLF14</i> <sup>§</sup>
chr7:128,045,491-128,046,440	<i>IMPDH1</i>
chr10:118,896,501-118,897,800	<i>VAX1</i>

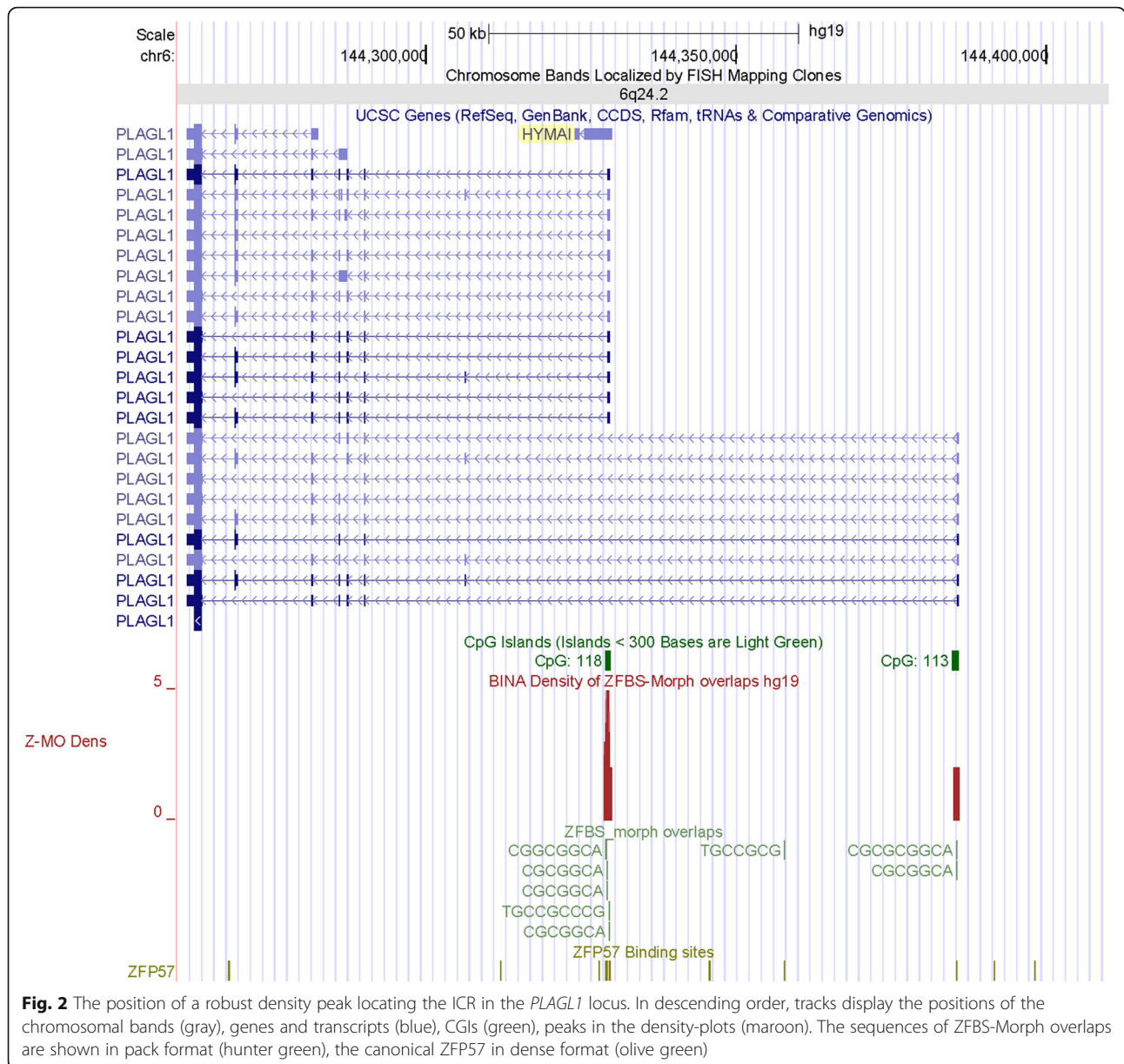
\* A known imprinted gene in mouse [29]

§ a known human imprinted gene [30]

the vicinity of *KCNQ1OT1* TSS. In closeup views, I observed a robust density peak within the CpG island that encompasses the *KCNQ1OT1* TSS of and the KvDMR1 (Fig. S5). Another robust density peak is between two noncoding RNA genes corresponding to *KCNQ1DN* and *KCNQ1-AS1* (Fig. 3). *KCNQ1DN* is an imprinted transcript within the WT2 critical region [28]. *KCNQ1-AS1* was reported in the genome assembly NR\_130721.1. Even though the function of *KCNQ1-AS1* is unknown, it is worth noting that it is transcribed antisense with respect to *KCNQ1OT1*. Therefore, the expression of *KCNQ1-AS1* might be a mechanism to impede leaky production of *KCNQ1OT1* in a subset of human tissues.

**Density-plots offer a strategy to discover potentially novel imprinted genes, candidate ICRs, and to determine their possible associations with clinical abnormalities**

In humans, defects in imprinted genes could cause severe diseases and developmental anomalies [1]. Therefore, I explored whether density-plots could help with finding unknown imprinting genes and whether any of these genes could be associated with genetic disorders. In exploratory studies, I also inspected peak resolutions in plots obtained for entire chromosomal DNA sequences. For example, examine the plot displaying the position of peaks in Chr6 (Fig. 4). The density peaks and

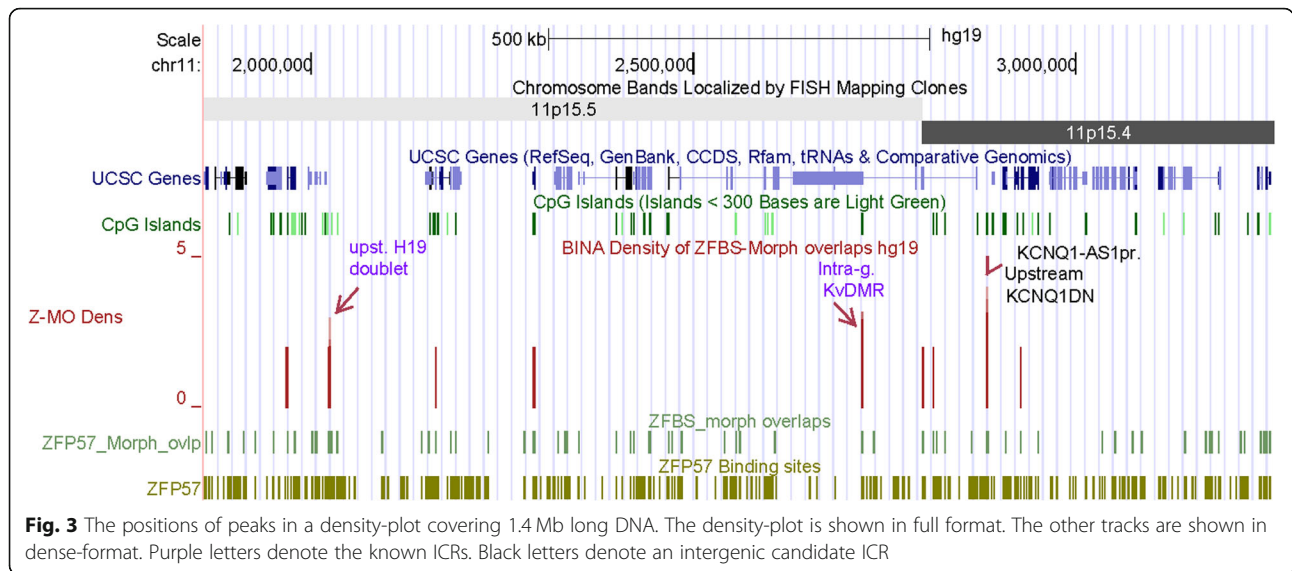


short clinical variants are primarily in gene rich regions. I noticed that Chr6 includes several robust peaks. Notably, one of the discernable peaks corresponds to the ICR of *ZAC1* and *HYMAI* in the *PLAGL1* locus (Fig. 4).

**A relatively long DNA sections from Chr6q includes candidate ICRs for several potential imprinted genes**

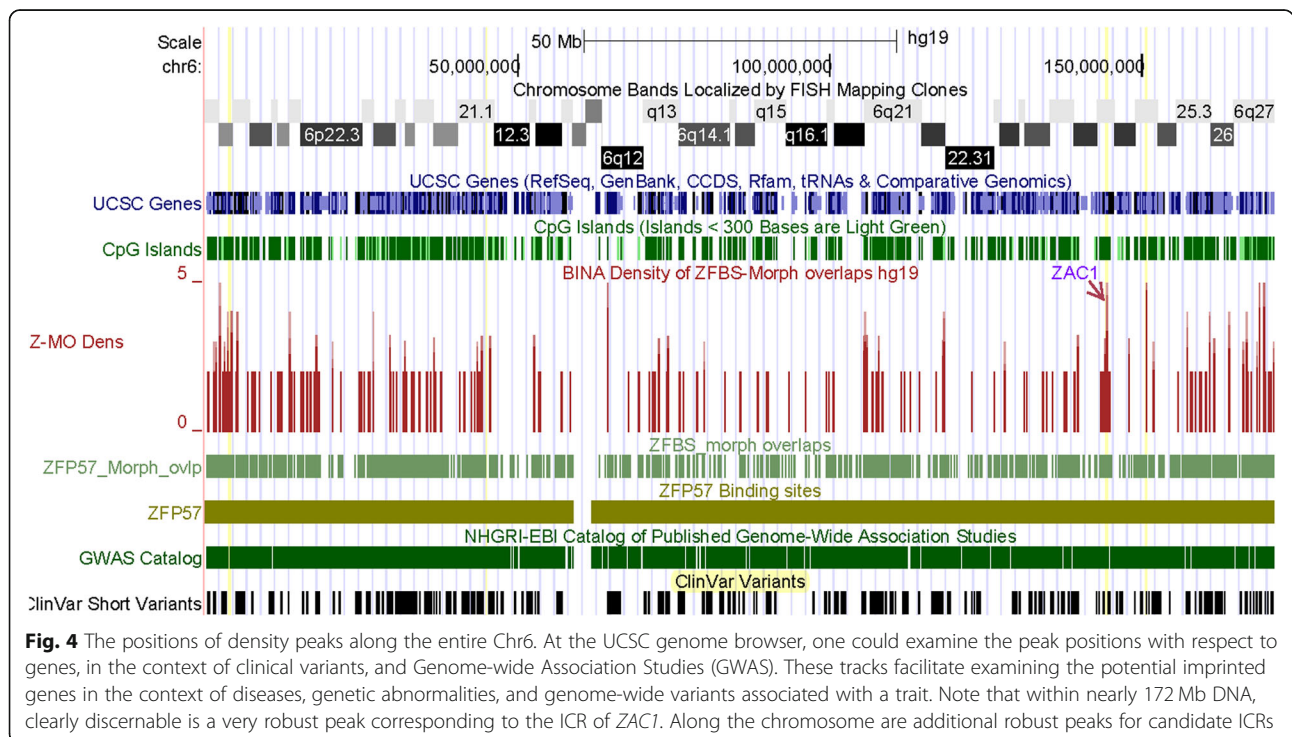
For close-inspections of plots with respect to clinical variants, I selected several relatively long DNA sections that included a known ICR and candidate ICRs for potential imprinted genes. Initially I zoomed in a segment in Chr6 that included the known ICR of *ZAC1*. The displayed segment contains 25 Mb DNA encompassing

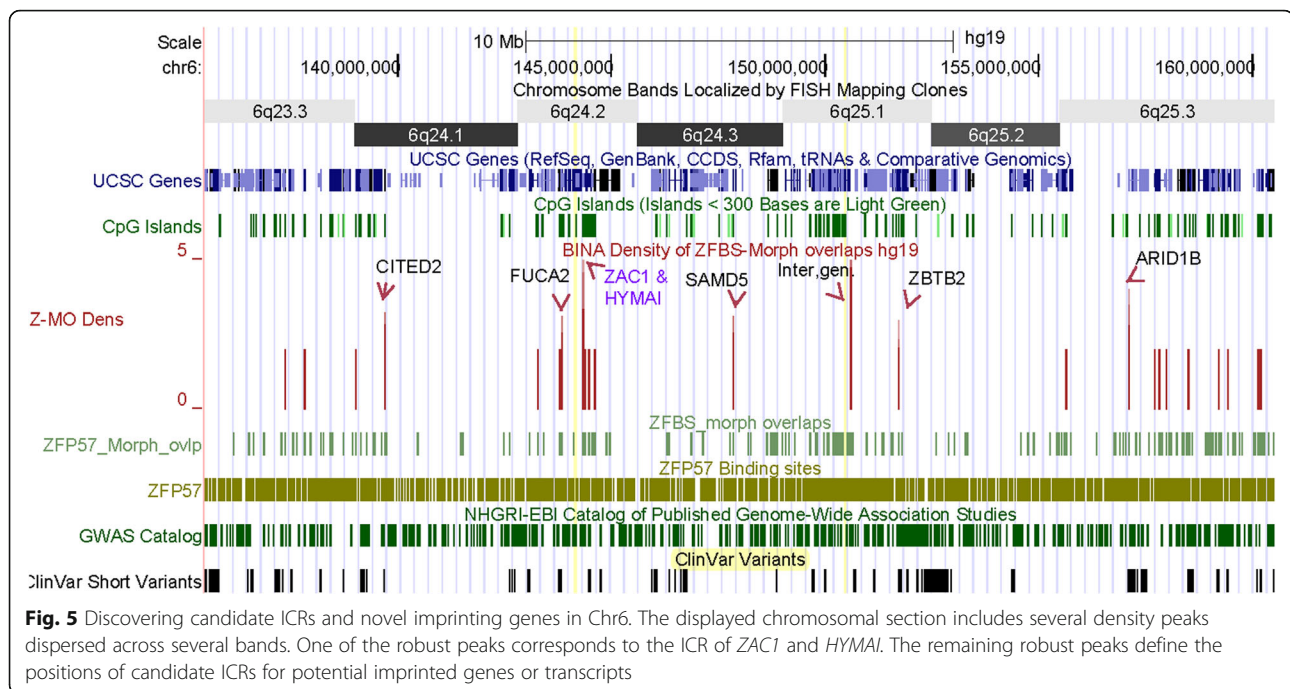
several chromosomal bands (Fig. 5). Even within such a long genomic DNA section, the density peaks appear clearly resolved. In that view, I observed 7 robust density peaks (1 per ~ 3.6 Mb) demonstrating that robust peaks occurred sporadically in human genomic DNA. One of the robust peaks pinpointed the known intragenic ICR of *ZAC1* and *HYMAI* (Figs. 4 and 5). The remaining 6 reflect the positions of candidate ICRs dispersed in various genomic locations. For example, in Chr6q25.1 I noticed an intergenic candidate ICR between *PPP1R14C* and *IYD*. Additional candidates ICRs are dispersed along the 25 Mb DNA mapping to *CITED2*, *FUCA2*, *SAMD5*, *ZBTB2*, and *ARID1B* (Fig. 5, Table 2).



Chr6q24.1 includes the candidate ICR for *CITED2*. The corresponding density peak is in an intragenic CpG island in the last *CITED2* exon (Fig. S6). *CITED2* (CBP/p300-interacting transactivator 2) regulates transcription. Absence of *Cited2* in mouse embryos caused congenital heart disease by perturbing left-right patterning of the body axis [31]. In Chr6q24.2, a candidate ICR maps to an intragenic CpG island near the 5' end of the longest of *FUCA2* transcript (Fig. S7). Genome-wide analyses

have identified *FUCA2* and *IL-18* as novel genes associated with diastolic function in African Americans with sickle cell disease [32]. A previous study deduced that *FUCA2* was biallelically expressed [19]. As listed in a zip file, the analyzed SNP (rs72992630) and primers were from an intragenic *FUCA2* exon [19]. In contrast, the candidate ICR is in a CpG island (CpG52) that encompasses the 1st exon of the longest of *FUCA2* transcript. This exon is at ~9.5 kb upstream of the analyzed SNP





and the primers selected for expression analyses; for details see (Fig. S7).

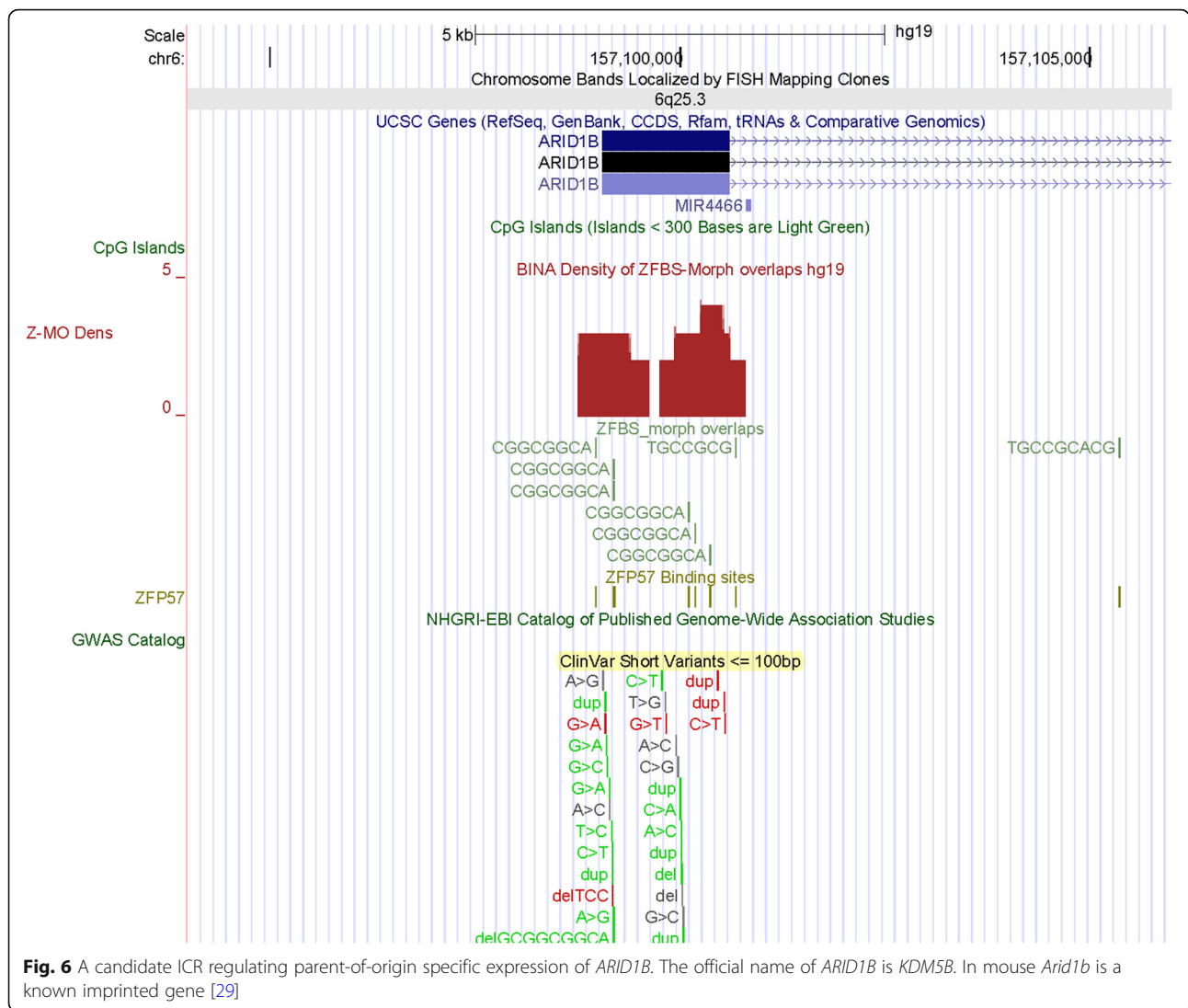
Density-plots include a candidate ICR in chr6q24.3. The corresponding peak maps to a CpG island that encompasses *SAMD5* TSS (Fig. S8). A study found that *SAMD5* was overexpressed in prostate cancer and had powerful prognostic ability for predicting post-operative biochemical recurrence after radical prostatectomy [33]. In Chr6q25.1, a candidate ICR maps to a CpG island at the 5' end of *ZBTB2* (Fig. S9). *ZBTB2* binds DNA and is among the master regulators of the p53 pathway [34]. In mouse embryonic stem cells, *ZBTB2* dynamically interacted with nonmethylated CpG island promoters and regulated differentiation [35]. In colorectal cancer, the abnormal forms of *ZBTB2* increased cell proliferation [36].

Chr6q25.3 includes 1 candidate ICR (Fig. 5). This ICR corresponds to 2 density peaks (Fig. 6). One peak encompasses *ARID1B* TSS. The other maps to the *ARID1B* 1st exon. Since *Arid1b* is a known imprinted gene in mouse [29], from my data one could deduce that its human ortholog also is an imprinted gene. *ARID1B* encodes an enzyme that removes H3K4 methyl-marks from chromatin [37]. In mouse, *Arid1b* haploinsufficiency impacted and disrupted cortical interneuron development [38]. In Chr6q25.1, a candidate ICR is between *PPP1R14C* and *IYD* (Fig. 7). The corresponding density peak is in the vicinity of the 3' end *PPP1R14C* in a DNA segment far upstream of *IYD* TSS. *PPP1R14C* regulates the enzymatic activity of protein phosphatase

1. *IYD* encodes an enzyme that functions in iodide salvage in the thyroid [39]. Thyroid dysmorphogenesis-4 (TDH4) is caused by homozygous mutations in *IYD*. Patients with this defect lack the ability to deiodinate radiolabeled monoiodotyrosine and diiodotyrosine [40]. Notably, the thyroid hormone pathway includes an enzyme (*DIO3*) encoded by an imprinted gene [41].

#### A DNA section from Chr7q includes a candidate ICR for a known imprinted gene (*KLF14*) and a candidate ICR for a potential imprinted gene (*IMPDH1*)

Chr7 contains several known imprinted genes [42, 43]. From Chr7qA, I selected a 4 Mb long DNA covering 3 robust density peaks (1 per 1.3 Mb). These peaks map to *IMPDH1*, *MEST*, and *KLF14* (Fig. 8). The *MEST* and *KLF14* loci includes known imprinted transcripts [30, 44]. In the *MEST* locus, an intragenic ICR regulates parent-of-origin-specific expression of a subset of *MEST* transcripts and *MESTIT1* –a noncoding RNA gene [45, 46]. An enlarged view of density-plots shows a peak within an intragenic CpG island that encompasses the TSSs of both *MEST* and *MESTIT1* (Fig. S10). Thus, a peak at that position correctly pinpointed the ICR in the *MEST* locus. The displayed view also shows the peak that corresponds to *KLF14* (Fig. S10). *KLF14* is a known human imprinted gene. It is the first example of an imprinted gene that has undergone accelerated evolution in the human lineage [30]. To date, I have not found a report locating the ICR of *KLF14*. Notably, density plots



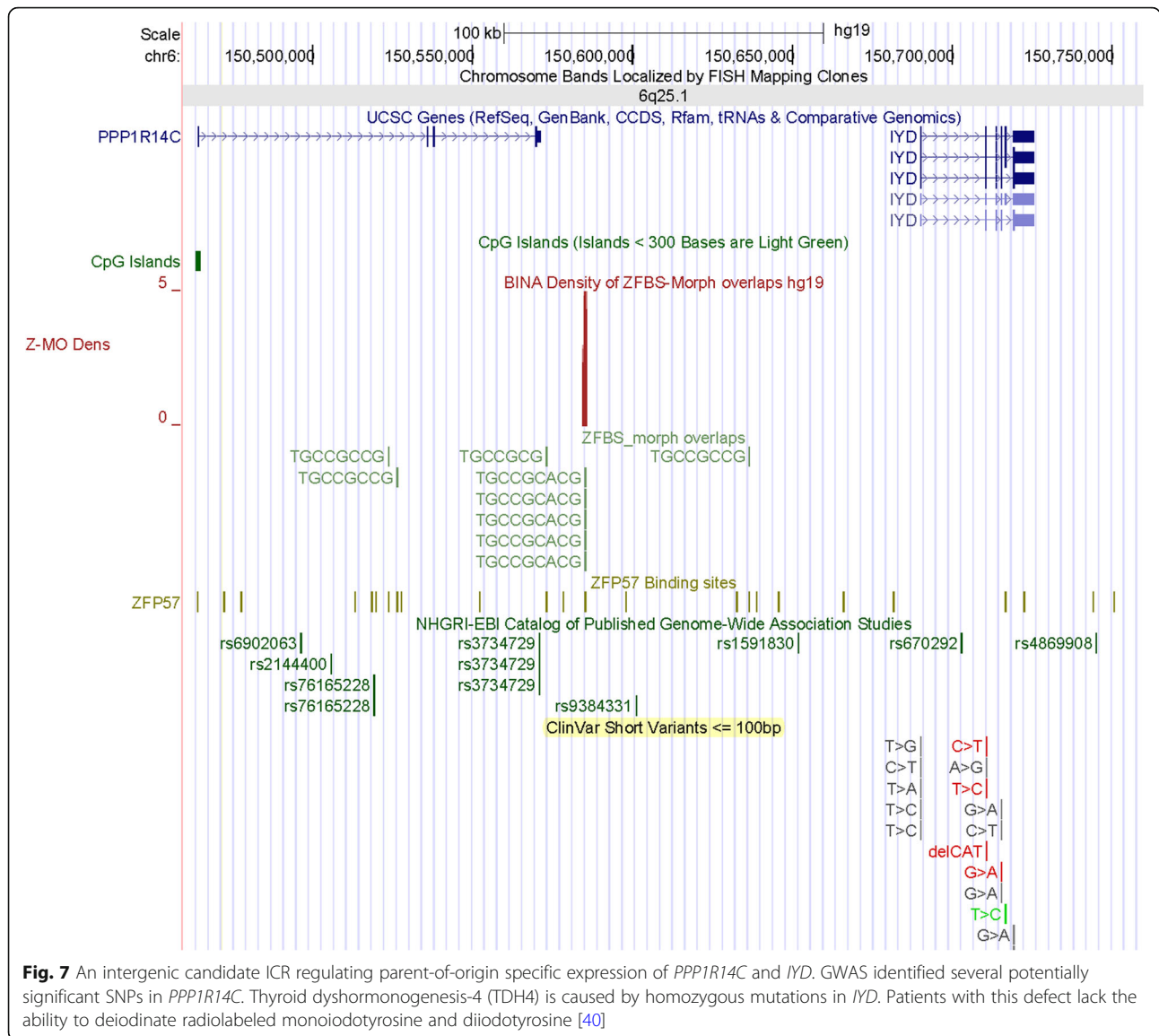
predicted a candidate ICR within a CpG island that encompasses *KLF14* TSS (Fig. S10, Table 2).

Next, I obtained a closeup view to inspect the position of a candidate ICR that mapped to *IMPDH1* (Fig. 9). As observed for the *MEST* locus, this ICR is intragenic and maps to a CpG island that encompasses the TSSs of several short *IMPDH1* transcripts (Fig. 9). Even though *IMPDH1* is expressed in many tissues, its predominant transcripts are produced in the inner segment and synaptic terminals of retinal photoreceptors [47]. The IMPDH proteins form active homo-tetramers that catalyze the rate-limiting step for de novo guanine synthesis by converting inosine monophosphate to xanthosine monophosphate [47]. Deleterious mutations in *IMPDH1* cause Leber congenital amaurosis 11. Manifestations of this genetic anomaly include a group of early-onset childhood retinal dystrophies [48].

**A long DNA section from Chr10q includes a candidate ICR for a potential imprinted gene (*VAX1*)**

In Chr10q, a nearly 8.6 Mb DNA encompasses several chromosomal bands and 2 robust density peaks (1 per 4.3 Mb). One of the peaks is within the *INPP5F* locus. The other maps to *VAX1* (Fig. 10). From *INPP5F* are produced several transcriptional variants. In mouse, one of the variants (*Inpp5f\_v2*) is imprinted in the brain [49]. The TSS of *Inpp5f\_v2* is within a differentially methylated CpG island [49]. In closeup views of density-plots, I observed an intragenic density peak that pinpointed the ICR for parent-of-origin specific expression of human *INPP5F\_v2* (Fig. S11). A candidate ICR is in a CpG island that encompasses TSSs of 2 *VAX1* transcripts (Fig. 11). This gene encodes a transcription factor with a homeobox for binding DNA. *VAX1* is expressed in the pituitary, hypothalamus, and testis [50]. A study





implicated two homozygous mutations in *VAX1* causing microphthalmia associated with cleft lip and palate and agenesis of the corpus callosum [51].

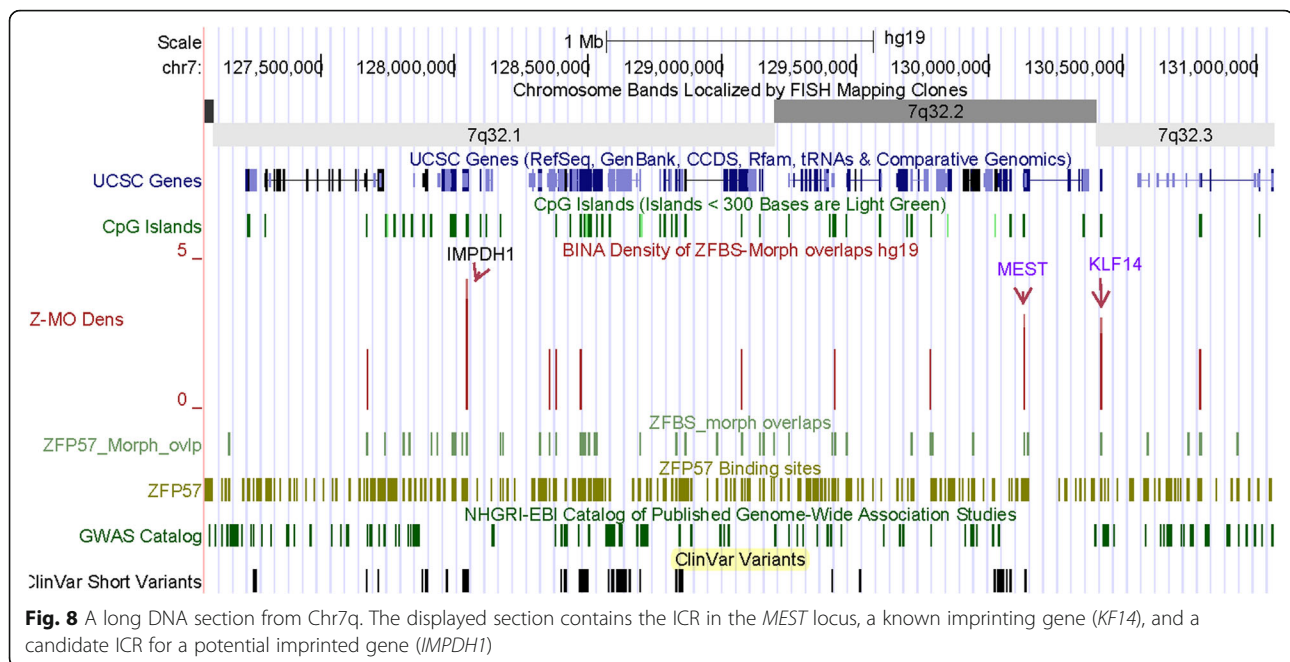
**Examination of candidate ICRs with respect to several previously predicted imprinted genes**

The literature includes various computational strategies for prediction of novel imprinted genes. Examples include sequence or epigenetic features [52, 53]. I looked at a listing of predicted human imprinted genes [54] to assess whether density-plots included candidate ICRs in their vicinity. I located candidate ICRs for several of the predicted genes (Table 3). Several of these ICRs corresponded to density peaks encompass 2 ZFBS-morph overlaps. Therefore, I am not confident whether the predicted ICRs could be true or a false-positive (Table 3). I

did not find a candidate ICR for several of the predicted imprinted genes including *HOXA2*, *HOXA3*, *HOXA5*, *HOXC9*, *HOXC4*, *IFITM1*, *PKP3*, *SLC26A10*, *CDH15*, *RASGRF1*, and *ZRSR1*.

**Discussion**

The discovery of unknown ICRs would facilitate pinpointing the genes in their vicinity and thus uncovering novel imprinted genes. Furthermore, examination of the imprinted genes in the context of clinical variants gives clues into their impacts on human embryonic development, disease states, and genetic anomalies including syndromes [1]. However, to date I could not find any genome-wide studies to methodically discern the genomic positions of the ICRs and to obtain nearly complete listings of novel imprinted genes and



transcripts. Towards these and related goals, I developed a predictive genome-wide strategy. My approach pinpointed several of the known ICRs within relatively long DNA sections. Examples include the ICRs of *H19* – *IGF2* and *KCNQ1* imprinted domains in 1.4 Mb long DNA (Fig. 3); the ICR of *MEST* and *MESTIT1* transcripts in 4.0 Mb DNA (Fig. 8), and the ICR of *INPP5F\_V2* within a nearly 8.6 Mb DNA (Fig. 10). Even along the entire Chr6, I could discern the ICR of *ZAC1* and *HYMAI* in the *PLAGL1* locus (Fig. 4).

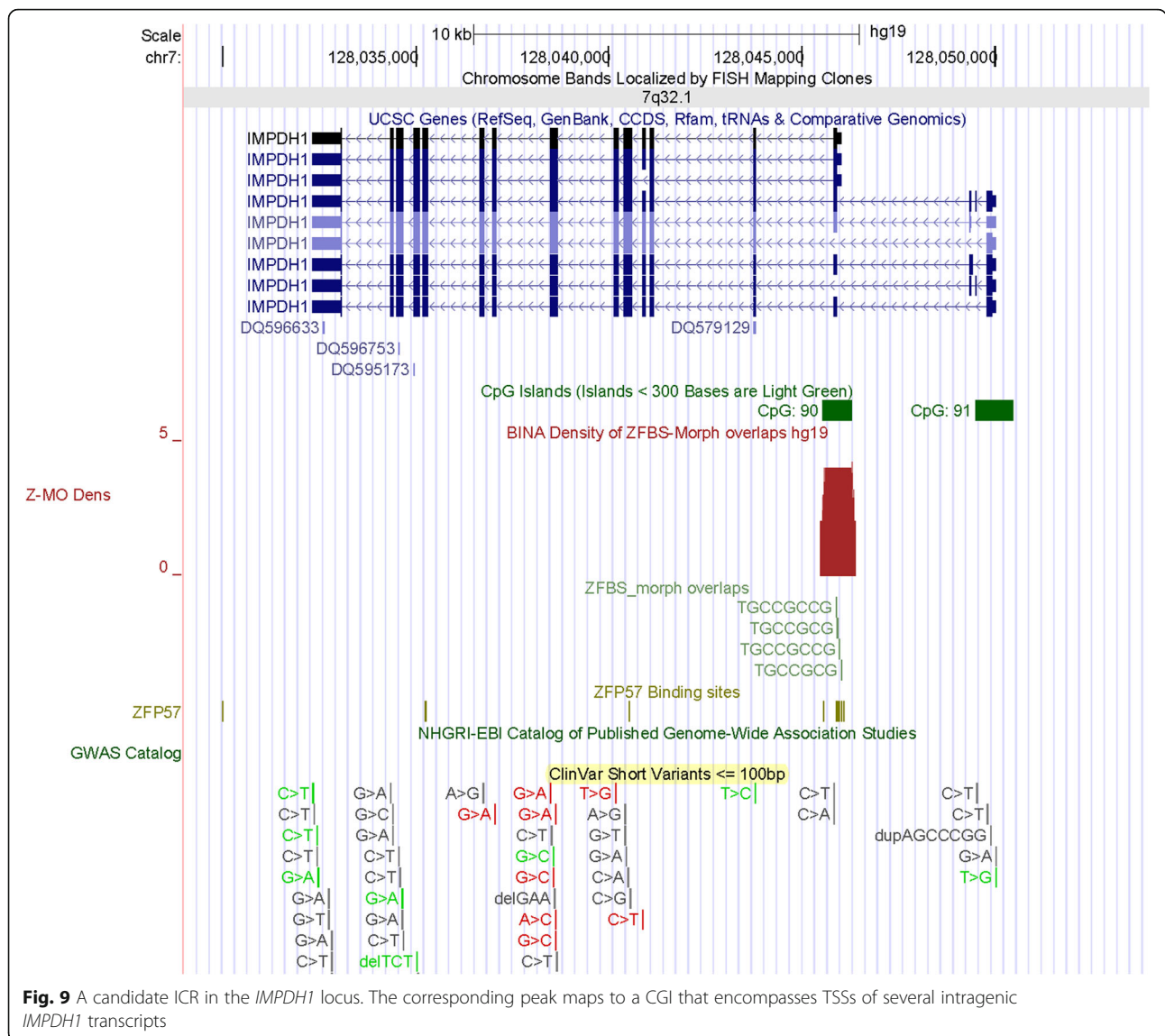
My strategy involves creating density-plots to detect the genomic DNA segments that contain clusters of 2 or more ZFBS-morph overlaps [14, 55, 56]. Previously, I showed that occurrences of such clusters pinpointed ~90% of the fully characterized ICRs/gDMRs in the mouse genome [7, 8, 57]. Even though my approach is predictive, it is based on reports demonstrating the importance of ZFP57 in maintaining allele-specific gene repression [5, 12, 58, 59]. Furthermore, my discovery of ZFBS-morph overlaps has offered mechanistic clues into why ZFP57 is selectively recruited to the ICRs but not elsewhere in the genomic DNA. Briefly, essential to genomic imprinting is a protein complex consisting of DNMT3A and DNMT3L [2, 3, 60]. This complex methylates DNA processively [4]. Since clusters of ZFBS-morph overlaps are CpG-rich, they could provide sites for DNMT3A-DNMT3L to processively methylate the ICRs. Subsequently, the methylated ZFBS-morph overlaps would recruit ZFP57 to associate selectively with ICRs to maintain parent-of-origin specific gene expression [7].

#### The majority of candidate ICRs maps to CpG islands; a subset maps to specific gene transcripts

The UCSC genome browser is highly suitable for examining results of predictive methods in the context of landmarks including the positions of chromosomal bands, genes, transcripts, and the CpG islands [61]. ICR associated islands could be intergenic, encompass promoters, TSSs, and the 1st exon of genes [3, 62]. The ICRs also occur in intragenic CpG [45, 63, 64]. Similarly, in density-plots the analyzed robust peaks primarily map to CpG islands at various genomic locations. Examples include the candidate ICRs for *PPP1R14C*, *IYD*, *CITED2*, *IMPDH1*, and *VAX1* (Fig. 7, S6, 8, 9, and 11). Several of the known ICRs correspond to single transcripts or to transcriptional variants [62]. I observed similar patterns for several of the candidate ICRs. For example, the candidate ICRs for *SAMD5* and *ZBTB2* correspond to isolated transcripts (Figs. S8 and S9). The candidate ICR for *FUCA2* corresponds to the gene longest transcript (Fig. S7). The candidate ICR for *IMPDH1* encompasses intragenic transcripts (Fig. 9).

#### Density-plots revealed candidate ICRs for potential imprinted genes associated with syndromes and disease states

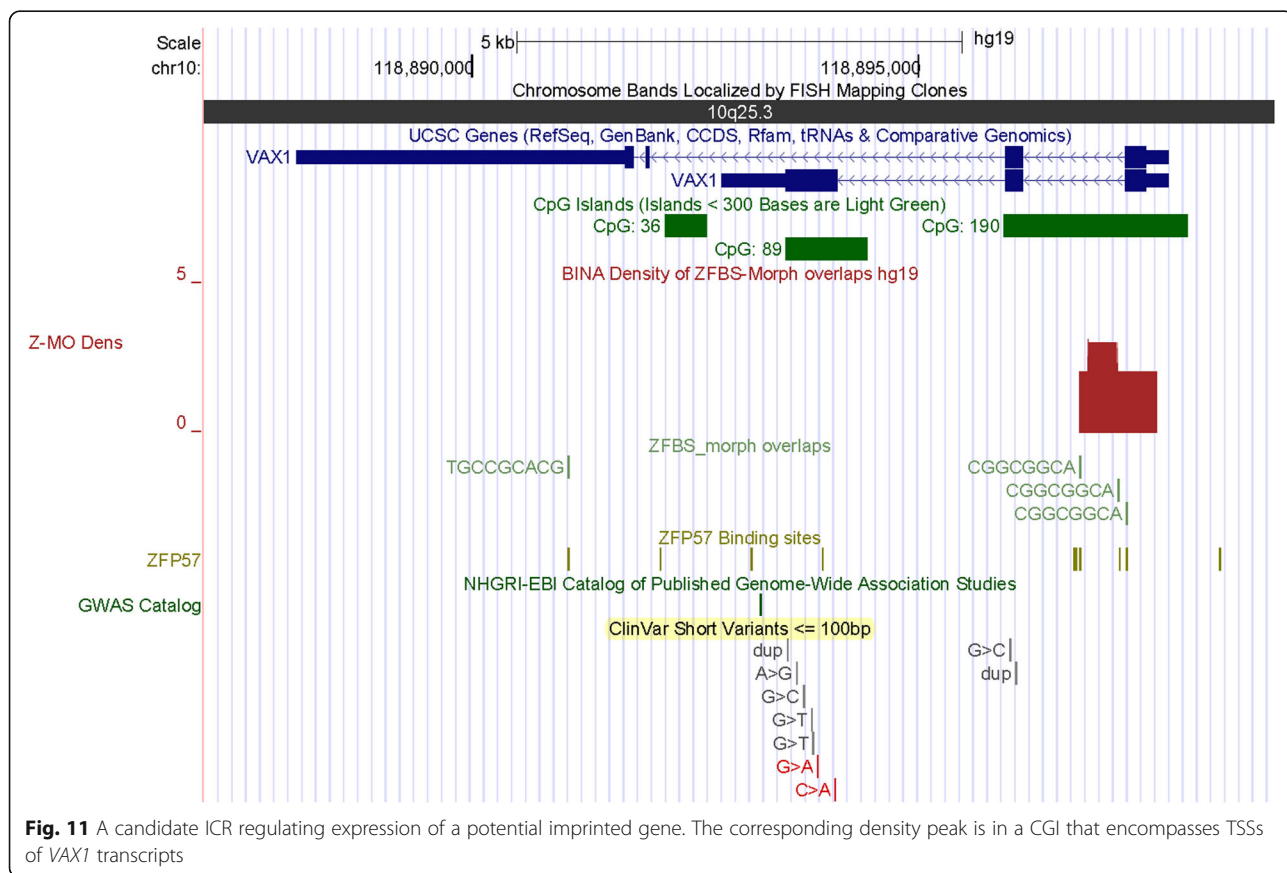
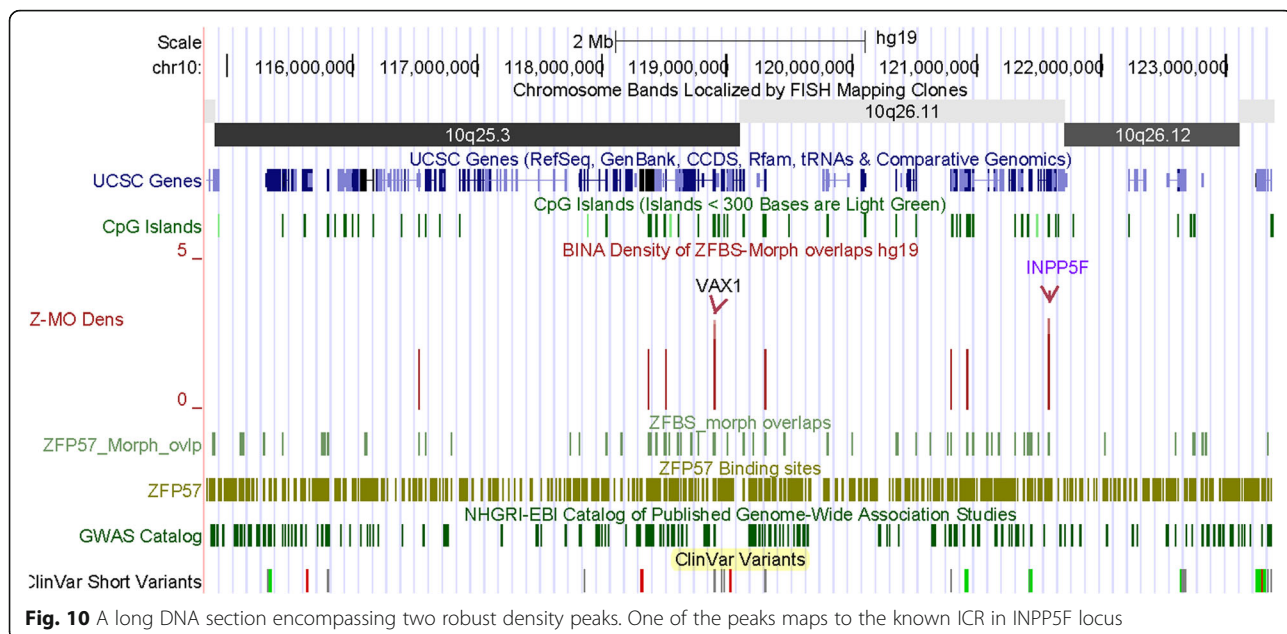
With animal model systems, it is possible to determine whether knockout of a gene would produce a phenotype. In humans, one could examine adverse effects of anomalous loci in the context of clinical variants that produce discernable phenotypes. Examples include developmental disorders, neurological disorders, malformation of



body parts, and syndromes. One could identify these phenotypes from literature surveys, the track displaying the clinical variants at the UCSC browser, or both. For example, examine the figure that displays the positions of short clinical variants with respect to peaks in the density-plot obtained for the entire Chr6 (Fig. 4). As these peaks, the clinical variants are primarily dispersed in gene-rich genomic DNA sections.

*ARID1B* is among the potential human imprinted genes identified by my approach (Fig. 6). In mouse *Arid1b* is a known imprinted gene [29]. My data predicts that human *ARID1B* also is an imprinted gene. This gene encodes an enzyme that removes activating H3K4 methyl marks from chromatin [65]. In human, deleterious variations in *ARID1B* are thought to contribute to Coffin-Siris syndrome. This abnormality is

a multiple malformation syndrome characterized by mental retardation associated with coarse facial features, hypertrichosis, sparse scalp hair, and hypoplastic or absent fifth fingernails or toenails. Other features may include poor overall growth, craniofacial abnormalities, spinal anomalies, and congenital heart defects [66]. Mechanistically, within the protein-networks *ARID1B* interacts with a complex Known as NuRD (Nucleosome Remodeling and Deacetylase); for details see Fig. 1 in reference [7]. In the networks, NuRD is a central node for receiving or transmitting signals via protein-protein interactions. For example, while one of the subunits in NuRD (Mi-2 $\alpha$ ) interacts with TRIM28, its HDAC1 subunit interacts with MLL1, DNMT3A, DNMT3L, and H3K4 demethylases including *ARID1B* [7].



**Table 3** Examination of several predicted human imprinted genes with respect to candidate ICRs observed in density-plots

ICRs Positions in the build hg19	Neighboring transcripts
chr7:42,275,801-42,277,360	<i>GLI3</i>
chr7:150,783,541-150,784,530 *	<i>FASTK</i>
chr7:27,169,571-27,170,500 #	<i>HOXA4</i>
chr7:27,224,001-27,224,600	<i>HOXA11</i>
chr7:27,291,381-27,292,450 &	<i>EVX1</i>
chr6:105,387,976-105,388,975 +	<i>LIN28B</i>
chr7:77,648,446-77,649,555 #	<i>MAGI2</i>

\* This predicted ICR could be a true or a false positive. It is ~ 5400 bp upstream of *FASTK* at the 5' end of *AGAP3*, transcript variant 3

# This predicted ICR encompasses 2 ZFBs-morph overlaps. Therefore, it could be a true or a false positive

& This density predicted ICR is far downstream *EVX1* and upstream of *EVX1-AS*

+ This predicted ICR is far upstream of *LIN28B* and near *LIN28B-AS1*

Several of the potential imprinted genes identified by my approach correspond to a subset of candidate-imprinted genes discovered by experimental techniques [15]. Examples include *SQSTM1*, *PRDM8*, and *NM\_006031/PCNT* (Figs. 1, S1, and S2). A domain in *PRDM8* methylates H3K9 and thus impacts the chromatin structure [67]. Monoallelic expression of *PRDM8* was detected in placental tissues [15]. Genetic studies have observed association of *PRDM8* with progressive myoclonic epilepsy-10. This recessive neurodegenerative disorder is characterized by onset of progressive myoclonus, ataxia, spasticity, dysarthria, and cognitive decline in the first decade of life [68]. *PCNT* (pericentrin) is a component of pericentriolar material that surrounds the two centrioles of a centrosome [69]. Absence of *PCNT* results in disorganized mitotic spindles and missegregation of chromosomes [70]. Heterozygous mutation in *PCNT* caused microcephalic osteodysplastic primordial dwarfism type II [71]. This disorder (MOPD2) is characterized by intrauterine growth retardation, severe proportionate short stature, and microcephaly. Adults with these inherited distinctive physical features have an average height of 100 cm and a brain size comparable to that of a 3-month-old baby. Otherwise, they have near-normal intelligence [70].

Potential imprinted genes discovered de novo by approach include *IMPDH1* (Fig. 9), *CITED2* (Fig. S6), *SAMD5* (Fig. S8), *ZBTB2* (Fig. S9), and *VAX1* (Fig. 11). *IMPDH1* (inosine monophosphate dehydrogenase 1) catalyzes the synthesis of xanthine monophosphate [72]. This reaction is the rate-limiting step in the de novo synthesis of guanine nucleotides. Deleterious mutations in *IMPDH1* often cause Leber congenital amaurosis 11 (LCA11). This anomaly consists of a group of early-onset childhood retinal dystrophies characterized by vision loss, nystagmus, and severe retinal dysfunction [48]. Mutations in *IMPDH1* also may cause a disorder known

as retinitis pigmentosa-10 (RP10). In most patients, RP10 is manifested by early onset and rapid progression of ocular symptoms, initially with night blindness in childhood [73]. The impairment tends to produce stable visual field constriction, although it may worsen very slowly over time. Another potential imprinted gene (*CITED2*) transactivates transcription through interactions with CBP/P300 [74]. Deleterious mutations in *CITED2* cause ventricular septal defect 2 (VSD2). As the most common form of congenital cardiovascular anomaly, VSD2 has affected nearly 50% of all infants with a congenital heart defect and accounts for ~ 15% of cardiac anomalies requiring invasive treatment within the first year of infant's life [75]. Furthermore, congenital VSDs may arise alone or in combination with other cardiac malformations. In that context, it seems relevant that loss of *Cited2* in mouse causes congenital heart disease by perturbing left-right patterning of the body axis [31]. Therefore, it seems plausible that VSDs in infants might also stem from defects in left-right patterning of the body axis in the course of human embryonic development.

*SAMD5* is another potential imprinted gene discovered by approach. In the context of disease-related anomalies, a study found that *SAMD5* was overexpressed in prostate cancer and had powerful prognostic ability on predicting post-operative biochemical recurrence after radical prostatectomy [33]. Another potential imprinted gene (*ZBTB2*) is among the genes identified in approximately 15% of all colorectal cancers [36]. Abnormal forms of *ZBTB2* increased cell proliferation [36]. *ZBTB2* is a transcription factor. Several members of the ZBTB family have emerged as critical factors in the lineage commitment, differentiation, and function of lymphoid cells as well as many other developmental events [76]. Furthermore, *ZBTB2* is among the master regulators of the p53 pathway [34]. In mouse embryonic stem cells, *ZBTB2* dynamically interacted with unmethylated CpG island promoters and regulated differentiation. Another potential imprinted gene (*VAX1*) encodes a transcription factor that controls developmental processes. The structure of *VAX1* includes a homeodomain for binding DNA. In mouse, this homeobox-containing gene is expressed in the developing anterior ventral forebrain [77]. *Vax1* expression was observed in the pituitary, hypothalamus, and testis [50]. From studies of 70 patients, a report found associations of two homozygous mutations in *VAX1* with microphthalmia, cleft lip and palate, and agenesis of the corpus callosum [51]. For an overview [78].

## Conclusion

In this report, I have offered a predictive genome-wide strategy to discover candidate ICRs and novel imprinted

genes. I gave evidence for robustness of my strategy by pinpointing several of the well-known ICRs in relatively long DNA sections. I also gave examples showing that my strategy predicted ICRs for several of the candidate imprinted genes discovered by experimental strategies. The finding that several of the potential imprinted genes impact developmental processes, lends additional support for the robustness of my approach. I also covered examples of how I could deduce the phenotypes of the potential imprinted genes discovered by my approach. Nonetheless, only experimental validations could demonstrate the strength of my approach. Therefore, I offer links for accessing and downloading my data on the positions of ZFBS and ZFBS-Morph overlaps [55], peaks in density-plots [56], and the MLL1 morphemes in the build hg19 of the human genome [79]. Links are also available for accessing data pertaining to the build mm9 of the mouse genome [16, 57, 80].

## Methods

### Marking the genomic positions of ZFP57 binding site and the ZFBS-morph overlaps

From the UCSC genome browser, I downloaded the nucleotide sequences reported for the build hg19 of the human genome. Next, I created two text files: one file containing the nucleotide sequences of the ZFBS-Morph overlaps [8]; the other the hexameric ZFP57 binding site [5]. Using a Perl script, initially I determined the genomic positions the ZFBS-Morph overlaps along the chromosomal DNA sequences. That script opened the file containing the nucleotide sequence of a specified chromosome and the file containing the sequences of the ZFBS-Morph overlaps. After that step, the script moved along the DNA to report the genomic positions of the overlaps. Next, I wrote a subroutine to combine the outputs obtained for various chromosomes. Another subroutine produced a file to create a custom track for displaying the genomic positions of the ZFBS-Morph overlaps at the UCSC genome browser. I followed similar procedures to obtain a file to display the genomic positions of the hexameric ZFP57 binding site at the UCSC genome browser [8]. Reference [55] gives a link for accessing the datafile containing the positions of the ZFBS-Morph overlaps and the hexameric ZFP57 binding site in the build hg19.

### Creating plots of the density of ZFBS-morph overlaps in genomic DNA

With another Perl script, I obtained the genomic positions of DNA segments that covered 2 or more closely spaced ZFBS-Morph overlaps along the chromosomal DNA sequences. That script opened the file containing the positions of ZFBS-Morph overlaps for a specified chromosome. Subsequently, the script scanned the file

to count and to report the number of ZFBS-Morph overlaps within an 850-base window. I chose the window-size by trial and error [14]. Windows covering less than 850 bases tended to give superfluous spikes. Larger windows tended to produce false-peaks. By ignoring their isolated occurrences, the script removed background noise. Next, I combined and tailored the outputs of the program for display as a custom track at the UCSC genome browser. In exploratory studies, I found that density peaks corresponding to 3 or more ZFBS-Morph overlaps appeared reliable. Peaks covering 2 overlaps could be true or false-positive [57]. Reference [56] gives a link for accessing the datafile of density-plots. For an overview about how to use the UCSC genome browser, see [81, 82].

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12864-020-6688-8>.

**Additional file 1 Figure S1.** A candidate ICR for *PRDM8*. Large scale experimental studies listed *PRDM8* as a candidate imprinted gene [15]. Due to its vicinity to a robust density peak, my strategy also predicts that *PRDM8* is potential imprinted gene. **Figure S2.** A candidate ICR for *PCNT*. Large scale experimental studies listed *PCNT* as a candidate imprinted gene [15]. Since *PCNT* includes 2 intragenic robust density peaks, my strategy predicts that *PCNT* is a potential imprinted gene. **Figure S3.** A density peak in the vicinity of *WDR60*. Results of large-scale experimental studies listed *WDR60* as a candidate imprinted gene [15]. Within *WDR60*, I noticed a peak covering 2 ZFBS-morph overlaps. This peak could be a false or a true-positive. If false-positive, then another peak -far upstream of (*WDR60*)- could be a candidate ICR for regulating parent-of-origin specific expression of both *WDR60* and *ESYT2*. The latter gene encodes a protein (synaptotagmin-like protein 2) that belongs to a family of membranous Ca<sup>2+</sup> -sensors. **Figure S4.** The positions of density peaks with respect to the ICR of *H19* - *IGF2* imprinted domain. In a few cases, an ICR encompasses two density peaks. The figure below includes a track displaying "Updated CTCF Binding sites predictions". See references [25, 26] for updated positions of the unique repeats and CTCF sites upstream of *H19* TSS. Results of the ENCODE ChIPs do not support the existence of the predicted CTCF site 5 described previously [24]. Furthermore, in results of ChIPs displayed at the UCSC genome browser, I noticed a chromatin boundary consisting of CTCF, RAD21, and SMC3 in a CpG island upstream of *H19* TSS [25]. I named the predicted site in that island CTCF site 8. **Figure S5.** A peak in the density plots correctly locating the KvDMR in the *KCNQ1* imprinted domain. **Figure S6.** A candidate ICR for a potential imprinted gene (*CITED2*). This gene encodes a regulator of transcription. Absence of *Cited2* in mouse embryos caused congenital heart disease by perturbing left-right patterning of the body axis [31]. Deleterious mutations in *CITED2* cause VSD2 -ventricular septal defect 2 [75]. **Figure S7.** A candidate ICR for a potential imprinted gene (*FUCA2*). Note the position of the candidate ICR with respect to the SNP (rs72992630) and primers used to deduce that *FUCA2* is biallelically expressed gene [19]. The SNP and primers are not in the vicinity of the 1st exon of the transcript that is associated with a candidate ICR. **Figure S8.** A candidate ICR for a potential imprinted gene (*SAMD5*). A study found that *SAMD5* was overexpressed in prostate cancer and had powerful prognostic ability for predicting postoperative biochemical recurrence after radical prostatectomy [33]. **Figure S9.** A density peak predicted a candidate ICR for a potential imprinted gene (*ZBTB2*). In mouse embryonic stem cells, *ZBTB2* dynamically interacted with nonmethylated CpG island promoters and regulated differentiation [35]. In colorectal cancer, the abnormal forms of *ZBTB2* increased cell proliferation [36]. **Figure S10.** Density peaks mapping to known parent-of-origin specific transcripts. One peak correctly located the ICR at the *MEST* locus. This ICR is

intragenic and encompasses the TSSs of *MEST11* (a noncoding RNA gene) and a subset of *MEST* transcripts. *KLF14* is a known imprinted gene [30]. Density-plots predicted a candidate ICR regulating its imprinted expression. **Figure S11**. In density-plots, a peak correctly located the ICR regulating the expression of *INPP5\_2*.

### Abbreviations

CGIs: CpG islands

### Acknowledgments

I thank Arnold Stein for helpful discussions.

### Authors' contributions

MB is the solo author who carried out this study designed the approach, performed the analyses, and wrote the manuscript.

### Funding

None.

### Availability of data and materials

You can access the data via the following links.

The positions of ZFBs and ZFBs-Morph overlaps in the build hg19 of the human genome: <https://purr.purdue.edu/publications/3208/1>

Density of ZFBs-Morph overlaps in the build hg19 of the human genome: <https://purr.purdue.edu/publications/2967/1>

The positions of the MLL morphemes in the build hg19 of the human genome: <https://purr.purdue.edu/publications/1639/1>

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable

### Competing interests

The author declares that she has no competing interests.

Received: 18 October 2019 Accepted: 18 March 2020

Published online: 31 May 2020

### References

- Grafodatskaya D, Choufani S, Basran R, Weksberg R. An update on molecular diagnostic testing of human imprinting disorders. *J Pediatr Genet*. 2017;6(1):3–17.
- Strogantsev R, Ferguson-Smith AC. Proteins involved in establishment and maintenance of imprinted methylation marks. *Brief Funct Genomics*. 2012; 11(3):227–39.
- Stewart KR, Veselovska L, Kelsey G. Establishment and functions of DNA methylation in the germline. *Epigenomics*. 2016;8(10):1399–413.
- Holz-Schietinger C, Reich NO. The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by DNMT3L. *J Biol Chem*. 2010;285(38):29091–100.
- Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell*. 2011;44(3):361–72.
- Bird AP. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res*. 1980;8(7):1499–504.
- Bina M. Imprinted control regions include composite DNA elements consisting of the ZFP57 binding site overlapping MLL1 morphemes. *Genomics*. 2017;109:361–72.
- Bina M, Wyss P, Song XC. Datasets on the genomic positions of the MLL1 morphemes, the ZFP57 binding site, and ZFBs-Morph overlaps in the build mm9 of the mouse genome. *Data Brief*. 2017;13:202–7.
- Bina M, Wyss P, Novorolsky E, Zulkelfi N, Xue J, Price R, Fay M, Gutmann Z, Fogler B, Wang D. Discovery of MLL1 binding units, their localization to CpG Islands, and their potential function in mitotic chromatin. *BMC Genomics*. 2013;14:927.
- Birke M, Schreiner S, Garcia-Cuellar MP, Mahr K, Titgemeyer F, Slany RK. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic Acids Res*. 2002;30(4):958–65.
- Bach C, Mueller D, Buhl S, Garcia-Cuellar MP, Slany RK. Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene*. 2009;28(6):815–23.
- Riso V, Cammisa M, Kukreja H, Anvar Z, Verde G, Sparago A, Acurzio B, Lad S, Lonardo E, Sankar A, et al. ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells. *Nucleic Acids Res*. 2016;44(17):8165–78.
- Bina M, Demmon S, Pares-Matos EI. Syndromes associated with Homo sapiens pol II regulatory genes. *Prog Nucleic Acid Res Mol Biol*. 2000;64:171–219.
- Bina M, Wyss P. Simultaneous discovery of candidate imprinted genes and Imprinting Control Regions in the mouse genome. *bioRxiv*. 2019; <https://doi.org/10.1101/780551>.
- Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, Forrest MS, Campino S, Clark TG, Stanier P, Kwiatkowski D, et al. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genet*. 2010;11:25.
- Bina M, Wyss PJ, Song XC. Density of ZFBs-Morph overlaps in the build mm9 of the mouse genome. *Purdue Univ Res Repository*. 2018. <https://doi.org/10.4231/R7W37TJH> <https://purr.purdue.edu/publications/2961/1>.
- OMIM. <https://www.omim.org/entry/615503>.
- Arima T, Dreweil RA, Oshimura M, Wake N, Surani MA. A novel imprinted gene, HYMAI, is located within an imprinted domain on human chromosome 6 containing ZAC. *Genomics*. 2000;67(3):248–55.
- Iglesias-Platas I, Court F, Camprubi C, Sparago A, Guillaumet-Adkins A, Martin-Trujillo A, Riccio A, Moore GE, Monk D. Imprinting at the PLAGL1 domain is contained within a 70-kb CTCF/cohesin-mediated non-allelic chromatin loop. *Nucleic Acids Res*. 2013;41(4):2171–9.
- Arima T, Yamasaki K, John RM, Kato K, Sakumi K, Nakabeppu Y, Wake N, Kono T. The human HYMAI/PLAGL1 differentially methylated region acts as an imprint control region in mice. *Genomics*. 2006;88(5):650–8.
- Verona RI, Mann MR, Bartolomei MS. Genomic imprinting: intricacies of epigenetic regulation in clusters. *Annu Rev Cell Dev Biol*. 2003;19:237–59.
- Choufani S, Shuman C, Weksberg R, Beckwith-Wiedemann syndrome. *Am J Med Genet C: Semin Med Genet*. 2010;154C(3):343–54.
- Frevell MA, Sowerby SJ, Petersen GB, Reeve AE. Methylation sequencing analysis refines the region of H19 epimutation in Wilms tumor. *J Biol Chem*. 1999;274(41):29331–40.
- Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*. 2000;405(6785):482–5.
- Bina M. Assessment of the CTCF Binding Sites and Repeat-Positions Upstream the Human H19 Gene. *bioRxiv*. 2018; <https://doi.org/10.1101/250407>.
- Bina M. Positions of predicted CTCF binding sites and unique A and B-repeats in the build hg19 of the human genome. *Purdue Univ Res Repository*. 2018; <https://purr.purdue.edu/publications/2900/1>.
- Shin JY, Fitzpatrick GV, Higgins MJ. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. *EMBO J*. 2008;27(1):168–78.
- Xin Z, Soejima H, Higashimoto K, Yatsuki H, Zhu X, Satoh Y, Masaki Z, Kaneko Y, Jinno Y, Fukuzawa R, et al. A novel imprinted gene, KCNQ1DN, within the WT2 critical region of human chromosome 11p15.5 and its reduced expression in Wilms' tumors. *J Biochem*. 2000;128(5):847–53.
- Gigante S, Gouil Q, Lucattini A, Keniry A, Beck T, Tinning M, Gordon L, Woodruff C, Speed TP, Blewitt ME, et al. Using long-read sequencing to detect imprinted DNA methylation. *Nucleic Acids Res*. 2019.
- Parker-Katirae L, Carson AR, Yamada T, Arnaud P, Feil R, Abu-Amero SN, Moore GE, Kaneda M, Perry GH, Stone AC, et al. Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLoS Genet*. 2007;3(5):e65.
- Lopes Floro K, Artap ST, Preis JI, Fatkin D, Chapman G, Furtado MB, Harvey RP, Hamada H, Sparrow DB, Dunwoodie SL. Loss of Cited2 causes congenital heart disease by perturbing left-right patterning of the body axis. *Hum Mol Genet*. 2011;20(6):1097–110.
- Duarte JD, Desai AA, Sysol JR, Abbasi T, Patel AR, Lang RM, Gupta A, Garcia JG, Gordeuk VR, Machado RF. Genome-wide analysis identifies IL-18 and FUCA2 as novel genes associated with diastolic function in African Americans with sickle cell disease. *PLoS One*. 2016;11(9):e0163013.

33. Li F, Xu Y, Liu RL. SAMD5 mRNA was overexpressed in prostate cancer and can predict biochemical recurrence after radical prostatectomy. *Int Urol Nephrol*. 2019;51(3):443–51.
34. Jeon BN, Choi WI, Yu MY, Yoon AR, Kim MH, Yun CO, Hur MW. ZBTB2, a novel master regulator of the p53 pathway. *J Biol Chem*. 2009;284(27):17935–46.
35. Karemaker ID, Vermeulen M. ZBTB2 reads unmethylated CpG island promoters and regulates embryonic stem cell differentiation. *EMBO Rep*. 2018;19(4).
36. Gylfe AE, Kondelin J, Turunen M, Ristolainen H, Katainen R, Pitkanen E, Kaasinen E, Rantanen V, Tanskanen T, Varjosalo M, et al. Identification of candidate oncogenes in human colorectal cancers with microsatellite instability. *Gastroenterology*. 2013;145(3):540–543 e522.
37. Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, Xu G, Li E, Chen T. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature*. 2009;461(7262):415–8.
38. Jung EM, Moffat JJ, Liu J, Dravid SM, Gurumurthy C, Kim WY. Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. *Nat Neurosci*. 2017;20(12):1694–707.
39. Friedman JE, Watson JA Jr, Lam DW, Rokita SE. Iodotyrosine deiodinase is the first mammalian member of the NADH oxidase/flavin reductase superfamily. *J Biol Chem*. 2006;281(5):2812–9.
40. OMIM. <https://omim.org/entry/274800>.
41. Kota SK, Lleres D, Bouschet T, Hirasawa R, Marchand A, Begon-Pescia C, Sanli I, Arnaud P, Journot L, Girardot M, et al. ICR noncoding RNA expression controls imprinting and DNA replication at the Dlk1-Dio3 domain. *Dev Cell*. 2014;31(1):19–33.
42. Monk D, Wagschal A, Arnaud P, Muller PS, Parker-Katirae L, Bourc'his D, Scherer SW, Feil R, Stanier P, Moore GE. Comparative analysis of human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, TFPI2/Tfpi2, which requires EHMT2 and EED for allelic-silencing. *Genome Res*. 2008;18(8):1270–81.
43. Barbaux S, Gascoin-Lachambre G, Buffat C, Monnier P, Mondon F, Tonanny MB, Pinaud A, Auer J, Bessieres B, Barlier A, et al. A genome-wide approach reveals novel imprinted genes expressed in the human placenta. *Epigenetics*. 2012;7(9):1079–90.
44. Kobayashi S, Kohda T, Miyoshi N, Kuroiwa Y, Aisaka K, Tsutsumi O, Kaneko-Ishino T, Ishino F. Human PEG1/MEST, an imprinted gene on chromosome 7. *Hum Mol Genet*. 1997;6(5):781–6.
45. Kosaki K, Kosaki R, Craigen WJ, Matsuo N. Isoform-specific imprinting of the human PEG1/MEST gene. *Am J Hum Genet*. 2000;66(1):309–12.
46. Nakabayashi K, Bentley L, Hitchins MP, Mitsuya K, Meguro M, Minagawa S, Bamforth JS, Stanier P, Preece M, Weksberg R, et al. Identification and characterization of an imprinted antisense RNA (MEST1T1) in the human MEST locus on chromosome 7q32. *Hum Mol Genet*. 2002;11(15):1743–56.
47. Bowne SJ, Liu Q, Sullivan LS, Zhu J, Spellacy CJ, Rickman CB, Pierce EA, Daiger SP. Why do mutations in the ubiquitously expressed housekeeping gene IMPDH1 cause retina-specific photoreceptor degeneration? *Invest Ophthalmol Vis Sci*. 2006;47(9):3754–65.
48. OMIM. <https://www.omim.org/entry/613837>.
49. Choi JD, Underkoffler LA, Wood AJ, Collins JN, Williams PT, Golden JA, Schuster EF Jr, Loomes KM, Oakey RJ. A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG island. *Mol Cell Biol*. 2005;25(13):5514–22.
50. Hoffmann HM, Tamrazian A, Xie H, Perez-Millan MI, Kauffman AS, Mellon PL. Heterozygous deletion of ventral anterior homeobox (vax1) causes subfertility in mice. *Endocrinology*. 2014;155(10):4043–53.
51. Slavotinek AM, Chao R, Vacik T, Yahyavi M, Abouzeid H, Bardakjian T, Schneider A, Shaw G, Sherr EH, Lemke G, et al. VAX1 mutation associated with microphthalmia, corpus callosum agenesis, and orofacial clefting: the first description of a VAX1 phenotype in humans. *Hum Mutat*. 2012;33(2):364–8.
52. Luedi PP, Hartemink AJ, Jirtle RL. Genome-wide prediction of imprinted murine genes. *Genome Res*. 2005;15(6):875–84.
53. Brideau CM, Eilertson KE, Hagarman JA, Bustamante CD, Soloway PD. Successful computational prediction of novel imprinted genes from epigenomic features. *Mol Cell Biol*. 2010;30(13):3357–70.
54. geneimprint. <http://www.geneimprint.com/site/genes-by-species>.
55. Bina M, Wyss PJ, Song XC. The positions of ZFBS and ZFBS-Morph overlaps in the build hg19 of the human genome. *Purdue Univ Res Repository*. 2019. <https://doi.org/10.4231/SK8Y-EA21> <https://purr.purdue.edu/publications/3208/1>.
56. Bina M, Wyss PJ, Song XC. Density of ZFBS-Morph overlaps in the build hg19 of the human genome. *Purdue Univ Res Repository*. 2019. <https://doi.org/10.4231/PR3C-J556> <https://purr.purdue.edu/publications/2967/1>.
57. Bina M, Wyss PJ, Song XC. The positions of ZFBS and ZFBS-Morph overlaps in the build mm9 of the mouse genome. *Purdue Univ Res Repository*. 2017. <https://doi.org/10.4231/R7C82782> <https://purr.purdue.edu/publications/2473/1>.
58. Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC. A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev Cell*. 2008;15(4):547–57.
59. Strogantsev R, Krueger F, Yamazawa K, Shi H, Gould P, Goldman-Roberts M, McEwen K, Sun B, Pedersen R, Ferguson-Smith AC. Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol*. 2015;16:112.
60. Aref-Eshghi E, Laila C, Schenkel LC, Carere DE, David I, Rodenhiser DE, Sadikovic B. Epigenomic mechanisms of human developmental disorders, *Translational Epigenetics*, vol. 6. 2nd ed; 2018.
61. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. The UCSC genome browser database: 2019 update. *Nucleic Acids Res*. 2019;47(D1):D853–8.
62. Ferguson-Smith AC, Bourc'his D. The discovery and importance of genomic imprinting. *Elife*. 2018;7.
63. Wood AJ, Bourc'his D, Bestor TH, Oakey RJ. Allele-specific demethylation at an imprinted mammalian promoter. *Nucleic Acids Res*. 2007;35(20):7031–9.
64. Abdollahi A. LOT1 (ZAC1/PLAGL1) and its family members: mechanisms and functions. *J Cell Physiol*. 2007;210(1):16–25.
65. Vallianatos CN, Iwase S. Disrupted intricacy of histone H3K4 methylation in neurodevelopmental disorders. *Epigenomics*. 2015;7(3):503–19.
66. OMIM. <https://www.omim.org/entry/135900>.
67. Eom GH, Kim K, Kim SM, Kee HJ, Kim YJ, Jin HM, Kim JR, Kim JH, Choe N, Kim KB, et al. Histone methyltransferase PRDM8 regulates mouse testis steroidogenesis. *Biochem Biophys Res Commun*. 2009;388(1):131–6.
68. OMIM. <https://www.omim.org/entry/616640>.
69. Doxsey SJ, Stein P, Evans L, Calarco PD, Kirschner M. Pericentrin, a highly conserved centrosome protein involved in microtubule organization. *Cell*. 1994;76(4):639–50.
70. Rauch A, Thiel CT, Schindler D, Wick U, Crow YJ, Ekici AB, van Essen AJ, Goecke TO, Al-Gazali L, Chranowska KH, et al. Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. *Science*. 2008;319(5864):816–9.
71. OMIM. <https://www.omim.org/entry/210720>.
72. Carr SF, Papp E, Wu JC, Natsumeda Y. Characterization of human type I and type II IMP dehydrogenases. *J Biol Chem*. 1993;268(36):27286–90.
73. OMIM. <https://www.omim.org/entry/180105>.
74. Bhattacharya S, Michels CL, Leung MK, Arany ZP, Kung AL, Livingston DM. Functional role of p35srj, a novel p300/CBP binding protein, during transactivation by HIF-1. *Genes Dev*. 1999;13(1):64–75.
75. OMIM. <https://www.omim.org/entry/614431>.
76. Zhu C, Chen G, Zhao Y, Gao XM, Wang J. Regulation of the development and function of B cells by ZBTB transcription factors. *Front Immunol*. 2018;9:580.
77. Hallonet M, Hollemann T, Wehr R, Jenkins NA, Copeland NG, Pieler T, Gruss P. Vax1 is a novel homeobox-containing gene expressed in the developing anterior ventral forebrain. *Development*. 1998;125(14):2599–610.
78. OMIM. <https://www.omim.org/entry/614402>.
79. Bina M, Wyss PJ, Wang D, Zulkefl N, Novorolsky E. Position of MLL1 morphemes in human genomic DNA. *Purdue Univ Res Repository*. 2014. <https://doi.org/10.4231/R7H41PBT> <https://purr.purdue.edu/publications/1639/1>.
80. Bina M, Wyss P, Wang D, Song X. Localization of MLL1 morphemes in mouse mm9 genomic DNA. *Purdue Univ Res Repository*. 2014. <https://doi.org/10.4231/R7KW5CXF> <https://purr.purdue.edu/publications/1648/1>.
81. Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. UCSC genome browser tutorial. *Genomics*. 2008;92(2):75–84.
82. Bina M. The genome browser at UCSC for locating genes, and much more! *Mol Biotechnol*. 2008;38(3):269–75.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.