

RESEARCH

Open Access



# The mismatch-repair proteins MSH2 and MSH6 interact with the imprinting control regions through the ZFP57-KAP1 complex

Basilia Acurzio<sup>1,2†</sup>, Francesco Cecere<sup>1,2†</sup>, Carlo Giaccari<sup>1,2</sup>, Ankit Verma<sup>1,2</sup>, Rosita Russo<sup>1</sup>, Mariangela Valletta<sup>1</sup>, Bruno Hay Mele<sup>3</sup>, Claudia Angelini<sup>4</sup>, Angela Chambery<sup>1</sup> and Andrea Riccio<sup>1,2\*</sup>

## Abstract

**Background:** Imprinting Control Regions (ICRs) are CpG-rich sequences acquiring differential methylation in the female and male germline and maintaining it in a parental origin-specific manner in somatic cells. Despite their expected high mutation rate due to spontaneous deamination of methylated cytosines, ICRs show conservation of CpG-richness and CpG-containing transcription factor binding sites in mammalian species. However, little is known about the mechanisms contributing to the maintenance of a high density of methyl CpGs at these loci.

**Results:** To gain functional insights into the mechanisms for maintaining CpG methylation, we sought to identify the proteins binding the methylated allele of the ICRs by determining the interactors of ZFP57 that recognizes a methylated hexanucleotide motif of these DNA regions in mouse ESCs. By using a tagged approach coupled to LC-MS/MS analysis, we identified several proteins, including factors involved in mRNA processing/splicing, chromosome organization, transcription and DNA repair processes. The presence of the post-replicative mismatch-repair (MMR) complex components MSH2 and MSH6 among the identified ZFP57 interactors prompted us to investigate their DNA binding profile by chromatin immunoprecipitation and sequencing. We demonstrated that MSH2 was enriched at gene promoters overlapping unmethylated CpG islands and at repeats. We also found that both MSH2 and MSH6 interacted with the methylated allele of the ICRs, where their binding to DNA was mediated by the ZFP57/KAP1 complex.

**Conclusions:** Our findings show that the MMR complex is concentrated on gene promoters and repeats in mouse ESCs, suggesting that maintaining the integrity of these regions is a primary function of highly proliferating cells. Furthermore, the demonstration that MSH2/MSH6 are recruited to the methylated allele of the ICRs through interaction with ZFP57/KAP1 suggests a role of the MMR complex in the maintenance of the integrity of these regulatory regions and evolution of genomic imprinting in mammalian species.

**Keywords:** LC-MS/MS, Chip-seq, Allele-specific analysis, CpG islands, Methyl CpG, Genomic imprinting, Mismatch-repair, Transcription factor binding, Cytosine deamination

## Background

Genomic imprinting is a gene regulatory mechanism of mammals based on differential acquisition of epigenetic marks during female and male gametogenesis [1]. These marks oppositely influence the expression of the maternally inherited and paternally inherited alleles of nearby genes at later stages of development. The best characterized form of imprinted gene modification is CpG

<sup>†</sup>Basilia Acurzio and Francesco Cecere contributed equally

\*Correspondence: andrea.riccio@unicampania.it

<sup>1</sup> Department of Environmental Biological and Pharmaceutical Sciences and Technologies (DISTABIF), Università Degli Studi Della Campania "Luigi Vanvitelli", 81100 Caserta, Italy  
Full list of author information is available at the end of the article



© The Author(s) 2022. **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.

methylation, and most imprinting control regions (ICRs) identified so far overlap CpG-rich sequences (germline-derived Differentially Methylated Regions), whose methylation is differentially established in female and male germ cells and faithfully maintained in somatic cells throughout development.

Maintenance of differential methylation of the ICRs is particularly critical during pre-implantation development, a stage in which epigenetic reprogramming occurs. Several human diseases (overall known as imprinting disorders) arise from loss or gain of methylation at specific imprinted loci at this stage [2]. A multi-protein complex including the zinc-finger protein ZFP57 and its cofactor KAP1 specifically recognizes the methylated allele of the ICRs and is required for maintaining their DNA methylation in mouse embryos and embryonic stem cells [3–6]. ZFP57/KAP1 binding is also necessary for maintaining different histone marks on the maternal and paternal alleles of the ICRs and allele-specific expression of the imprinted genes [7–10]. In humans, biallelic loss of function mutations of *ZFP57* lead to Transient Neonatal Diabetes type 1 and multi-locus imprinting disturbances [11]. While ZFP57 recognizes multiple methylated CpG-containing motifs on one allele of the ICRs, the other and non-methylated allele of these regulatory regions is bound by several other transcription factors required for preventing de novo methylation and promoting transcription of the imprinted genes [2].

The mutation rate of CpGs in mammalian genomes is generally elevated because these dinucleotides are often methylated, and methylated cytosine is unstable as it undergoes deamination to thymine, which if uncorrected yields a C to T transition [12]. This mechanism is mostly responsible for the lower frequency of the CpG dinucleotide in mammalian genomes. An exception is represented by the CpG islands that overlap gene promoters and are generally non-methylated in the germline, thus preserving their CpG-richness. Although the ICRs are methylated on one allele in the germline, they maintain their characteristics of 2–4 kb long CpG-rich regions and show conservation of CpG-containing transcription factor binding sites in mammals [13, 14].

The mismatch-repair (MMR) complex recognizes misincorporated bases in double-stranded DNA and is required to repair G:T mis-matches resulting from deamination of 5-methylcytosine (5mC) [15]. It is worth noting that mutations arising from unrepaired 5mC deamination events are prevalent in MMR-deficient cancers, particularly those deficient in the MutS $\alpha$  heterodimer that is composed by the MSH2-MSH6 heterodimer [15].

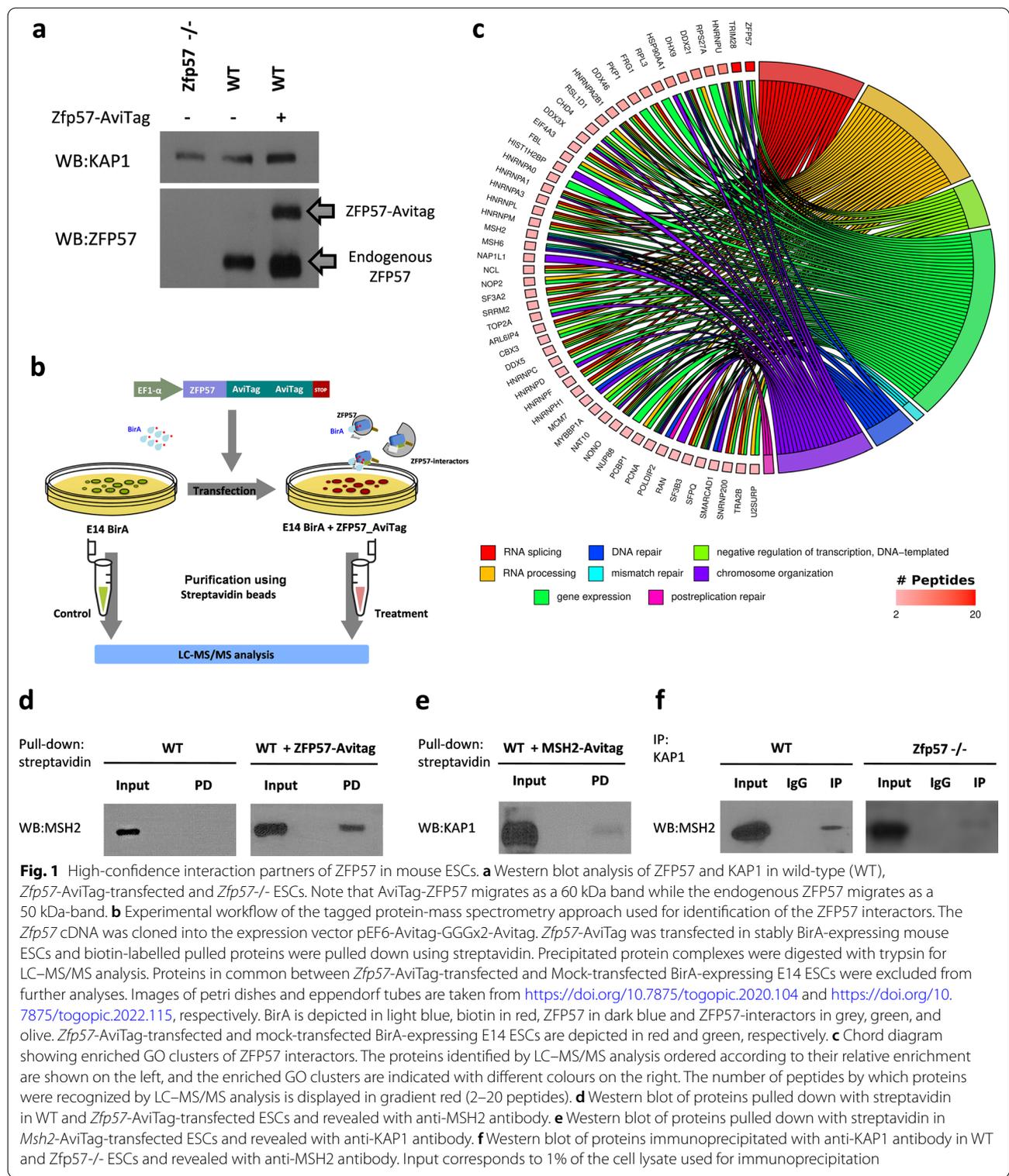
To gain further insights into the mechanisms of maintenance of methyl CpGs in the ICRs, we sought to identify the proteins interacting with the methylated allele

of the ICRs, by determining the ZFP57 interactors in mouse ESCs. We were able to identify several factors involved in mRNA processing/splicing, chromosome organization, transcription and DNA repair processes. We focused on the components of the mismatch-repair complex MSH2 and MSH6 and demonstrated their interactions with the ICRs by investigating their DNA binding profile by chromatin immunoprecipitation and sequencing. Although MSH2 was found preferentially enriched at non-methylated CpG islands, both MSH2 and MSH6 also showed interaction with the methylated allele of the ICRs through the ZFP57/KAP1 complex.

## Results

### Identification of ZFP57-interacting proteins by LC-MS/MS analysis

Because of the poor specificity and high background detected with commercial ZFP57-specific antibodies, we used a tagged approach to identify the proteins interacting with ZFP57. We chose the Avi-Tag that can be labeled with biotin in cells expressing the bacterial biotin ligase BirA and subsequently pulled down with streptavidin [16]. This method was applied to the mouse embryonic stem cells (ESCs) E14, in which the endogenous *Zfp57* gene is highly expressed and functional [4]. Thus, BirA-expressing E14 ESCs were transfected with the tagged *Zfp57* gene (*Zfp57*-AviTag) (Fig. 1a) and the proteins interacting with the biotinylated ZFP57-AviTag were pulled down in cell lysates with streptavidin-coated beads, purified and analyzed by mass spectrometry. To exclude the proteins that could be aspecifically biotinylated by BirA, we also analyzed mock-transfected BirA-expressing E14 ESCs. Only the proteins exclusively detected in the *Zfp57*-AviTag-transfected cells (Additional file 1: Table S1a) were considered for further analyses. A schematic outline of this procedure is shown in Fig. 1b. LC-MS/MS analysis allowed us to identify 60 high-confidence ZFP57-interacting proteins (Additional file 1: Table S1a). This list includes KAP1 and HP1 $\gamma$ , which were already reported as ZFP57 interactors [4, 17], as well as several novel proteins. A gene ontology (GO) analysis allowed classifying the ZFP57 interactors into several functional categories, including mRNA processing/splicing, chromosome organization, transcription and DNA repair processes (Fig. 1c). By using the STRING database [18], we reconstructed a network model of the ZFP57-interacting proteins (Additional file 2: Fig. S1a). This approach allowed us to identify several protein families or complexes potentially interacting with ZFP57, including histone H1 variants, Heterogeneous Nuclear Ribonucleoproteins and the post-replicative DNA Mismatch-Repair (MMR) complex including MSH2, MSH6 and PCNA.



We were particularly interested in the interaction of the MSH2/MSH6 heterodimer of the MMR complex with ZFP57, because of its potential role in CpG conservation

at the ICRs. We then validated the interaction of MSH2 with the ZFP57-AviTag by co-precipitation and Western blotting (Fig. 1d). Similarly, we demonstrated that KAP1

interacts with the Avi-tagged MSH2 (Fig. 1e), that the endogenous MSH2 and KAP1 proteins interact in the wildtype E14 ESCs and that this interaction is maintained in the *Zfp57*<sup>-/-</sup> ESCs (Fig. 1f). Overall, these results confirm the interaction of the DNA mismatch repair complex with ZFP57 and indicate that this is probably mediated by KAP1.

### DNA-binding profile of MSH2

After validating the interaction between MSH2 and the ZFP57-KAP1 complex, we determined the DNA binding profile of MSH2 in mouse ESCs. We used the approach based on proteins fused to AviTag and pulldown with streptavidin. This method has been demonstrated to be efficient for chromatin immunoprecipitation and sequencing (Bio-ChIP-seq), particularly when the available antibodies raised against the endogenous protein produce high background [19]. Application of the Bio-ChIP-seq protocol to determine MSH2 binding in the E14 ESCs revealed 4444 shared peaks between two replicates (Additional file 1: Tables S2a and S2b). The genomic regions mostly enriched by MSH2 corresponded to gene promoters (52%) and CpG islands (CpGI, 51,4%) and most of the peaks overlapped transcription start sites (TSS, Fig. 2a–c and Additional file 2: Fig. S2), including the promoters of several cell growth-controlling genes, such as *Jun*, *Jund*, *Fos* and *Nras* (Additional file 2: Fig. S3). In addition, about 45% of the MSH2 peaks overlapped repetitive sequences, of which the most abundant categories were SINEs and LTRs (Fig. 2d, e).

To better characterize the relationship between the MMR proteins and the ZFP57/KAP1 complex, we analysed the distribution of MSH2 ChIP-seq reads along 1343 KAP1 ChIP-seq peaks [7] in the E14 ESCs. This analysis revealed co-localization of many MSH2 peaks with ZFP57 and KAP1 peaks (Fig. 2f). The overlap between MSH2 and KAP1 primarily involved intergenic and noncoding regions and generally no promoter CpGI (Additional file 2: Figs. S3 and S4). However, enrichment of both MSH2 and KAP1 was particularly intense at the ICRs and non-ICR ZFP57 binding sites (Additional file 2: Figs. S5 and S6). In addition, MSH2 bound many KAP1

binding sites that are not targeted by ZFP57 supporting the hypothesis that MSH2 interacts with KAP1 independently of ZFP57 (Additional file 2: Fig. S5). We also performed a ChIP-seq experiment for MSH6 in the E14 ESCs by using the antibodies raised against the native protein. The cells were treated with oxygen peroxide to increase the recruitment of MSH6 to DNA [20]. Although very few peaks could be discriminated with this antibody, the alignment of the MSH6 ChIP-seq reads along the KAP1 peaks demonstrated MSH6 enrichment on the KAP1 binding sites, confirming the results obtained with MSH2 (Fig. 2f and Additional file 2: Figs. S4 and S5).

These results demonstrate that MSH2 is mostly enriched on promoter CpGIs and repeats. Furthermore, we showed that the sequences bound by both MSH2 and MSH6 overlap those bound by ZFP57 and KAP1 in mouse ESCs.

### Interaction of MSH2 and MSH6 with the ICRs

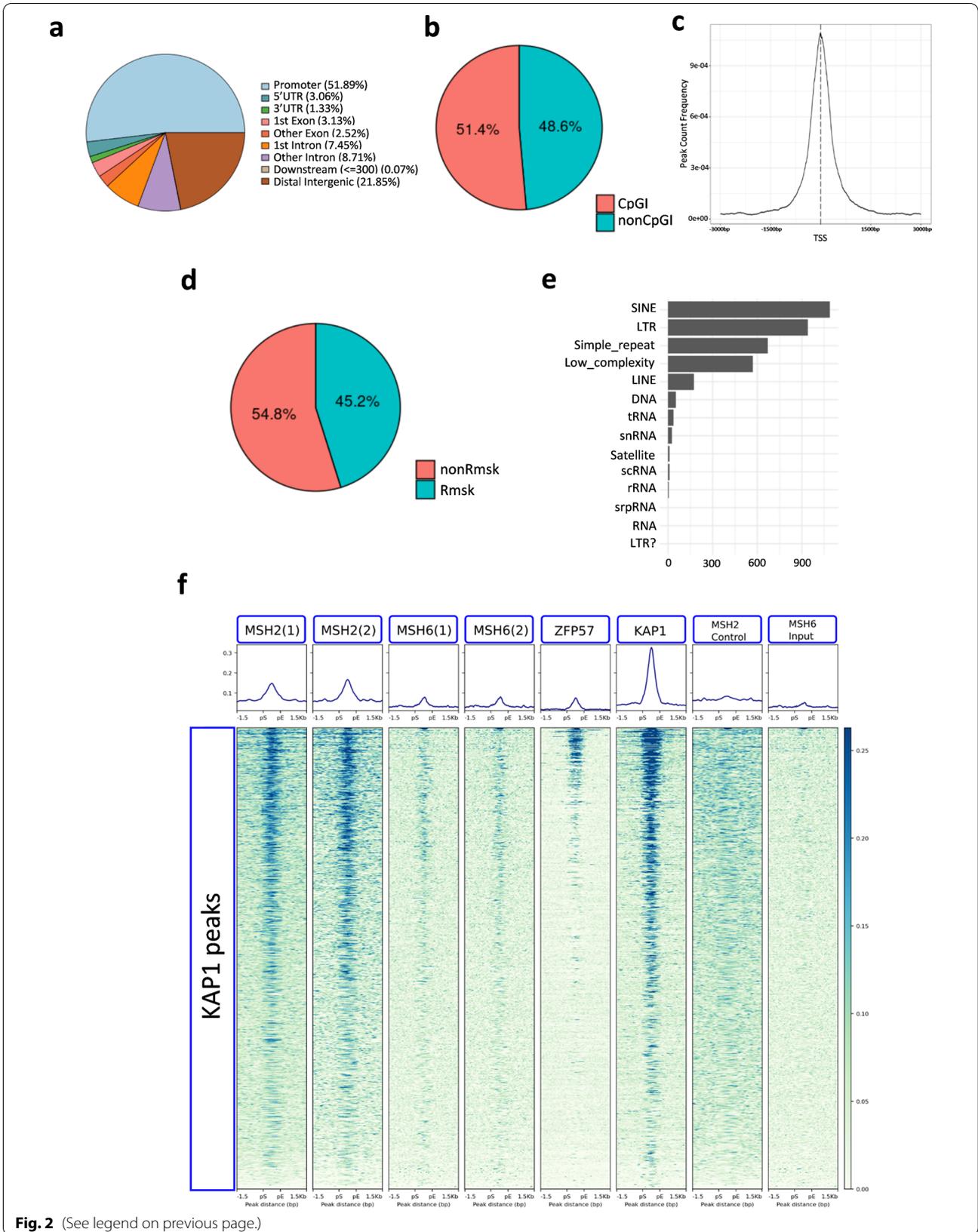
We then focused on the imprinted loci and found that MSH2 binding revealed by Bio-ChIP-seq was enriched on 10 ICRs in at least one replicate and the MSH2 peaks coincided with the ZFP57 and KAP1 peaks in these regions (Fig. 3a and Additional file 1: Table S2a, b). Because of the relatively high background observed with the MSH2 Bio-ChIP-seq, we validated the binding of MSH2 and tested that of MSH6 to the ICRs by quantitative PCR after precipitation of MSH2 by Bio-ChIP and MSH6 by conventional ChIP, respectively. The results demonstrated higher enrichment of MSH2 and MSH6 on 16 and 14 ICRs, respectively, when compared with the *Hoxa3* gene that does not overlap with CpGI (Fig. 3b) and ICR-adjacent sequences (Fig. 3c).

### MSH2 and MSH6 bind the methylated ICR allele in a *Zfp57*-dependent manner

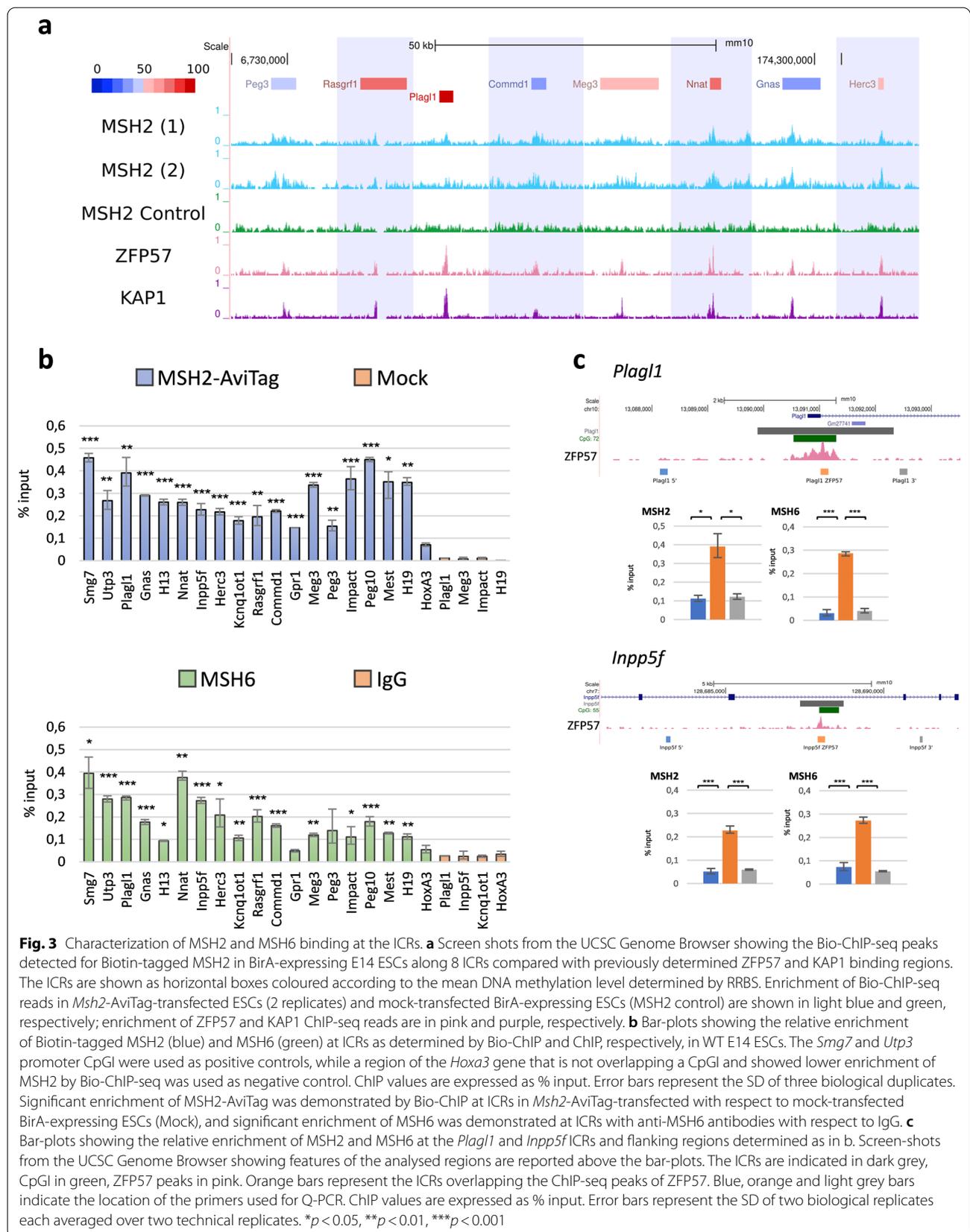
To investigate the methylation status of the sequences bound by MSH2 and MSH6, we determined the whole-genome methylation profile of the E14 ESCs by RRBS. We found that most of the non-repetitive MSH2-bound regions that could be analyzed with this method showed low methylation levels (Fig. 4a). The only exceptions were

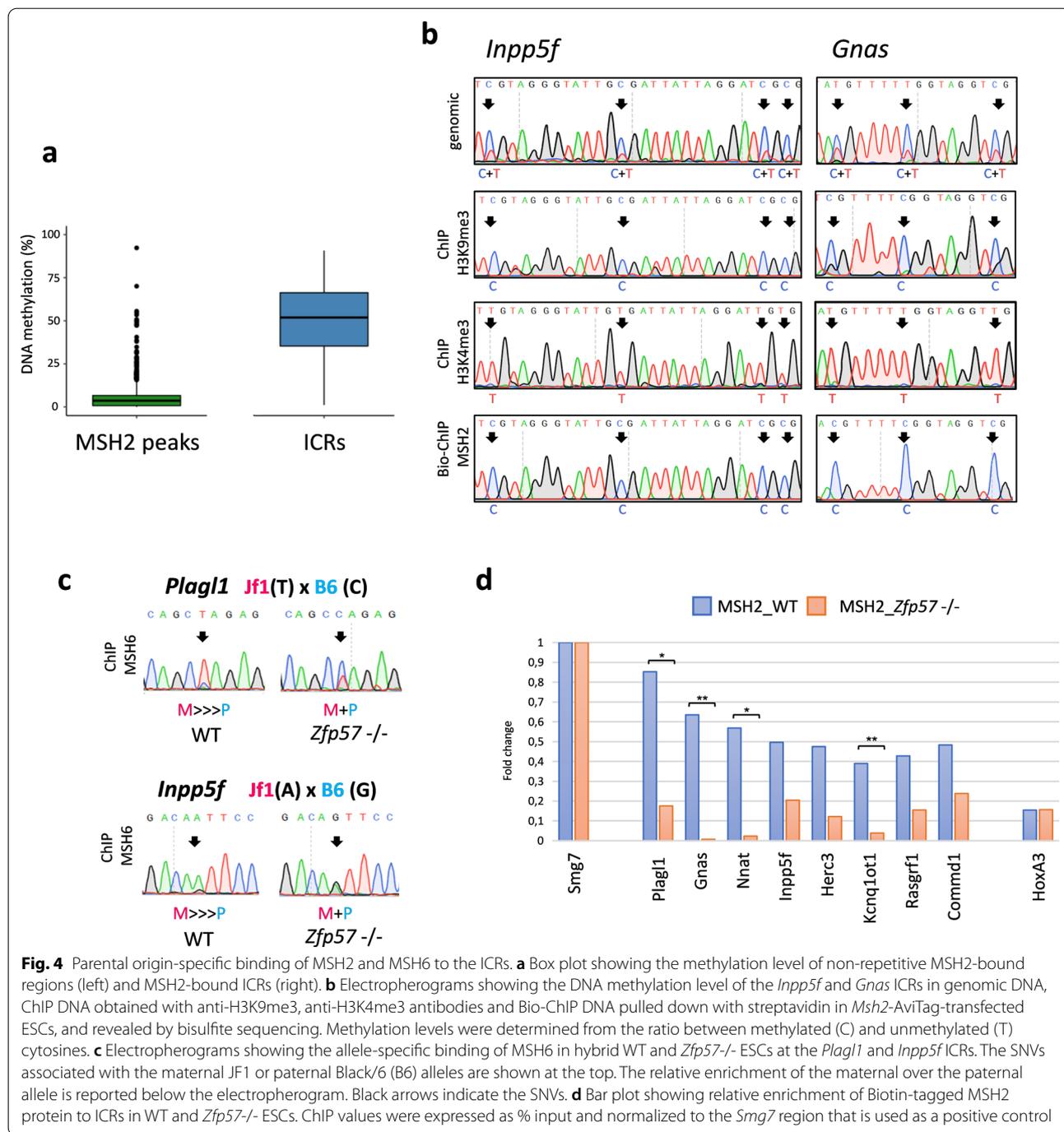
(See figure on next page.)

**Fig. 2** Genome-wide binding profiles of MSH2 and MSH6 in mouse ESCs as revealed by Bio-ChIP-seq and ChIP-seq. **a** Pie chart showing the distribution of genomic elements within the DNA regions covered by MSH2 Bio-ChIP-seq peaks. MSH2 peaks were defined as DNA regions that were enriched in the *Msh2*-AviTag cells compared to mock-transfected BirA-expressing ESCs. For feature annotation, we used the plotAnnoPie function with annoDb = "org.Mm.eg.db" parameter". **b** Pie chart representing the percentage of MSH2 Bio-ChIP-seq peaks overlapping CpG islands (CpGI) and genomic regions not including CpGI (non CpGI). **c** Average Profile of MSH2 peaks overlapping Transcription Start Sites (TSS). The plot shows the peak frequency between -3000 bp and +3000 bp from TSS. **d** Pie chart representing the percentile distribution of MSH2 peaks overlapping or not overlapping mask repetitive elements. **e** Bar plot showing the absolute frequency of repetitive elements covered by MSH2 peaks. **f** Heatmaps showing the read enrichment of MSH2 (2 replicates), MSH6 (2 replicates), ZFP57 and KAP1 in the genomic regions overlapping ( $\pm 1.5$  kbp) the KAP1 ChIP-seq peaks. As controls, we added the Bio-ChIP-seq reads of untransfected E14 ESCs (MSH2 control) and input DNA reads of the MSH6 ChIP (MSH6 input). KAP1 peaks are sorted in descending order based on the mean read enrichment value per region



**Fig. 2** (See legend on previous page.)





**Fig. 4** Parental origin-specific binding of MSH2 and MSH6 to the ICRs. **a** Box plot showing the methylation level of non-repetitive MSH2-bound regions (left) and MSH2-bound ICRs (right). **b** Electropherograms showing the DNA methylation level of the *Inpp5f* and *Gnas* ICRs in genomic DNA, CHIP DNA obtained with anti-H3K9me3, anti-H3K4me3 antibodies and Bio-CHIP DNA pulled down with streptavidin in *Msh2*-AviTag-transfected ESCs, and revealed by bisulfite sequencing. Methylation levels were determined from the ratio between methylated (C) and unmethylated (T) cytosines. **c** Electropherograms showing the allele-specific binding of MSH6 in hybrid WT and *Zfp57*<sup>-/-</sup> ESCs at the *Plagl1* and *Inpp5f* ICRs. The SNVs associated with the maternal JF1 or paternal Black/6 (B6) alleles are shown at the top. The relative enrichment of the maternal over the paternal allele is reported below the electropherogram. Black arrows indicate the SNVs. **d** Bar plot showing relative enrichment of Biotin-tagged MSH2 protein to ICRs in WT and *Zfp57*<sup>-/-</sup> ESCs. ChIP values were expressed as % input and normalized to the *Smg7* region that is used as a positive control

the ICRs that displayed an average methylation level of 50%, consistent with their imprinting status, and a few other non-imprinted ZFP57-bound loci in most of which RRBS analysis demonstrated higher methylation levels (Figs. 3a and 4a and Additional file 2: Fig. S6).

Because the ICRs have one allele methylated and the other allele non-methylated, we asked if there was any bias in the binding of MSH2 and MSH6 to these genomic

regions. To discriminate between methylated and unmethylated cytosines, we treated the MSH2-bound DNA precipitated with the Bio-CHIP protocol in the E14 ESCs with sodium bisulfite and analyzed it by Sanger sequencing. We observed that the methylated allele of the *Inpp5f* and *Gnas* ICRs was preferentially enriched in the MSH2-bound DNA (Fig. 4b). As a control, we tested the methylation status of the *Inpp5f* and *Gnas* ICR sequences

that were recovered after ChIP with anti-H3K9me3 and anti-H3K4me3 antibodies, and found that H3K9me3 and H3K4me3 were preferentially enriched on the methylated and non-methylated allele, respectively. This observation is consistent with previous results [7]. To determine if also MSH6 interacts with the DMRs in allele-specific manner, we performed a ChIP experiment in an ESC line (JB1) derived from an intra-specific mouse hybrid (JF1x C57-Black/6). In this ChIP, the maternal and paternal alleles of the ICRs can be discriminated through JF1- and C57-Black/6-specific SNVs [10]. Using this approach, we showed that MSH6 was preferentially enriched on the maternal JF1 allele of the *Plagl1* and *Inpp5f* ICRs in the JB1 ESCs (Fig. 4c). Because the *Plagl1* and *Inpp5f* ICRs are methylated on the maternal allele [10], these results confirm those obtained with MSH2.

Since the binding sites of MSH2 and MSH6 overlap those of ZFP57, we asked if the enrichment of these proteins at the ICRs depended on this zinc-finger protein. To address this issue, we first investigated MSH2 binding to the ICRs in wildtype and *Zfp57*<sup>-/-</sup> E14 ESCs by Bio-ChIP. The results demonstrated that MSH2 enrichment was significantly reduced on eight ICRs, upon losing the *Zfp57* gene (Fig. 4d). The role of ZFP57 on MSH6 binding was investigated in the JB1 ESCs by conventional ChIP. Sanger sequencing of the DNA immunoprecipitated with anti-MSH6 antibodies demonstrated that the allelic bias in MSH6 binding to the *Plagl1* and *Inpp5f* ICRs observed in the wildtype JB1 ESCs was lost in the *Zfp57*<sup>-/-</sup> JB1 ESCs (Fig. 4c).

In summary, these results demonstrate the preferential interaction of MSH2 and MSH6 with the methylated allele of the ICRs and indicate that DNA binding is mediated by ZFP57.

## Discussion

The ICRs are DNA sequences with unique properties [1, 2]. Maintenance of parental origin-dependent methylation and allele-specific expression of the imprinted genes require the presence of multiple CpGs and the binding sites of transcription factors recognizing either methylated or non-methylated DNA on the two parental alleles of these regulatory sequences. We used the ZFP57 zinc-finger protein as bait to identify the proteins interacting with the methylated allele of the ICRs in mouse ESCs. By high-resolution mass spectrometry, members of the MMR complex were identified in the ZFP57 interactome. We explored the DNA binding profile of MSH2 and MSH6 and found that they were enriched at non-methylated CpGI, but they also bound the methylated allele of the ICRs. The analysis of *Zfp57*-deficient cells allows

the conclusion that MSH2/MSH6 binding to the ICRs is mediated by ZFP57.

By employing an LC-MS/MS-based approach, we identified 60 potential ZFP57-interactors in mouse ESCs. Interestingly, all these proteins have previously been pulled down with anti-KAP1 antibodies in human ESCs and K562 cells, confirming that they are bona fide interactors of the ZFP57-KAP1 complex [21]. Furthermore, some of these interactions, such as those of the H3K9me3-interacting protein HP1 $\gamma$  and the chromatin remodelling factor CHD4 were characterized in previous studies [4, 22]. Among the novel ZFP57 interactors discovered in this study, histone H3.3 variants have been previously associated with the ICRs, while Histone H1 variants have been reported to recruit DNMTs to these regulatory regions in mouse ESCs [23]. Particularly interesting is also the finding of multiple Heterogeneous Nuclear Ribonucleoproteins including HNRNP-U, which may be involved in shaping the large-scale chromatin structures controlling parent-of-origin-dependent allelic gene expression at the imprinted loci [24, 25].

Our study identified three components (MSH2, MSH6 and PCNA) of the post-replicative MMR complex as ZFP57 interactors by LC-MS/MS analysis. The MMR pathway is essential for repairing 5mC deamination because the MSH2/MSH6 heterodimer recognizes G:T mismatches and recruits downstream proteins for correction [15, 26]. Indeed, defects of MMR genes increase the genome-wide mutation rate of methylated CpGs in cancer [15, 27]. Also, *Msh2/Msh6* knockout mice show higher cancer predisposition, microsatellite instability and mutator phenotype [28–31]. Thus, the interaction of MSH2/MSH6 with the methylated allele of the ICRs may reduce their mutation rate and preserve imprinting in rapidly dividing embryonic cells. A few methylated sequences were found among the MSH2 target sites. All of them correspond to ZFP57 binding sequences and show particularly high enrichment of MSH2 and KAP1, indicating specific MMR complex recruitment at the imprinted loci mediated by the ZFP57-KAP1 complex.

Upon oxidative damage, the MSH2-MSH6 heterodimer contributes to the recruitment of epigenetic silencing proteins, including DNMT1, SIRT1, and EZH2, to promoter CpGI in human cells [20]. This recruitment results in a transient reduction of transcription while the repair occurs. Our findings of MSH2 enrichment on TSS overlapping CpGI are consistent with these results. Interestingly, preferential MSH2/MSH6 binding to the promoters of cell growth-controlling genes suggests a safeguarding mechanism for rapidly proliferating cells. Furthermore, unlike the study performed by Ding and collaborators

[20], our cells were untreated with oxygen peroxide before the MSH2 Bio-ChIPseq experiment, suggesting that MSH2/MSH6 also recognizes CpG-rich promoters in absence of oxidative damage. However, it should be considered that the increased MSH2 concentration after gene transfer may have favored its interaction with DNA in our cells. So, it is not possible to exclude that interaction of MSH2/MSH6 with CpG-rich promoters and ICRs occurs preferentially after DNA damage. Targeting of MSH2 to promoter CpGI appears generally not mediated by ZFP57/KAP1 and its underlying mechanisms need to be elucidated in future studies. Our results are not consistent with the interaction of MSH6 with gene bodies via H3K36me3 demonstrated in human cells, suggesting the occurrence of species- or cell type-specific differences [32].

The MSH2 ChIP-seq also revealed enrichment of MSH2 at repetitive elements, mostly SINEs, LTRs, low-complexity and simple sequence repeats. Some of these interactions may be mediated by KAP1, which is known to bind repetitive sequences and contribute to genomic stability [33, 34]. It is possible that these interactions contribute to the anti-mutator and anti-recombination functions of the MMR proteins [35].

In conclusion, by determining the genomic binding profile of MSH2/MSH6, our study provides novel insights into the function of the MMR complex and reveals an unexpected role exerted through recognition of the methylated CpGs of the ICRs that may have an essential role in maintaining the integrity of these regulatory regions and in the evolution of genomic imprinting in mammalian species.

## Materials and methods

### Cell lines and culture conditions

Mouse wild-type, *Zfp57*<sup>-/-</sup> and *Msh2*<sup>-/-</sup> E14 ESCs were cultured under standard feeder-free conditions on gelatinized tissue culture dishes and maintained in DMEM (Gibco, Thermo Fisher Scientific) supplemented with 100  $\mu$ M 2-mercaptoethanol (Sigma), 1 mM sodium pyruvate, 2 mM L-glutamine, 1  $\times$  penicillin–streptomycin, 15% fetal calf serum (HyClone) and 10<sup>3</sup> U/ml leukemia inhibitory factor (LIF, Millipore). The wild-type hybrid ESC line JB1, which is (JF1  $\times$  C57BL/6) F1, and the JB1-derived *Zfp57*<sup>-/-</sup> ESC line were described previously [7, 36]. Wild-type and *Zfp57*<sup>-/-</sup> JB1 ESCs were cultured under standard feeder-free conditions on gelatinized tissue culture dishes with ESGRO Complete<sup>TM</sup> Plus Serum-free Clonal Grade 1i Medium (Merck Millipore) in the presence of 3  $\mu$ M Gsk3 inhibitor CHIR99021. For H<sub>2</sub>O<sub>2</sub> exposure, 0.5 mM H<sub>2</sub>O<sub>2</sub> was diluted in PBS immediately before adding it to the media and cells were collected

30 min later. Cells were cultured at 37 °C under an atmosphere of 5% CO<sub>2</sub>.

### Cloning and transfections

The cDNAs encoding the full-length *Zfp57* and *Msh2* genes were cloned into the expression vector containing the AviTag sequence, under the control of the elongation factor-1 alpha (*EF-1 alpha*) promoter. For plasmid transfection, cells were transfected with pEF1-BirA-V5\_His and the pEF-AviTag-Zfp57/Msh2 plasmids using Lipofectamine LTX according to the manufacturer's protocol (Thermo Fisher Scientific). Stably BirA transfected cells were selected with 0.25 mg/ml G418 (Life Technologies) and maintained in 0.1 mg/ml G418. Mock-transfected BirA-expressing E14 ESCs were used as control for MS analysis and Bio-ChIP. The primers used for cloning are listed in the Additional file 1: Table S4.

### Protein immunoprecipitation analysis

Two 10 cm dishes of cells were pelleted and resuspended in NP40 buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.5% NP-40) on ice and incubated for 20 min at 4 °C. Isolated nuclei were lysed in SDS lysis buffer (50 mM Tris–HCl pH 8.0, 10 mM EDTA and 1% SDS) for 10 min at 4 °C. After incubation at 95 °C for 10 min, the nuclei lysate was sonicated on ice and centrifuged for 10 min at maximum speed. 1 mg of proteins (for KAP1, MSH6 and IgG IP) were pre-cleared with 30  $\mu$ l protein A/G agarose beads (SantaCruz) for 2 h at 4 °C on a rotating wheel. Anti-KAP1 antibody (Abcam ab10483), anti-MSH6 (Santa Cruz sc-137015) or mouse IgG were added to the pre-cleared lysate and incubated overnight at 4 °C on a rotating wheel. Proteins were precipitated with 50  $\mu$ l protein A/G agarose beads for 1 h at 4 °C with rotation. For biotin tagged-ZFP57 and -MSH2 immunoprecipitation, 1 mg of proteins was incubated with 100  $\mu$ l of streptavidin beads overnight at 4 °C on a rotating wheel. The agarose and streptavidin beads were then washed five times with 500  $\mu$ l RIPA buffer (10 mM Tris–HCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 1% Triton X-100 and 0.1% SDS, protease inhibitor (cOmplete, EDTA-free, Roche Life Science). Beads elution was followed by Mass Spectrometry or Western Blot analyses. High resolution nano-LC–tandem mass spectrometry and MS data processing were carried out as previously reported [37]. Proteins that were pulled down by streptavidin and identified by mass spectrometry in mock-transfected BirA-expressing E14 ESCs (Additional file 1: Table S1b) were excluded from further analysis. The list of ZFP57 interactors identified by LC–MS/MS was imported into the STRING

database [18]. Cluster analysis was performed using the “kmean clustering” option with “number of clusters” parameter set to 4. The edges between clusters were set as “Dotted line”. To explore the processes in which those proteins are involved, we performed a GO:BP enrichment analysis using the function `gost` of `gProfiler2` R package v 0.2.1 [38] with default options. We selected all the terms involved in either DNA or RNA processes with an adjusted  $p$ -values  $< 0.05$ . The selected terms were plotted using the `GOplot` package v 1.0.2 [39].

### Western blotting

Proteins were eluted from beads by incubating them at 95 °C for 10 min in Laemmli buffer and resolved by 8% acrylamide gel. Samples were transferred onto PVDF membranes (Biorad Transblot). After 1 h of blocking in 5% w/v milk/TBST at RT, membranes were incubated with the primary antibodies anti-ZFP57 (Abcam ab45341), anti-KAP1 (Abcam ab10483), anti-MSH6 antibody (BD Biosciences, 610,918), anti-MSH2 (Calbiochem, NA27) and anti-Actin (Sigma-Aldrich, A2066) overnight at 4 °C. The membranes were washed 3 × with TBST and incubated with the secondary antibody. Signals were visualized using an ECL method.

### Chromatin immunoprecipitation (ChIP)

ChIP for the analysis of biotin tagged-MSH2, MSH6, H3K9me3 and H3K4me3 binding, was performed on formaldehyde cross-linked chromatin isolated from cells grown on 10 cm dishes to ~80% confluency. Briefly, the cells were detached by adding 0.05% trypsin at 37 °C for 3 min. Formaldehyde was added to approximately  $3 \times 10^7$  cells resuspended in Phosphate Buffered Saline (PBS) at final concentration of 1% and the cells were incubated at room temperature for 10 min with shaking. The reaction was stopped by the addition of glycine to a final concentration of 0.125 M. Cells were washed twice in ice-cold PBS, centrifuged and resuspended in lysis buffer 1 (50 mM HEPES pH 8, 10 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40 and 0.25% Triton X-100) for 10 min at 4 °C. Isolated nuclei were lysed in lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA) for 10 min at 4 °C. The chromatin was sheared in a sonication buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate and 0.5% N-lauroylsarcosine) to an average size of 100–400 bp using the S220 Focused-ultrasonicator (Covaris). Sonicated chromatin was diluted in sonication buffer. Chromatin for MSH6, H3K9me3 and H3K4me3 ChIP was pre-cleared with 30 µl protein A/G agarose beads (SantaCruz) for 4 h at 4 °C on a rotating wheel. Streptavidin beads (Dyna-beads MyOne streptavidin T1, ref 65,601, Invitrogen)

and anti-MSH6 (Santa Cruz sc-137015), anti-H3K9me3 (Abcam 8898), anti-H3K4me3 (Abcam ab8580) antibodies or rabbit/mouse IgG were added to the pre-cleared chromatin and incubated overnight at 4 °C on a rotating wheel. Chromatin for MSH6, H3K9me3 and H3K4me3 ChIP was precipitated with 30 µl protein A/G agarose beads for 4 h at 4 °C with rotation. The beads were then washed five times with 500 µl RIPA buffer (10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 1% Triton X-100 and 0.1% SDS) and once with each of the following buffers: WASH buffer (50 mM HEPES, 0.5% sodium deoxycholate, 1% Triton X-100, 1 mM EDTA, 500 mM NaCl and 0.2% NaN<sub>3</sub>), LiCl buffer (0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA and 10 mM Tris pH 8) and TE buffer (10 mM Tris pH 8, 1 mM EDTA). The bound chromatin was eluted in 100 µl TE buffer. Cross-links were reversed by incubation at 1 h at 37 °C with 1 µl RNase cocktail (Ambion) and O/N at 60 °C after addition of 2.5 µl 20% SDS and 2.5 µl 20 mg/ml proteinase K (Sigma). The DNA was extracted by using the QIAquick Gel Extraction Kit (Qiagen). The immunoprecipitated or 1% input DNAs were analysed by real-time PCR using SBYR Green PCR Master Mix (Bio-Rad) on a CFXCONNECT Thermal Cycler (Bio-Rad). Each reaction was performed in biological duplicate with each duplicate averaged over at least two technical replicates. To evaluate significance, we applied an unpaired Student's T-test. Asterisks indicate statistical significant differences in enrichment of DNA sequences with adjusted  $p$ -value: \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Primers are listed in Additional file 1: Table S4.

### Allele-specific ChIP

MSH2 binding as well as histone H3K9me3 and H3K4me3 enrichment on the methylated and non-methylated alleles of *Inpp5f* and *Gnas* loci was assessed after direct sequencing of the bisulfite-treated and PCR-amplified immunoprecipitated DNA. MSH6 binding on the JF1 and B6 alleles of *Plagl1*, and *Inpp5f* was determined by typing the immunoprecipitated DNA for the SNPs present between the two parental genomes. The amplification products were sequenced (Eurofins Genomics) and the ratio between allele-specific DNAs was determined from the electropherogram. All the primers are listed in the Additional file 1: Table S4.

### NGS analysis

For the RRBS library preparation, we used 100–200 ng of genomic DNA according to Illumina's instructions. Libraries were generated and sequenced at IGA Technology Services (Italy), by using the NuGEN Ovation RRBS Methyl-Seq Library System and paired-end

150 bp sequencing mode on NovaSeq6000 (Illumina, San Diego, CA). After adapter trimming and quality step using TrimGalore v. 0.6.6 [<https://github.com/FelixKrueger/TrimGalore>] with the RRBS mode (`-rrbs`), we aligned the reads against the *Mus musculus* genome (mm10) using Bismark v.0.23.0 with default parameters [40]. We removed the duplicate reads using UMI-tools v.1.1.1 with default parameters [41] and then used Bismark methylation extractor to obtain the methylation level of the CpGs covered.

For ChIP-seq analysis, two nanograms of DNA from immunoprecipitated and input chromatin were used for Illumina library preparation. Libraries were generated and sequenced at IGA Technology Services (Italy), by using the NuGen Ovation Ultralow Library System v2 Kit and 150 bp paired-end sequencing on the Illumina NovaSeq6000 platform (Illumina, San Diego, CA). After checking that there were no bad quality base calls and adapter contaminations in the raw data, we aligned the reads against the *Mus musculus* genome (mm10) using Bowtie2 short read aligner v2.3.5.1 with default parameters [42]. We removed duplicate reads using Picard MarkDuplicates v2.22.9 [<http://broadinstitute.github.io/picard/>] and the multiple mapping reads, and we used only uniquely mapped reads for the rest of the study.

The DNA binding profiles of ZFP57 and KAP1 in the E14 ESCs were obtained from the GSE77744 dataset. The coordinates were converted into the mm10 genome by using CrossMap Python script [43]. For visualization in UCSC, we used tracks normalized by reads per million (RPM) generated by the GenomeCoverageBed tools of the BEDtools suite v2.29.2 [44]. To define the enriched regions, we used MACS2 algorithm with PE and `-broad` parameters [45]. We performed the intersection of the peaks using Bedtools intersect function and evaluated its statistical significance using the function `enrichPeakOverlap` of the Bioconductor package ChIPseeker with `nShuffle = 1000`. The MSH2 peaks shared by the two replicates were 33% of the replicate 1 peaks and 62% of the replicate 2 peaks. The overlap (4444 peaks) was considered statistically significant. ( $p = 0.0001$ ). We annotated the peaks using the ChIPseeker package [46] to the TxDb.Mmusculus.UCSC.mm10.knownGene database [47]. Using the `plotAvgProf` function with default parameters, we plotted the peaks frequency profile near the TSS. Moreover, for feature annotation, we used the `plotAnnoPie` function with `annoDb = "org.Mm.eg.db"` parameter. The statistical significance of the overlap between MSH2 peaks and promoters ( $-1000 + 0$  of TSS) was calculated using the `enrichPeakOverlap` function and found to be highly significant ( $p = 0.0001$ ). Concerning

annotation of the CpG islands and Rmsk regions, we intersected the peaks with the coordinates of those regions downloaded from UCSC. We plotted the heatmap of scores associated with genomic regions using the `computeMatrix` and `plotHeatmap` function of the deepTools v3.4.3 [48]. The raw and processed files are deposited in GEO under the accession number GSE205043. Publicly available datasets of KAP1 and ZFP57 ChIPseq were sourced from GEO database: GSE77444.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13072-022-00462-7>.

**Additional file 1: Table S1.** List of high-confidence proteins pulled down with ZFP57-AviTag in E14 ESCs and identified by nano-LC-MS/MS. **Table S1b.** List of shared proteins pulled down with streptavidin in E14 BirA (controls) and E14 BirA+ZFP57\_AviTag (samples). **Table S2a.** MSH2 peaks identified by Bio-ChIP-seq in E14 ESCs. Replicate 1. **Table S2b.** MSH2 peaks identified by Bio-ChIP-seq in E14 ESCs. Replicate 2. **Table S3.** Coordinates of the imprinted gDMRs in mouse ESCs\*. **Table S4.** Primer sequences.

**Additional file 2: Fig. S1.** STRING network model of ZFP57-interacting proteins. 41 of the 60 candidate ZFP57-interacting proteins were mapped on an interconnected network constructed by STRING analysis. The model revealed key sub-network clusters connected to ZFP57 and KAP1 (TIF1B). Colours represent different subnetworks based on K-means clustering. Edge thickness is representative of the confidence in interaction based on database mining, experimental evidence and text mining. **Fig. S2.** Heatmap showing that the great majority of the MSH2 Bio-ChIP-seq peaks overlapping with promoters are centered on Transcription Start Sites (TSS). **Fig. S3.** Screenshots from the UCSC Genome Browser showing the ChIP-seq signals detected for biotin-tagged MSH2 in BirA-expressing E14 ESCs along eight cell growth-controlling genes with promoters overlapping CpGI. DNA methylation and binding profiles of MSH2 (2 replicates), ZFP57 and KAP1 are reported as in Figure 3a. **Fig. S4.** (related to Figure 2f) Heatmaps showing the read enrichment of MSH2, MSH6, ZFP57 and KAP1 in the genomic regions overlapping ( $\pm 1.5$  kbp) the KAP1 ChIP-seq peaks sorted on the basis of their overlap with various genomic elements. **Fig. S5.** (related to Figure 2f) Heatmaps showing the read enrichment of MSH2, MSH6, ZFP57 and KAP1 in the genomic regions overlapping ( $\pm 1.5$  kbp) the KAP1 ChIP-seq peaks sorted on the basis of their overlap with ZFP57 peaks. **Fig. S6.** Screenshots from the UCSC Genome Browser showing the ChIP-seq signals detected for the Biotin-tagged MSH2 in BirA-expressing E14 ESCs along five *noICR* regions bound by ZFP57 [7]. DNA methylation and binding profiles of MSH2 (2 replicates), ZFP57 and KAP1 are reported as in Figure 3a.

## Acknowledgements

We thank Salvatore Oliviero for the AviTag vector and protocol of Bio-ChIP-seq.

## Author contributions

Conceptualization, AR and BA; methodology, BA, FC, CG, AV, RR, MV; validation, FC, CG and BHM; formal analysis, BA and FC; data curation, BA and FC; writing, review and editing, BA, FC and AR; supervision, CA, AC and AR; funding acquisition, AR. All authors read and approved the final manuscript.

## Funding

The work was supported by the following grants: Associazione Italiana Ricerca sul Cancro IG 2020 N. 24405, "Progetti per la ricerca oncologica della Regione Campania" Grant: I-Cure, "Progetti competitivi intraAteneo" Programma V:ALERE (VANviteLLi pER la RicERca) 2019 – grant MIRIAM from Università degli Studi della Campania "Luigi Vanvitelli".

**Availability of data and materials**

Raw data supporting the findings of this study have been deposited under accession code GSE205043 in the Gene Expression Omnibus repository.

**Declarations****Consent for publication**

Consent for publication of clinical and molecular data were obtained from all participants involved in this study.

**Competing interests**

The authors declare that they have no potential competing interests to disclose. The authors declare that they have no financial relationships relevant to this article to disclose.

**Author details**

<sup>1</sup>Department of Environmental Biological and Pharmaceutical Sciences and Technologies (DiSTABIF), Università Degli Studi Della Campania "Luigi Vanvitelli", 81100 Caserta, Italy. <sup>2</sup>Institute of Genetics and Biophysics (IGB) "Adriano Buzzati-Traverso", Consiglio Nazionale Delle Ricerche (CNR), 80131 Naples, Italy. <sup>3</sup>Department of Biology, Università Degli Studi Di Napoli "Federico II", 80126 Naples, Italy. <sup>4</sup>Istituto Per Le Applicazioni Del Calcolo "Mauro Picone" (IAC), CNR, 80131 Naples, Italy.

Received: 13 June 2022 Accepted: 25 July 2022

Published online: 02 August 2022

**References**

- Barlow DP, Bartolomei MS. Genomic imprinting in mammals. *Cold Spring Harb Perspect Biol.* 2014;6: a018382.
- Monk D, Mackay DJG, Eggermann T, Maher ER, Riccio A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat Rev Genet.* 2019;20:235–48.
- 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;15:547–57.
- Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, Trono D. In embryonic stem cells, *ZFP57/KAP1* recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell.* 2011;44:361–72.
- 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.
- Anvar Z, Cammisa M, Riso V, Baglivo I, Kukreja H, Sparago A, Girardot M, Lad S, De Feis I, Cerrato F, Angelini C, Feil R, Pedone PV, Grimaldi G, Riccio A. *ZFP57* recognizes multiple and closely spaced sequence motif variants to maintain repressive epigenetic marks in mouse embryonic stem cells. *Nucleic Acids Res.* 2016;44:1118–32.
- Riso V, Cammisa M, Kukreja H, Anvar Z, Verde G, Sparago A, Acurzio B, Lad S, Lonardo E, Sankar A, Helin K, Feil R, Fico A, Angelini C, Grimaldi G, Riccio A. *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;44:8165–78.
- Shi H, Strogantsev R, Takahashi N, Kazachenka A, Lorincz MC, Hemberger M, Ferguson-Smith AC. *ZFP57* regulation of transposable elements and gene expression within and beyond imprinted domains. *Epigenetics Chromatin.* 2019;12:49.
- Jiang W, Shi J, Zhao J, Wang Q, Cong D, Chen F, Zhang Y, Liu Y, Zhao J, Chen Q, Gu L, Zhou W, Wang C, Fang Z, Geng S, Xie W, Chen LN, Yang Y, Bai Y, Lin H, Li X. *ZFP57* dictates allelic expression switch of target imprinted genes. *Proc Natl Acad Sci USA.* 2021;118: e2005377118.
- Acurzio B, Verma A, Polito A, Giaccari C, Cecere F, Fioriniello S, Della Ragione F, Fico A, Cerrato F, Angelini C, Feil R, Riccio A. *Zfp57* inactivation illustrates the role of ICR methylation in imprinted gene expression during neural differentiation of mouse ESCs. *Sci Rep.* 2021;11:13802.
- Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP, Hahnemann JM, Kordonouri O, Masoud AF, Oestergaard E, Storr J, Ellard S, Hattersley AT, Robinson DO, Temple IK. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in *ZFP57*. *Nat Genet.* 2008;40:949–51.
- Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. *Nat Rev Genet.* 2011;12:756–66.
- Kim J. Evolution patterns of *Peg3* and *H19-ICR*. *Genomics.* 2019;111:1713–9.
- Takahashi N, Coluccio A, Thorball CW, Planet E, Shi H, Offner S, Turelli P, Imbeault M, Ferguson-Smith AC, Trono D. *ZNF445* is a primary regulator of genomic imprinting. *Genes Dev.* 2019;33:49–54.
- Fang H, Zhu X, Yang H, Oh J, Barbour JA, Wong JWH. Deficiency of replication-independent DNA mismatch repair drives a 5-methylcytosine deamination mutational signature in cancer. *Sci Adv.* 2021;7:eabg4398.
- Kim J, Cantor AB, Orkin SH, Wang J. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nat Protoc.* 2009;4:506–17.
- Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G, Li X. Zinc finger protein *ZFP57* requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J Biol Chem.* 2012;287:2107–18.
- Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 2009;37:D412–6.
- Krepelova A, Neri F, Maldotti M, Rapelli S, Oliviero S. Myc and max genome-wide binding sites analysis links the Myc regulatory network with the polycomb and the core pluripotency networks in mouse embryonic stem cells. *PLoS ONE.* 2014;9: e88933.
- Ding N, Bonham EM, Hannon BE, Amick TR, Baylin SB, O'Hagan HM. Mismatch repair proteins recruit DNA methyltransferase 1 to sites of oxidative DNA damage. *J Mol Cell Biol.* 2016;8:244–54.
- Jang SM, Kauzlaric A, Quivy JP, Pontis J, Rauwel B, Coluccio A, Offner S, Duc J, Turelli P, Almouzni G, Trono D. *KAP1* facilitates reinstatement of heterochromatin after DNA replication. *Nucleic Acids Res.* 2018;46:8788–802.
- Hoffmeister H, Fuchs A, Erdel F, Pinz S, Gröbner-Ferreira R, Bruckmann A, Deutzmann R, Schwartz U, Maldonado R, Huber C, Dendorfer AS, Rippe K, Längst G. *CHD3* and *CHD4* form distinct NuRD complexes with different yet overlapping functionality. *Nucleic Acids Res.* 2017;45:10534–54.
- Yang SM, Kim BJ, Norwood Toro L, Skoultschi AI. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc Natl Acad Sci USA.* 2013;110:1708–13.
- Michieletto D, Gilbert N. Role of nuclear RNA in regulating chromatin structure and transcription. *Curr Opin Cell Biol.* 2019;58:120–5.
- Grin I, Ishchenko AA. An interplay of the base excision repair and mismatch repair pathways in active DNA demethylation. *Nucleic Acids Res.* 2016;44:3713–27.
- Noordermeer D, Feil R. Differential 3D chromatin organization and gene activity in genomic imprinting. *Curr Opin Genet Dev.* 2020;61:17–24.
- Poulos RC, Olivier J, Wong JWH. The interaction between cytosine methylation and processes of DNA replication and repair shape the mutational landscape of cancer genomes. *Nucleic Acids Res.* 2017;45:7786–95.
- Wei K, Kucherlapati R, Edelmann W. Mouse models for human DNA mismatch-repair gene defects. *Trends Mol Med.* 2002;8:346–53.
- de Wind N, Dekker M, Berns A, Radman M, te Riele H. Inactivation of the mouse *Msh2* gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell.* 1995;82:321–30.
- Reitmair AH, Schmits R, Ewel A, Bapat B, Redston M, Mitri A, Waterhouse P, Mittrücker HW, Wakeham A, Liu B, et al. *MSH2* deficient mice are viable and susceptible to lymphoid tumours. *Nat Genet.* 1995;11:64–70.
- Kolodner RD, Marsischky GT. Eukaryotic DNA mismatch repair. *Curr Opin Genet Dev.* 1999;9:89–96.
- Huang Y, Gu L, Li GM. *H3K36me3*-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J Biol Chem.* 2018;293:7811–23.
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB,

- Trono D. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*. 2010;463:237–40.
34. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev Cell*. 2016;36:611–23.
  35. Harfe BD, Jinks-Robertson S. DNA mismatch repair and genetic instability. *Annu Rev Genet*. 2000;34:359–99.
  36. Kota SK, Llères D, Bouschet T, Hirasawa R, Marchand A, Begon-Pescia C, Sanli I, Arnaud P, Journot L, Girardot M, Feil R. ICR noncoding RNA expression controls imprinting and DNA replication at the Dlk1-Dio3 domain. *Dev Cell*. 2014;31:19–33.
  37. Marino MM, Rega C, Russo R, Valletta M, Gentile MT, Esposito S, Baglivo I, De Feis I, Angelini C, Xiao T, Felsenfeld G, Chambery A, Pedone PV. Interactome mapping defines BRG1, a component of the SWI/SNF chromatin remodeling complex, as a new partner of the transcriptional regulator CTCF. *J Biol Chem*. 2019;294:861–73.
  38. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. Gprofiler2—an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res*. 2020;9:ELIXIR-709.
  39. Walter W, Sánchez-Cabo F, Ricote M. GOrbit: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015;31:2912–4.
  40. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27:1571–2.
  41. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. 2017;27:491–9.
  42. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
  43. Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*. 2014;30:1006–7.
  44. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
  45. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:1–9.
  46. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382–3.
  47. Bioconductor Core Team and Bioconductor Package Maintainer TxDb. *Mmusculus.UCSC.mm10.knownGene*: Annotation package for TxDb object(s). R package version 3.10.0. 2019.
  48. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44:W160–5.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

