Epigenetics & Chromatin

RESEARCH

Open Access

Dissecting the Kaiso binding profile in clear renal cancer cells

Alexey Starshin[1], Pavel Abramov[1], Yaroslava Lobanova[1], Fedor Sharko[1], Galina Filonova[1], Dmitry Kaluzhny[3], Daria Kaplun[1], Igor Deyev[4], Alexander Mazur[1], Egor Prokhortchou[1,2] and Svetlana Zhenilo[1*]

Abstract

Background There has been a notable increase in interest in the transcriptional regulator Kaiso, which has been linked to the regulation of clonal hematopoiesis, myelodysplastic syndrome, and tumorigenesis. Nevertheless, there are no consistent data on the binding sites of Kaiso in vivo in the genome. Previous ChIP-seq analyses for Kaiso contradicted the accumulated data of Kaiso binding sites obtained in vitro. Here, we studied this discrepancy by characterizing the distribution profile of Kaiso binding sites in Caki-1 cells using Kaiso-deficient cells as a negative control, and compared its pattern on chromatin with that in lymphoblastoid cell lines.

Results We employed Caki-1 kidney carcinoma cells and their derivative, which lacks the Kaiso gene, as a model system to identify the genomic targets of Kaiso. The principal binding motifs for Kaiso are CGCG and CTGCNAT, with 60% of all binding sites containing both sequences. The significance of methyl-DNA binding activity was confirmed through examination of the genomic distribution of the E535A mutant variant of Kaiso, which cannot bind methylated DNA in vitro but is able to interact with CTGCNA sequences. Our findings indicate that Kaiso is present at CpG islands with a preference for methylated ones. We identified Kaiso target genes whose methylation and transcription are dependent on its expression. Furthermore, Kaiso binding sites are enriched at CpG islands, with partial methylation at the 5′ and/or 3′ boundaries. We discovered CpG islands exhibiting wave-like methylation patterns, with Kaiso detected in the majority of these areas. Similar data were obtained in other cell lines.

Conclusion The present study delineates the genomic distribution of Kaiso in cancer cells, confirming its role as a factor with a complex mode of DNA binding and a strong association with CpG islands, particularly with methylated and eroded CpG islands, revealing a new potential Kaiso target gene—SQSTM1, involved in differentiation of acute myeloid leukemia cells. Furthermore, we discovered the existence of a new class of CpG islands characterized by wave-like DNA methylation.

Keywords DNA methylation, Methyl-DNA binding proteins, Kaiso, CpG islands erosion, SQSTM1

article note = online

The original online version of this article was revised: Author name "Egor Prokhortchou" has been updated.

*Correspondence:
Svetlana Zhenilo
szhenilo@gmail.com
Full list of author information is available at the end of the article

Introduction

DNA methylation and histone modifications are crucial regulatory mechanisms influencing gene transcription, X inactivation, imprinting, and chromatin architecture formation. Methylation of promoters is associated with transcriptional repression, especially in promoters containing CpG islands. The methylation status of CpG-poor promoters does not always indicate the active or repressed state of a gene; however, it establishes a specific profile enabling various transcription factors to regulate gene transcription [1, 2]. The same is true for regions adjacent to CpG islands shores and shelves, where methylation dynamics play a role in regulating tissue-specific gene expression and interactions between promoters and enhancers [3–6]. In recent years, a multitude of factors that interact with methylated DNA, often in conjunction with their primary consensus sequences, have been discovered. Many of these factors serve as transcriptional activators, with some also capable of influencing DNA methylation levels by recruiting dioxygenases or DNA methyltransferases [7–10]. These factors represent a novel class of transcriptional regulators that offer precise control over gene expression. Kaiso (ZBTB33) is a notable example of such a factor. It is a member of the BTB/POZ family and recognizes methylated DNA through zinc fingers located at the C-terminus. Hydroxymethylation inhibits the binding of Kaiso to DNA [11, 12]. Kaiso contains the BTB/POZ domain at the N-terminus, which interacts with the corepressors NcoR and SMRT, attracting histone deacetylases and leading to transcriptional repression [13–15]. Furthermore, Kaiso functions as a transcriptional activator [16, 17]. Notably, Kaiso also interacts with a specific DNA sequence known as the Kaiso binding site (KBS), which has the consensus sequence CTGCNA. Kaiso forms a complex with DNMT3a/3b and regulates the methylation of the ICR at the H19/IGF2 locus in human cells, as well as the promoter of Oct4 in MEFs [18, 19]. Like other methyl-DNA binding proteins, the knockout of Kaiso resulted in behavioral abnormalities in mice [20]. Additionally, Kaiso has been implicated in the development of inflammatory processes in the colon and in cellular oncogenesis and may be involved in clonal hematopoiesis [21–23]. There is currently limited knowledge about the role of Kaiso in kidney cancer. Kaiso has been identified as one of the most significantly activated transcription factors during a specific stage of oncotransformation in kidney cells [24]. Furthermore, Kaiso has been recognized as a protective factor for patients diagnosed with renal clear cell carcinoma [25]. Reports that determine Kaiso binding sites in vivo are clearly controversial. This is in part due to the quality of the Kaiso-specific antibodies. The recruitment of Kaiso to methylated DNA or KBS regulates several target genes [26, 27]. Previous ChIP-seq analyses of Kaiso contradicted the accumulated data concerning Kaiso binding sites obtained in vitro [28]. It revealed that the majority of Kaiso binding sites are located in the promoters of active genes that contain unmethylated CpG islands and are not enriched by KBS [28]. The antibodies used for ChIP-seq by Blatter et al. targeted the N-terminal region of the protein, encompassing the BTB/POZ domain shared by various BTB/POZ-containing proteins. To address this discrepancy, we conducted ChIP-seq analysis using antibodies specific to a peptide in the linker region. We characterized the Kaiso binding site profile in the human clear cell renal cell carcinoma Caki-1 cell line. Previously obtained Caki-1 cells with a frameshift mutation in the Kaiso coding sequence were employed in the ChIP-seq experiment as a control [8]. This methodology enables the exclusion of any nonspecific binding sites that might arise from antibody limitations. Furthermore, we determined Kaiso binding sites at the genome-wide level and elucidated alterations in their methylation patterns upon Kaiso removal. We have also demonstrated that the E535A point mutation in Kaiso's zinc finger domain allows us to selectively disrupt its interaction with methylated DNA while retaining its ability to bind to KBS in vitro [29, 30]. This approach facilitated the exploration of the relative importance of the DNA-binding functions of Kaiso within the cellular context. To integrate the findings from our study, we undertook a comprehensive analysis of the available ChIP-seq experiments from the ENCODE database, specifically targeting breast cancer cell line and two lymphoblastoid cell lines. The results of this study provide valuable insights into the functional implications of the interaction of Kaiso with methylated DNA and KBS in the context of cellular processes.

Materials and methods

Cell lines

The following cell lines were used: the human clear cell renal carcinoma cell line Caki-1 (ATCC Caki-1-HTB-46), previously obtained Kaiso-deficient Caki-1 cells, the human embryonic kidney cell line HEK293, and previously obtained Kaiso-deficient HEK293 cells [8, 17]. All cell lines were grown in Dulbecco's modified Eagle's medium (ThermoFisher Scientific) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, and 2 mM L-glutamine.

ChIP-seq

Chromatin immunoprecipitation was performed in duplicate via a magnetic kit (ab156907, Abcam) according to the manufacturer's instructions. Immunoprecipitation was performed with anti-Kaiso antibodies (Sigma

HPA005732) approved by ENCODE (ENCAB292USO). Libraries were constructed via the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB) and sequenced with a single read with a read length of 50 bp. The number of reads was adjusted to achieve peak saturation. Reads with a quality of less than 20 were filtered via Cutadapt and truncated to 50 nucleotides. The remaining reads were mapped to the human genome (GRCh37/hg19) via Bowtie2 with the 'end-to-end' option. Peak calling was performed via MACS2 [31] with the -q 0.01 parameter. Intersection and subtraction-related manipulations with peaks were performed via bedtools [32]. All subsequent analyses were carried out at the intersection of biological replicates. To minimize the number of false positive peaks, we subtracted Kaiso-knockout control peaks from those obtained from the wild type. The results of the ChIP-seq experiments for the cell lines MCF-7 (https://doi.org/10.17989/ENCSR231YFE), GM12878 (https://doi.org/10.17989/ENCSR542FLV), and K562 (https://doi.org/10.17989/ENCSR876GXA) were obtained from the ENCODE database. The peak sets obtained from two replicates, processed with the optimal IDR threshold filter and bigWig files, were derived from the ENCODE hg19 pipeline outputs.

DNA methylation analysis

Whole-genome methylation data for Caki-1 and Kaiso-deficient cells were obtained from GSE151787 and PRJNA734133, respectively [8, 33, 34]. The reads were aligned to the GRCh37/hg19 genome assembly via Bismark software [35]. The bisulfite conversion efficiency (>99%) was assessed using both lambda phage and non-CpG context methylation. Individual differential CpGs were identified via the beta-binomial regression approach implemented in the MethPipe pipeline [36]. The results of the WGBS experiments for the cell lines GM12878 (https://doi.org/10.17989/ENCSR890UQO), and K562 (https://doi.org/10.17989/ENCSR765JPC) were obtained from the ENCODE database. Then, they were processed by the steps described above. The level of DNA methylation in regions overlapping with Kaiso peaks was assessed as follows: Kaiso peaks were overlapped with annotations obtained from the R package AnnotationHub, and the distribution of CpG methylation levels for each region was then plotted.

The coverage of Kaiso binding regions with consensus sequences

The coverage of Kaiso binding regions with consensus sequences was analyzed within a range of ± 4 kb from the center of the Kaiso peak. The whole-genome distribution of the consensus sequences was generated via a Python script in the same way as that reported by Isagulieva et al. [37]. The Bedtools coverage function with nucleotide resolution was used to calculate the number of overlapping Kaiso peaks with consensus sequences genome-wide or for specific regions (intergenic, promoter, intron, exon, and intron–exon boundaries). For the enrichment of TCTCGCGAGA, a position weight matrix was downloaded from JASPAR (JASPAR ID: MA0527.1). The probability of finding a consensus sequence in the Kaiso region was calculated as the ratio of the number of overlaps to the total number of Kaiso peaks.

Data analysis

Data processing and alignment were performed via in-house computational pipelines. Statistical analyses were conducted with R statistical software. The distribution of peaks across genomic elements, CpG islands, shores, and shelves was performed using the ChIPseeker software. To calculate Kaiso enrichment around TSS (transcription start sites) and TES (transcription end sites), we considered all genes that intersect with Kaiso peaks. Using deepTools with the parameters scale-regions -m 3000 -b 5000 -a 5000, we generated a plot (with TSS marking the beginning of the gene and TES marking the end). Here, a and b represent the distances after TSS/TES, while m indicates the length of the region being normalized. The Bedtools coverage function was employed to determine the number of overlapping Kaiso peaks with CpG islands. The plot was created using R's ggplot2 package.

To generate CpG observed/expected heatmap we used awk to process an input BED file, calculating the center of each peak and adjusting it with the flanking distance of 5000 bp. Each line produced a new entry with chromosome, start coordinate, end coordinate, and a unique peak identifier. The output from awk is piped into bedtools makewindows, which divides the flanking regions into smaller windows of 300 bp size. Then, we used bedtools getfasta to extract the nucleotide sequences corresponding to the generated windows. The resulting file was read as pandas DataFrame and the observed-to-expected ratio of CpG sites were calculated with formula.

$$\text{Obs/Exp CpG} = \text{Number of CpG} * 300 / (\text{Number of C} * \text{Number of G}).$$

Observed to expected ratio of CpG sites in windows with GC content less than 0,5 was set to 0. Ratios in each window were smoothed with rolling means. Visualization was made with seaborn heatmap function.

Motif discovery by homer

We used the findMotifsGenome script of Homer [38] (http://homer.ucsd.edu/homer/motif/) tool for de novo motif identification. Fragment sizes were set according to the length of the Kaiso peaks, and other parameters were set to their defaults.

Methylation density and Kaiso reads enrichment

We normalized Kaizo reads to the input and calculated fold enrichment (FE) in a 50-base window using deepTools [39]. We then referenced the FE to exon, intron, intergenic, and promoter coordinates, calculating the average enrichment of these regions, including those 3000 bp from the center of each region. The clustered data were then visualized as a heatmap using DeepTools.

We selected only those bisulfite-treated reads that did not contain unmethylated CpGs and those that covered CpGs with a methylation level of 1. We then converted the BAM files to BigWig format and plotted the distributions of methylated reads in the selected genomic regions, as described above.

Computational definition of CpG island DNA methylation patterns

CpG islands were downloaded from the UCSC table browser. Promoter CpG islands were then defined by overlapping the CpG islands with transcription start sites obtained from the hg19 Ensembl annotation file. For normalization purposes, we split the promoter CpGIs into 40 bins. We then used the ScoreMatrixBin function from the R generation [40] package to calculate the average methylation of each bin. To obtain major groups on the basis of the methylation patterns of WT Caki-1 cells, we applied k-means clustering. Using this method, we identified clusters with high and low methylation: cluster 1—high methylation: median 0.8018, mean 0.7821, 1st Quartile 0.6939, 3rd Quartile 0.8792; cluster 2—low methylation: median 0.007103, mean 0.040291, 1st Quartile 0.002176, 3rd Quartile 0.064305. To obtain groups with 5' and 3' methylation erosion for regions not included in clusters 1 and 2, we calculated Pearson's r between the bin number and the respective average methylation level. This analysis was conducted at three different levels: 1) for all bins, 2) for the first and last halves of the bins separately, and 3) for all quarters of the bins separately. CpG islands with r values less than -0.5 for the first bin group at any level were defined as having 5' erosion. CpG islands with r values greater than 0.5 for the last bin group of any level were defined as having 3' erosion. If an island has negative and positive r values for the first and last bin groups of any level, it belongs to the category of 5'-3' erosion. Islands that do not fall under the above criteria but still have r values less than -0.5 and/or greater than 0.5 in at least one of the third-level bin groups were defined as wave shaped.

EMSA

Binding was performed via Kaiso without the BTB/POZ domain (ΔBTB Kaiso-GST, amino acids 117–692) purified from transformed BL21 *E. coli* via GST-Sepharose (Glutathione Sepharose 4B, Cytiva, UK). Probes were obtained by annealing methylated and unmethylated oligonucleotides labeled with FAM: 5′-FAM-GCAGC(meC)G(meC)GCCCAA(meC)GCTGGG AGATC-3' and met-SM 5′-TCCCCAG(meC)GTTGGG (meC)G(meC)GGCTGCGATC-3' (methylated probe M+, unmethylated M-); 5′-FAM-GTGTGCTTCCTG CCAACGATGA-3' and 5'-ACATCGTTATTGGCAGGA AGCACAC-3' (KBS probe); and 5'-FAM-AGAAGCCTC GCTGGGAAACAAGGAATCGGCGGG-3' and 5'-CCC GCCGATTCCTTGTTTCCCAGCGAGGCTTCT-3' (KBS-like probe). The binding reaction was performed via the LightShift EMSA Optimization and Control Kit (20148X) (Thermo Fisher Scientific). The mixture containing oligonucleotide probes and increasing amounts of ΔBTB Kaiso-GST protein was incubated for 30 min at 25 °C. The DNA–protein complex was loaded onto 5% PAAG (0.5X TBE), and the resolved complex was subsequently detected via Typhoon Trio+.

Differentially methylated regions (DMR)

Differentially methylated regions were identified utilizing the MethyLasso program, with a minimum CpG coverage threshold established at 4 [41].

Gene ontology enrichment analysis

Gene Ontology (GO) analysis was performed via ShinyGO 0.81 bioinformatics resources with an FDR < 0.05 [42, 43].

CRISPR/Cas9-based gene editing

The pX459-E535-ZBTB33 plasmid was obtained as follows: ligation of the BbsI-digested pSPCas9(BB)-2A-Puro (PX459) plasmid (Addgene #48139) [44] with an annealed sgRNA oligo insert: Crisp_E535_for 5′-CACCGTCCG TGCCGTTACTGTGAGA, and Crisp_E535_rev 5′-AAA CTCTCACAGTAACGGCACGGAC. The plasmid was verified by Sanger sequencing analysis. CRISPR/Cas9–based editing was performed as previously described [44]. Briefly, Caki-1 cells were seeded on a 12-well plate. The cells were transfected with the PX459-E535-ZBTB33 plasmid and ssODN5'- AGA AGA AGT ATC CGT GCC GTT ACT GTG AGA AGG TAT TTC CGC TAG CAG CAT ATC GCA CAA AGC ATG AAA TTC ATC ACA CAG GGG AGC GAA via Lipofectamine 3000 to generate the E535A mutation. Puromycin was added after

24 h at a concentration of 2 μg per ml. The cells were incubated for 72 h and passaged into a 96-well plate at a density of 1 cell per 2 wells in medium without puromycin. The mutation was confirmed by Sanger sequencing of the corresponding amplicons obtained from PCR with genomic DNA.

Results

Landscape of Kaiso-binding regions in human renal cancer cells

To investigate the DNA-binding activity of Kaiso, we performed chromatin immunoprecipitation with anti-Kaiso antibodies in Caki-1 human kidney cancer cells and Kaiso-deficient Caki-1 cells generated via CRISPR/Cas9 editing. These Kaiso-deficient cells were previously generated by introducing a frameshift mutation at lysine 42 in the BTB/POZ domain of Kaiso [8]. Western blot analysis of total cell lysates confirmed the depletion of Kaiso in these cells (Fig. 1a). Our ChIP-seq analysis revealed 23,153 Kaiso binding peaks in Caki-1 cells. After eliminating false positive peaks from Kaiso-deficient cells, we obtained 21,144 high-confidence Kaiso-interacting regions (Supplementary Fig. 1). Compared with a random set of genomic regions, these Kaiso-binding regions were distributed as follows: 32% were found in introns, 33% in promoters (< 3 kb), and 27% in intergenic regions (Fig. 1b). Kaiso binding was enriched upstream of the transcription start site (TSS) and downstream of the transcription end site (TES) up to 2 kb (Fig. 1c, d—example of Kaiso binding at the promoter (top) and near the TES (bottom)). Interestingly, 42% of the Kaiso binding sites overlapped with CpG islands or were located within 4 kb upstream and downstream of the CpG islands in regions known as shores and shelves (Fig. 1c, e, f). Shores are defined as the 2000 base pairs flanking CpG islands, whereas shelves are the regions farther up to 4000 base pairs (Fig. 1f, right panel). When we examined various genomic features, we found that in exons, exon–intron boundaries, and promoters, Kaiso binding preferentially occurred at CpG islands and shores (Fig. 1g). In contrast, in introns and intergenic regions, Kaiso binding was more common in CpG-poor regions. Among the 1167 Kaiso-bound promoters, 1031 contained CpG islands or shores, whereas 136 were CpG-poor (Fig. 1g).

Kaiso regulates expression of genes involved in mRNA processing

Next, we analyzed genes whose promoters were annotated to Kaiso binding regions and simultaneously exhibited changes in their transcription levels. Analysis of RNA-seq and ChIP-seq data revealed that 204 genes, whose promoters are associated with Kaiso binding, undergo transcriptional changes in Kaiso-deficient cells (padj < 0.05), of which 141 genes were upregulated, and 63 genes were downregulated in Kaiso deficient cells (Supplementary Fig. 2a). KEGG pathway analysis and Gene Ontology demonstrated that the upregulated genes are significantly enriched in pathways related to RNA processing and splicing factors (Supplementary Fig. 2b).

Identification of DNA binding consensus sequences in the Kaiso interacting regions

Kaiso binds in vitro to CTGCNA and DNA with single or multiple methylated CpGs [11, 26, 27]. Our results revealed that Kaiso binding regions are enriched in KBS and contain multiple CpGs compared with random regions (Fig. 2a, b). We observed enrichment of CGCG in the central part of the Kaiso-binding region (Fig. 2c), which is consistent with previous in vitro studies showing the high affinity of Kaiso for sequences containing methylated CGCG tetranucleotides [15, 26]. The known Kaiso binding motif from JASPAR (TCTCGCGAGA) was also detected at the central part of the peak but at a lower abundance (Fig. 2c). Although the KBS sequence is frequently found in the genome, there is a clear enrichment of KBS in the Kaiso-binding regions compared with the genome-wide distribution (Fig. 2b, c). Many researchers consider the longer variant of the Kaiso binding site (KBS) (TCCTGCNA) to be a consensus sequences [27, 45]. However, J. Daniel's work showed that the nucleotides T and C at the 5' end of this sequence don't actually play a role in Kaiso's recognition of DNA[27]. Our study yielded comparable results, as we did not detect any enrichment associated with this extended sequence (Fig. 2c). We analyzed whether the distribution of KBS and CGCG in Kaiso-binding regions depended on the type of genomic region in which this peak was located. Enrichment of KBS in Kaiso-binding regions was detected in introns and intergenic regions, whereas CGCG-containing peaks were abundant in all regions

(See figure on next page.)

Fig. 1 Kaiso binding landscape **A** western blot analysis of total cell lysates from Caki-1 and Kaiso deficient Caki-1, *-SUMOylated form of Kaiso protein. Distribution of Kaiso binding regions across the **B** genome by ChIPseeker, **C** around TSS, TES (by deepTools), and CpG islands (intersections were obtained using bedtools, then plot was constructed by ggplot2), **D** example of genomic tracks for Kaiso ChIP-seq data from Caki-1 and Kaiso KO Caki-1 cells, and input data,arrows indicate direction of transcription; **E** visualization of CpG islands in Kaiso binding regions via observed/expected CpG ratio (see materials and methods), **F** the overlay between Kaiso binding sites and CpG islands, shores, shelves, and **G** genomic regions

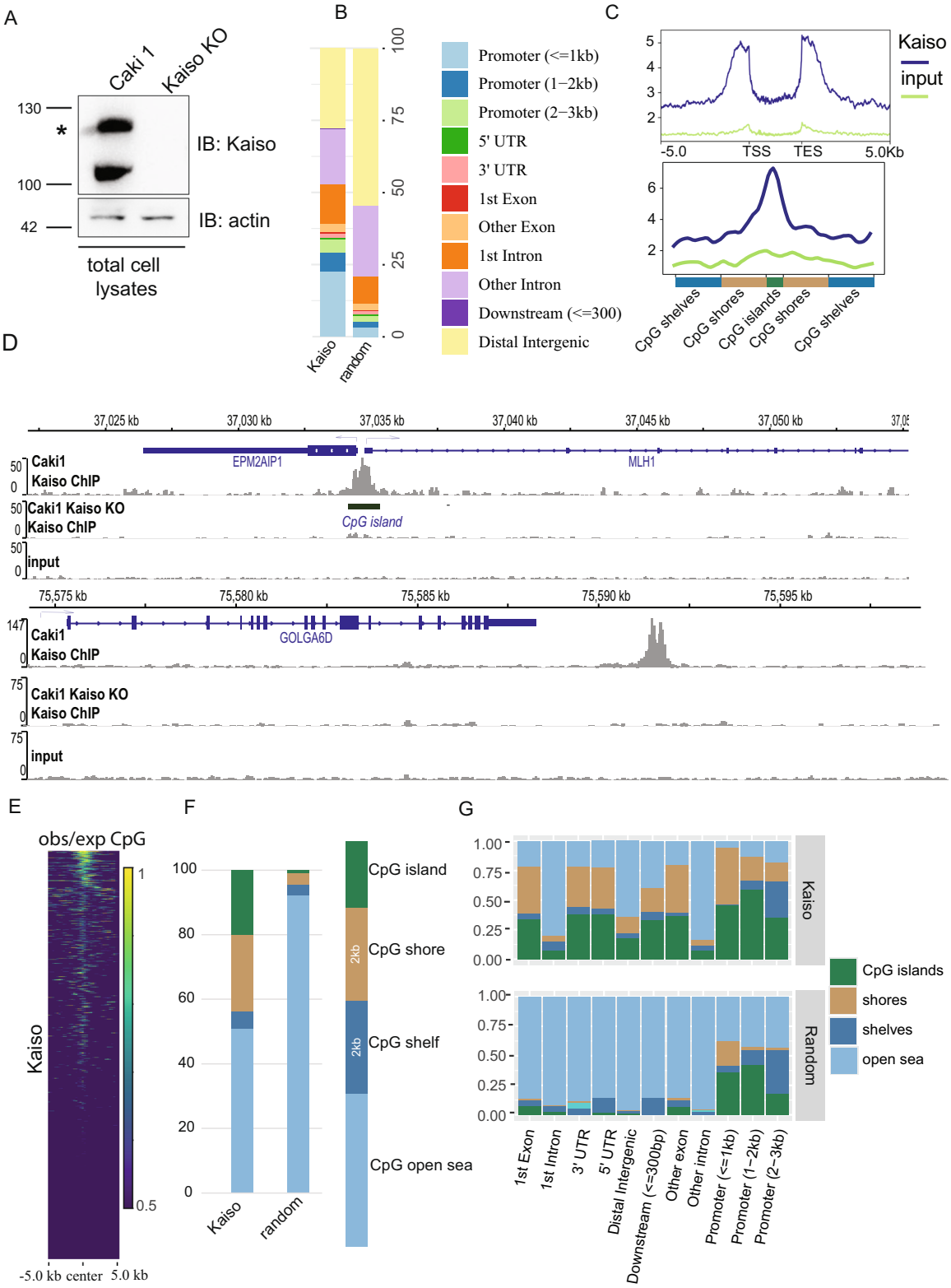
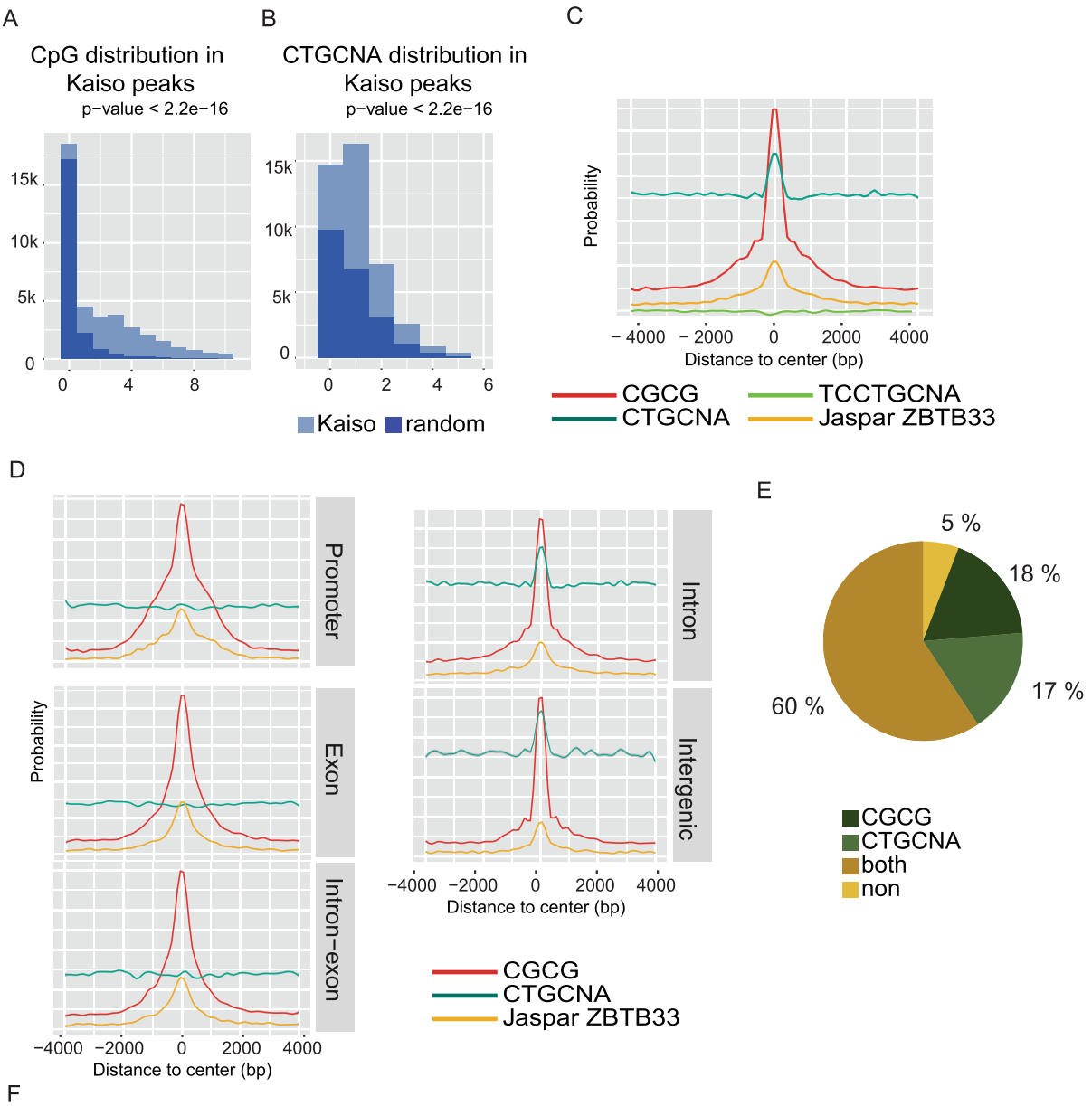


Fig. 1 (See legend on previous page.)



Rank Motif	P-value	log P-pvalue	% of targets	% of background
PROMOTER	1e-271	-6.248e+02	22.12%	4.06%
EXON	1e-128	-2.962e+02	57.62%	37.54%
INTRON	1e-1894	-4.362e+03	38.89%	8.27%
EXON-INTRON BOUNDARIES	1e-95	-2.197e+02	63.76%	43.53%
INTERGENIC	1e-1373	-3.164e+03	41.14%	5.60%

Fig. 2 Characterization of Kaiso binding regions. The distribution of CG (**A**), KBS (**B**), in Kaiso binding region via a stacked bar chart. Each bar is comprised of a Kaiso binding (light blue) and random regions (dark blue). Enrichment of indicated motifs in Kaiso binding regions for all peaks (**C**) and peaks sorted according to distinct genomic elements (**D**). **E** A circular diagram of CGCG and KBS presence in Kaiso binding regions. **F** De novo Homer analysis of sequences in Kaiso binding regions located in promoters, exons, exon–intron boundaries, introns and intergenic parts

(Fig. 2d). Overall, 60% of the Kaiso peaks contained both CGCG and KBS sequences, 17% contained only KBS, and 18% contained only CGCG. Five percent of the Kaiso binding sites contained neither CGCG nor KBS (Fig. 2e).

Next, we performed de novo motif analysis via Homer software for each category of Kaiso-binding regions. Motifs containing a single CpG were found for all genomic categories of Kaiso binding sites (Fig. 2f, Supplementary Fig. 3). Among the top motifs in the promoters, exon–intron boundaries, and exons, CGCG sequences were detected (Supplementary Fig. 3). The KBS site was found among the top DNA motifs in introns and intergenic regions, which is consistent with our analyses of KBS enrichment in Kaiso-binding regions. KBS was also detected in Kaiso peaks intersecting with exon–intron boundaries (Supplementary Fig. 2). Regions without CGCG and KBS sequences were enriched for motifs containing a single CpG (Supplementary Fig. 3).

In addition, we performed de novo motif analysis via MEME software (Supplementary Fig. 4). We observed that Kaiso-binding regions were enriched with T-rich sequences, CTGGGA, and CTGTAN. The last two contain sequences that are similar to the KBS site (CTGCNA). These KBS-like sequences were enriched in introns and intergenic regions, similar to KBS. Unlike KBS, they were found in promoters (Supplementary Fig. 5). We performed an EMSA experiment to test whether Kaiso can interact with the KBS-like sequence, using the methylated probe as a positive control. We were unable to detect DNA–protein complex formation with probes containing the KBS-like sequences CTGGGA and CTGTAN, whereas Kaiso effectively binds methylated and KBS-containing probes (Supplementary Fig. 6).

Consequently, those Kaiso binding sites that are located in intergenic regions and introns are characterized by the presence of both KBS and CGCG, whereas those binding sites that resemble promoters, exons, and intron–exon boundaries are predominantly enriched for CGCG alone.

Methylation-dependent localization of Kaiso

To characterize the Kaiso binding regions further, we analyzed the DNA methylation levels at the enriched loci via a dataset for genome-wide DNA methylation in Caki-1 cells [34]. Kaiso peaks were overlapped with genomic regions, and the distribution of CpG methylation levels for each region was then plotted. Our analysis revealed that Kaiso is predominantly localized to methylated regions (Fig. 3a). The Kaiso-binding sites exhibiting the highest methylation levels are predominantly located in introns and intergenic regions. In contrast, in promoters, exons, and exon–intron boundaries, Kaiso binding sites include a mix of hypo- and hypermethylated CpGs (Fig. 3a). We quantified the relationship between Kaiso binding and DNA methylation by calculating the methylation density for each 400 bp window, combining the methylation level with the density of the CpG dinucleotides. Methylation density reflects the local concentration of methylated CpGs as substrates for methyl-DNA binding proteins. Our analysis revealed that Kaiso enrichment is moderately positively correlated with methylation density (Pearson's $r = 0.28$) (Fig. 3b).

Further, we assessed the methylation status of Kaiso-binding regions by calculating fully methylated tag densities within these regions and extending 3 kb upstream and downstream from the center of the Kaiso peaks. The majority of Kaiso-binding regions were found to contain methylated DNA, with a more diffuse methylation density observed in promoters and exons within these regions (green panel) (Fig. 3c). Subsequently, we compared the enrichment levels of Kaiso in relation to the methylation status of CpG islands. The heatmap illustrates that Kaiso is more pronouncedly enriched in methylated CpG islands compared to unmethylated ones (Fig. 3d). Next, we evaluated the quantitative relationship of Kaiso-binding regions based on their methylation levels and genomic locations. The Kaiso-binding regions that intersected with CpG islands were sorted according to their methylation density. The distribution patterns indicate that highly methylated Kaiso-binding regions in intergenic regions constitute approximately 75% of the total, while those located within gene bodies account for 60%, and those in promoters represent 40% (see Supplementary Fig. 7).

The methylation levels of Kaiso binding sites in the shores and shelves were notably high (Fig. 3e). By adding RNA-seq data to our analysis, we found that Kaiso interacts with both the methylated and unmethylated promoters of active and inactive genes, respectively (Fig. 3f). Additionally, we analyzed Kaiso-bound regions that contained the regions predicted by the Homer motifs.

(See figure on next page.)

Fig. 3 Kaiso are targeted to highly methylated regions. DNA methylation in Kaiso binding peaks throughout **A** the various genomic regions, **E** CpG islands, shores, and shelves; **B** genome-wide correlation between Kaiso enrichment (blue) and methylation density, calculated at 400 bp windows ranked by methylation density (red line) CpG islands; **C** heatmap of signal density clustering on Kaiso peaks (3 kb up and downstream the center of Kaiso peaks, red) for Kaiso and fully methylated reads level (green), each line represents Kaiso binding region; Kaiso enrichment **D** in dependence of CpG island methylation status, and **F** around active and inactive TSS

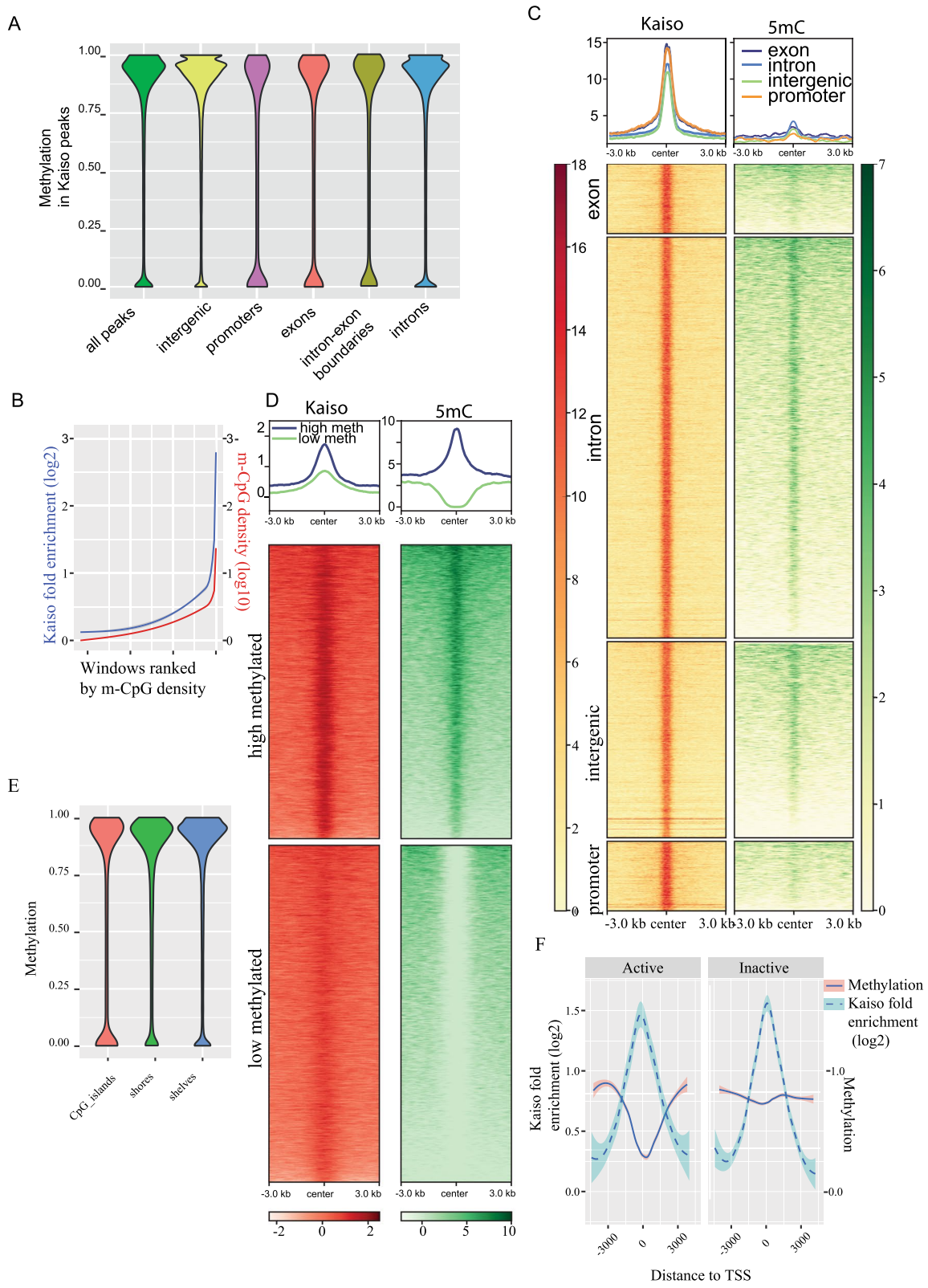


Fig. 3 (See legend on previous page.)

Notably, the degree of methylation around these regions was also quite high (Supplementary Fig. 8a, b).

Consequently, Kaiso clearly prefers highly methylated CpG-rich regions. However, dense methylation is not a prerequisite for its binding preferences, as is evident in the case of promoters.

In light of previous observations by D. Kaplun et al. [8], we sought to ascertain whether the knockdown of Kaiso might influence DNA methylation within its binding regions. We plotted the distribution of single CpG methylation levels within whole genome and various genomic elements (Supplementary Fig. 8c), as well as in Kaiso-bound regions in Caki-1 and Kaiso-KO cells (Fig. 4a). In Kaiso-deficient Caki-1 cells, a moderate increase in methylation was observed both across the genome and in various genomic elements, as previously reported in study (Supplementary Fig. 8c) [8]. In Kaiso binding regions, we noted a similar trend in methylation changes in Kaiso-deficient cells as observed on a genome-wide scale—a slight increase in DNA methylation (Fig. 4a). An exception to this trend was found in binding regions located within promoters, exons, and at exon–intron boundaries, where we detected a small decrease in methylation (Fig. 4a). When comparing the difference in methylation (delta methylation) across all Kaiso binding sites against randomly selected regions where Kaiso was not localized, no statistically significant difference was found (Supplementary Fig. 9a).

The analysis of the difference in DNA methylation within Kaiso-bound regions revealed very few regions in genes promoters: hypermethylated *TINCR, ETS2, RAB20, PLXNA3, KCNH2, HBQ1, UNCX, MKX, FBXO44, GABRE, and TMEM178B* and hypomethylated *HAND1, FTX, WNT1, PPP1R14A, PCDH7, FAM83G, CYP2E1, NDUFA4l2, IQgap2, TSC22D1, FAM255a, and FAM255b* in Kaiso-deficient cells exhibited a methylation difference of at least 20% (Fig. 4b and 4c and Supplementary Fig. 9b,c). The effect of Kaiso removal on the transcription of these genes was minimal. This is attributed to the fact that certain genes are not transcribed (WNT1) or have alternative promoters (PCDH7, FAM83G, PPP1R14A, IQgap2), and changes in methylation affect an inactive promoter, which does not lead to its reactivation. FTX, FAM225a and FAM225b are noncoding RNAs that, owing to the absence of polyadenylation, were not detected in the transcriptome analysis. Protocadherin 7 (PCDH7) and FAM83G were upregulated in Kaiso-deficient cells, whereas transcription of Hand1 was decreased. Some of the promoters of hypomethylated genes (Hand1, CYP2E1, NDUFA4l2, PPP1R14A, TSC22D1) that interact with Kaiso exhibit a moderate level of methylation, characteristic of monoallelically expressed genes.

We also conducted an additional analysis to investigate more localized changes in DNA methylation within Kaiso binding regions using DMR (Differential Methylation Region) analysis. Utilizing the latest DMR detection method, MethyLasso [41], we identified approximately 4,500 DMRs, of which 1,223 were found in the promoter regions. Notably, 273 of these DMRs intersected with Kaiso binding regions, with the majority located in promoters (Fig. 5a). Both hypo- and hypermethylated DMRs were found in Kaiso binding regions, most frequently situated in CpG islands and shores (Fig. 5a-c). Of these, 15 DMRs were located in the promoter regions of genes that exhibited changes in transcription (Fig. 5c). Hypomethylated DMRs were detected in 4 out of the 7 upregulated genes (SQSTM1, NACC2, RTL6, DEGS1), and these DMRs were situated in the downstream regions relative to the transcription start site (TSS). For the genes that were downregulated in Kaiso-deficient cells, we predominantly observed hypermethylation of DMRs located upstream of the TSS (Fig. 5c). Thus, we were able to identify regions characterized by both decreases and increases in DNA methylation.

In summary, we concluded that the alterations in DNA methylation within Kaiso-bound regions in Caki-1 cells, following the depletion of Kaiso, are predominantly similar to the changes observed in regions not associated with Kaiso, that Kaiso-dependent DMR are prominently located in promoters. Also, our analysis revealed a set of genes whose promoter methylation is dependent on Kaiso binding.

Kaiso is enriched at CpG islands with DNA methylation erosion

Given that Kaiso was predominantly detected within CpG islands in promoter regions, whether methylated or unmethylated, we conducted a more detailed analysis of the methylation profiles of these regions. Methylation patterns of bound and unbound Kaiso CpG islands were compared. K-means clustering suggested that the majority of the CpG islands in the Kaiso unbound promoters were unmethylated, with a very small fraction showing hypermethylation. Additionally, partial methylation at the 5' and/or 3' ends of the CpG islands was observed in 10% of the cases for the promoters (refer to Fig. 6a). Approximately 6% of the CpG islands presented intermediate methylation levels. A distinct methylation profile was observed for the CpG islands bound by Kaiso. We noted a significant increase in the number of methylated CpG islands, with notably fewer unmethylated CpG islands than those not bound by Kaiso. Furthermore, a substantial portion of the CpG islands displayed partial methylation at the 5', 3', or both ends, was detected (Fig. 6b). This consistent pattern of methylation in Kaiso-bound CpG

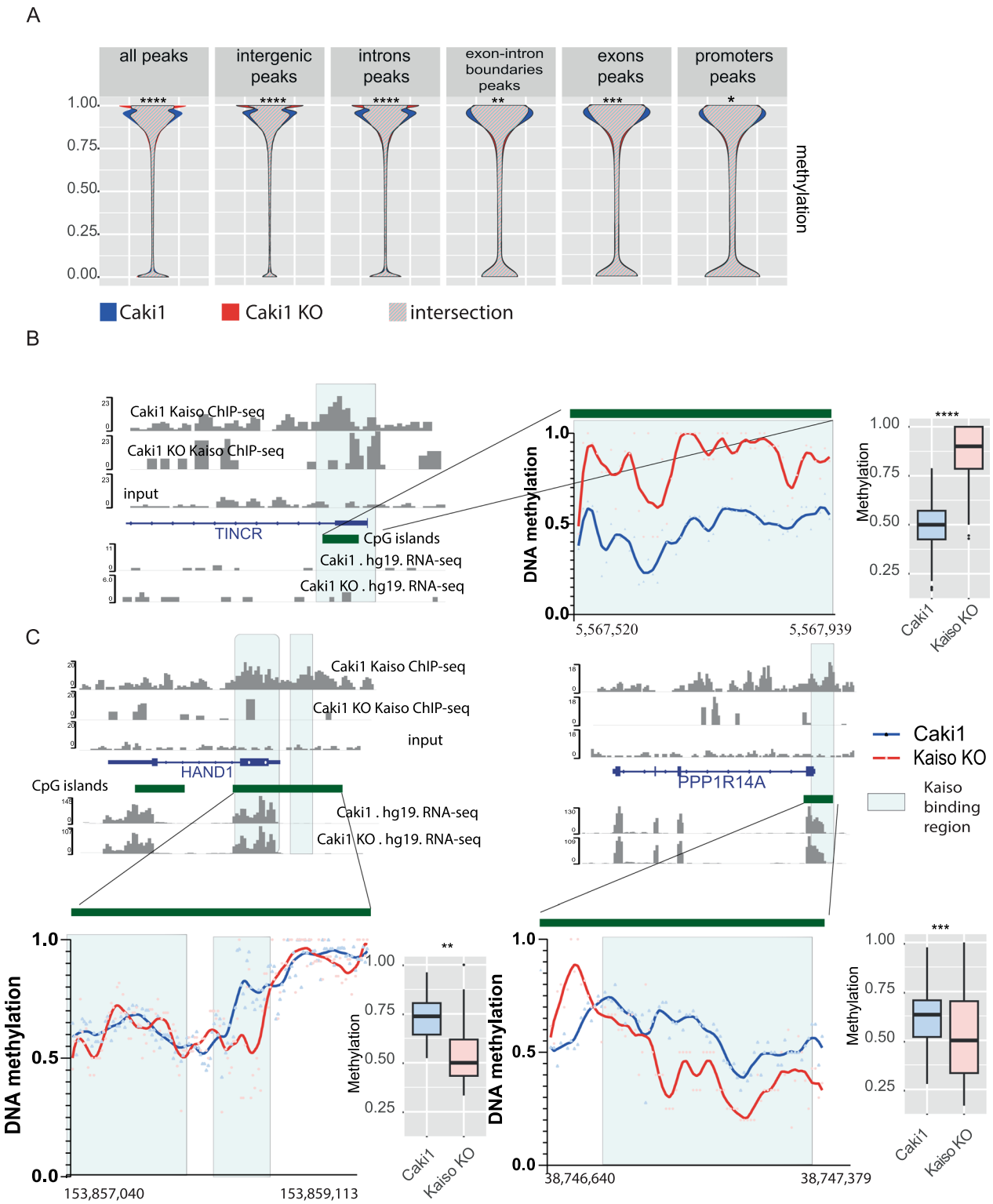
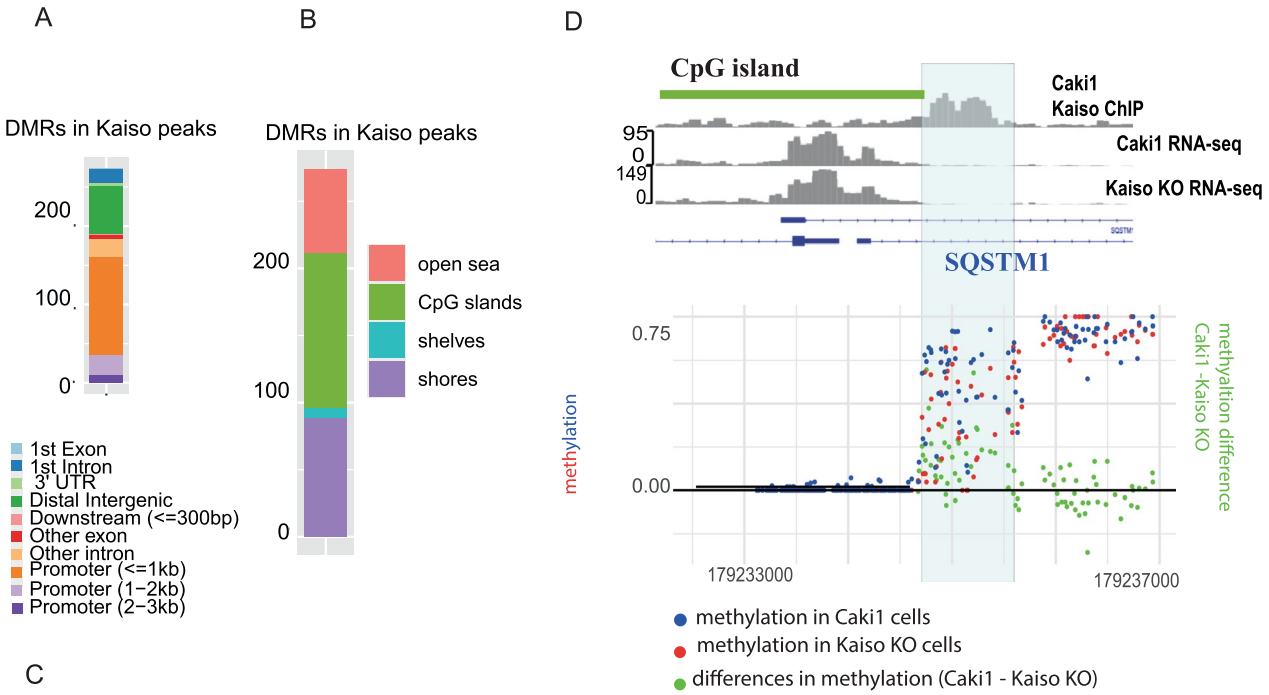


Fig. 4 DNA methylation changes in Kaiso deficient cells in Kaiso binding regions. **A** Distribution of DNA methylation across different genomic features in Caki-1 and Kaiso deficient cells in Kaiso binding regions. P-value of DNA methylation difference was determined by Wilcoxon signed-rank test, red-methylation in Caki-1, blue-methylation in Kaiso deficient cells, gray—intersection for Caki-1 and Kaiso deficient cells. Representative distribution of Chip-seq data and DNA methylation across CpG islands with increased (**B**) and decreased **C** DNA methylation in Kaiso binding regions at promoters CpG island, graphs were made by GraphPad software where methylation data were fitted with fit spline/LOWESS analysis. Boxplots followed by the Wilcoxon signed-rank test were plotted using ggplot2



C

	chr	start	end	distanceToTSS	gene	log2FoldChange	padj	
uprehulated	chr5	179234777	179234933	453	SQSTM1	0,669650287	6,8E-05	hypo
	chr9	138985225	138985628	1503	NACC2	0,575578201	0,01546	
	chr22	44893364	44893392	613	RTL6	0,531940464	0,027016	
	chr1	224373755	224373804	2845	DEGS1	0,464071299	0,019832	
	chr15	75248137	75248381	1394	RPP25	0,773595073	0,001238	hyper
	chr2	120981244	120981367	-260	TMEM185I	0,70118301	0,007792	
	chr17	74069256	74069378	-649	SRP68	0,486350259	0,007163	
downregulated	chr1	29562674	29562715	-313	PTPRU	-0,389222613	0,014704	
	chr16	88868753	88868786	-1400	CDT1	-0,505497215	0,012912	
	chr5	134073421	134073584	-586	CAMLG	-0,509020551	0,026881	
	chr15	75639467	75639502	136	NEIL1	-0,549390538	0,030976	
	chr9	123836877	123836944	-298	CNTRL	-0,607195442	0,044565	
	chr17	41466174	41466196	70	LINC0091C	-0,91156893	0,048391	
	chrX	107978977	107979515	92	IRS4	-0,997403634	1,04E-05	
	chr1	229476724	229476930	1758	CCSAP	-0,410966054	0,024638	hypo

Fig. 5 DMR analysis. **A** genomic distribution of Kaiso bound DMRs, **B** distrubution of Kaiso bound DMRs relative to CpG islands, **C** table of differentiallly expressed genes with Kaiso peaks at promoters with DMR, **D** SQSTM1 gene within Kaiso binding regions and hypomethylated in Kaiso deficent cells, light blue-Kaiso binding region

islands was observed in both the intergenic and promoter regions (Supplementary Fig. 10a). In addition to the previously described 5', 3', and 5'-3' eroded CpG islands [46], we detected CpG islands exhibiting wave-like methylation patterns (Fig. 6c). In Caki-1 cells, we detected 68 promoter wave-like methylated CpG islands, 48 of which were bound by Kaiso. The characteristic feature of CpG islands with wave-like methylation is their greater length, which can reach 14 kbp (Supplementary Fig. 10b).

Next, we performed an analysis of Kaiso enrichment on the basis of CpG island categories and demonstrated that the enrichment of Kaiso was correlated with the methylation status of the CpG islands, indicating that Kaiso preferentially binds to CpG islands that are highly methylated or with eroded parts of the CpG islands

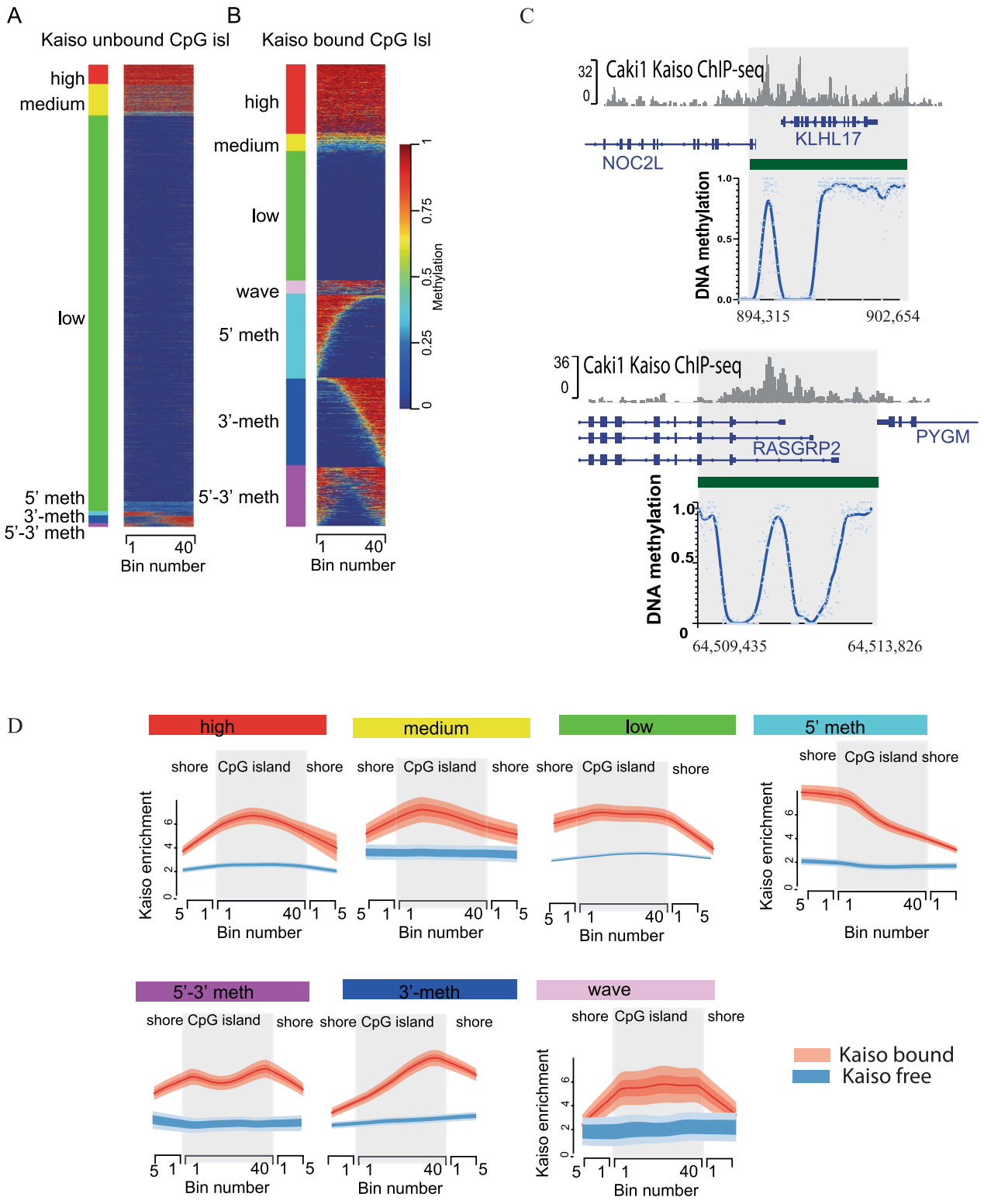


Fig. 6 DNA methylation at promoters CpG islands bound by Kaiso. **A**, **B** Heatmaps showing methylation patterns of promoter CpG islands in Caki-1 cells: **A** Kaiso unbound CpG islands, **B** Kaiso bound CpG islands. Each line represents a single CpG island. Each CpG island was binned into 40 equally sized bins and average methylation per bin was calculated. **C** Integrative Genomics Viewer (IGV) browser track showing Kaiso Chip-seq and DNA methylation for promoter CpG islands of KLHL17, and RASGPR2 genes with wave-like DNA methylation in the Caki-1 cells; **D** Enrichment of Kaiso at indicated types of CpG islands sorted by DNA methylation

(Fig. 6d). GO analysis of genes with erosive CpG islands bound by Kaiso revealed that the genes linked to promoters containing these distinct islands are associated with various transcription factors, including those involved in pluripotency maintenance, diverse signaling pathways, the neuronal system, and cancer pathways. KEGG pathway analysis revealed that CpG islands exhibiting partial methylation (erosion) at the 5′, 3′, or both ends, to which Kaiso binds, are enriched for genes involved in the establishment and maintenance of pluripotency, as well as the development of breast cancer, gastric cancer, and other types of cancer (Supplementary Fig. 10c). In contrast, KEGG analysis of genes with eroded CpG island promoters to which Kaiso does not bind revealed the presence of genes involved in various signaling pathways related to ion channel activity (Supplementary Fig. 10c). In summary, Kaiso binds to eroded CpG islands in the promoters of genes involved in cancer development and pluripotency, whereas Kaiso-unbound eroded CpG islands are associated with genes involved in signaling pathways and ion channel activity.

The E535A mutation in Kaiso leads to the loss of its ability to bind to methylated DNA

We proceeded to examine the significance of the methyl-DNA binding ability of Kaiso in terms of its genomic distribution. Previous studies have indicated that the introduction of the E535A point mutation within the second zinc finger of Kaiso effectively abolishes its binding to methylated sequences while simultaneously retaining its binding affinity for KBS-containing domains [29, 30]. Two clones (cl1 and cl2) were generated via CRISPR/Cas9 genome editing (Supplementary Fig. 11a, b). The mutation was validated through Sanger sequencing (Supplementary Fig. 11c). Western blot analysis revealed that the introduction of this mutation did not affect the level of Kaiso expression or its posttranslational modifications (Fig. 7a).

A detailed examination of the ChIP-seq data revealed that the E535A mutation resulted in a reduction or complete loss of Kaiso protein binding to the majority of its binding sites, as depicted in Fig. 6b. The genomic distribution of the E535A variant of the Kaiso protein indicated a reduction in the number of binding sites located in the CpG islands (Fig. 7c). In agreement with the loss of DNA methyl-binding activity observed for E535A Kaiso in vitro, the degree of enrichment of the CGCG sequence in regions where E535A Kaiso is bound was reduced, whereas that of the KBS and KBS-like sites remained unchanged (Fig. 7d, Supplementary Fig. 11d). Consequently, it may be reasonably deduced that methyl-DNA binding activity of Kaiso is a prevalent phenomenon in Caki-1 cells.

Genomic distribution of Kaiso in K562, MCF-7, and GM12878 cell lines

Next, we investigated the extent to which the identified features of Kaiso binding to chromatin are representative of other cancer cell lines. To achieve this, we conducted a search in the ENCODE database for ChIP-seq analyses performed using high-quality antibodies against Kaiso. We identified three datasets corresponding to breast cancer MCF-7 cells and two lymphoblastoid cell lines, GM12878 (non-cancerous cells) and K562 (cancer cells). In K562 cells, approximately 60,000 peaks were detected, whereas MCF-7 and GM12878 exhibited around 10,000 peaks each (Supplementary Fig. 12a,b). The highest overlap of binding regions was observed between Caki-1 and K562 cells (Supplementary Fig. 12b). For all cell lines analyzed, we observed an enrichment of binding regions in promoter regions and the first intron, consistent with our findings in Caki-1 cells (Fig. 8a). Kaiso-associated regions located in promoters and exons were prominently represented within CpG islands and shores compared to random sequences similar to that observed in Caki-1 cells (Fig. 8b, Supplementary Fig. 12c). Analysis of the enrichment of CGCG and CTGCNA in Kaiso binding regions for these three cell lines confirmed that CGCG is enriched across all genomic elements, while enrichment of CTGCNA was detected in introns and intergenic binding regions (Fig. 8c, Supplementary Fig. 13). Motifs containing a single CpG were identified in the HOMER analysis of binding regions for all tested cell lines (Fig. 8d).

Further, we investigated the methylation levels of Kaiso binding regions in these cell lines. We successfully obtained whole-genome bisulfite sequencing (WGBS) data for the K562 and GM12878 lines. In K562 cells, we observed a hypomethylated state of the genome, in contrast to Caki-1 cells, while the methylation level in the

(See figure on next page.)

Fig. 7 E535A mutation impaires methyl-DNA binding activity in vivo. **A** Western blot analyses of total cell extracts from two clones E535A Kaiso obtained after CRISPR/CAS9 genome editing.* -SUMOylated form of Kaiso protein, **B** Reads distribution in Kaiso binding regions of Caki-1 cells and E535A mutant cells after subtraction of data obtained in Kaiso deficient cells, **C** the overlay between Kaiso binding sites for Caki-1 cells and E535A cells and CpG islands, shores, and shelves, **D** The enrichment of indicated sequences in Kaiso binding regions in Caki-1 cells and cells expressing E535A Kaiso

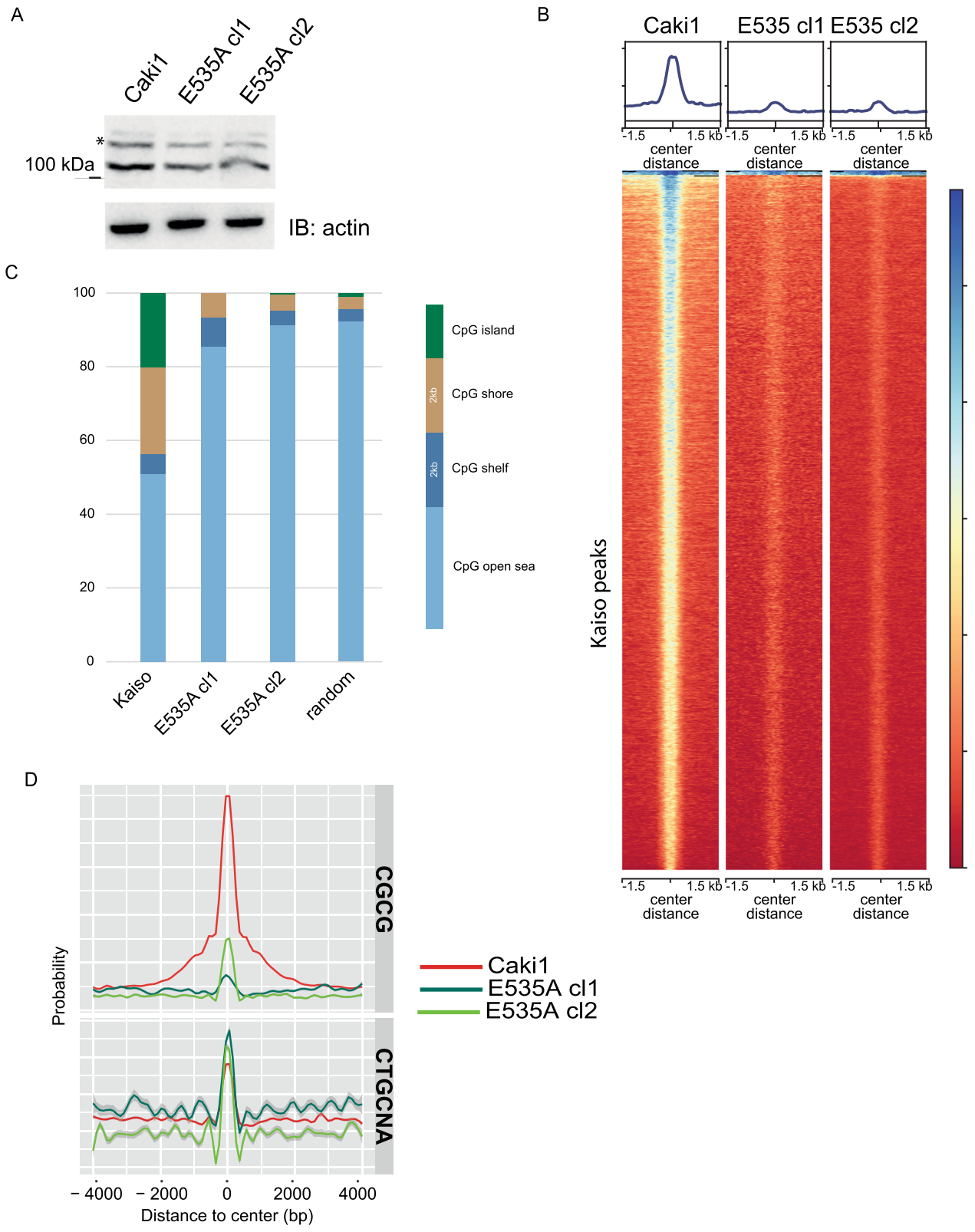


Fig. 7 (See legend on previous page.)

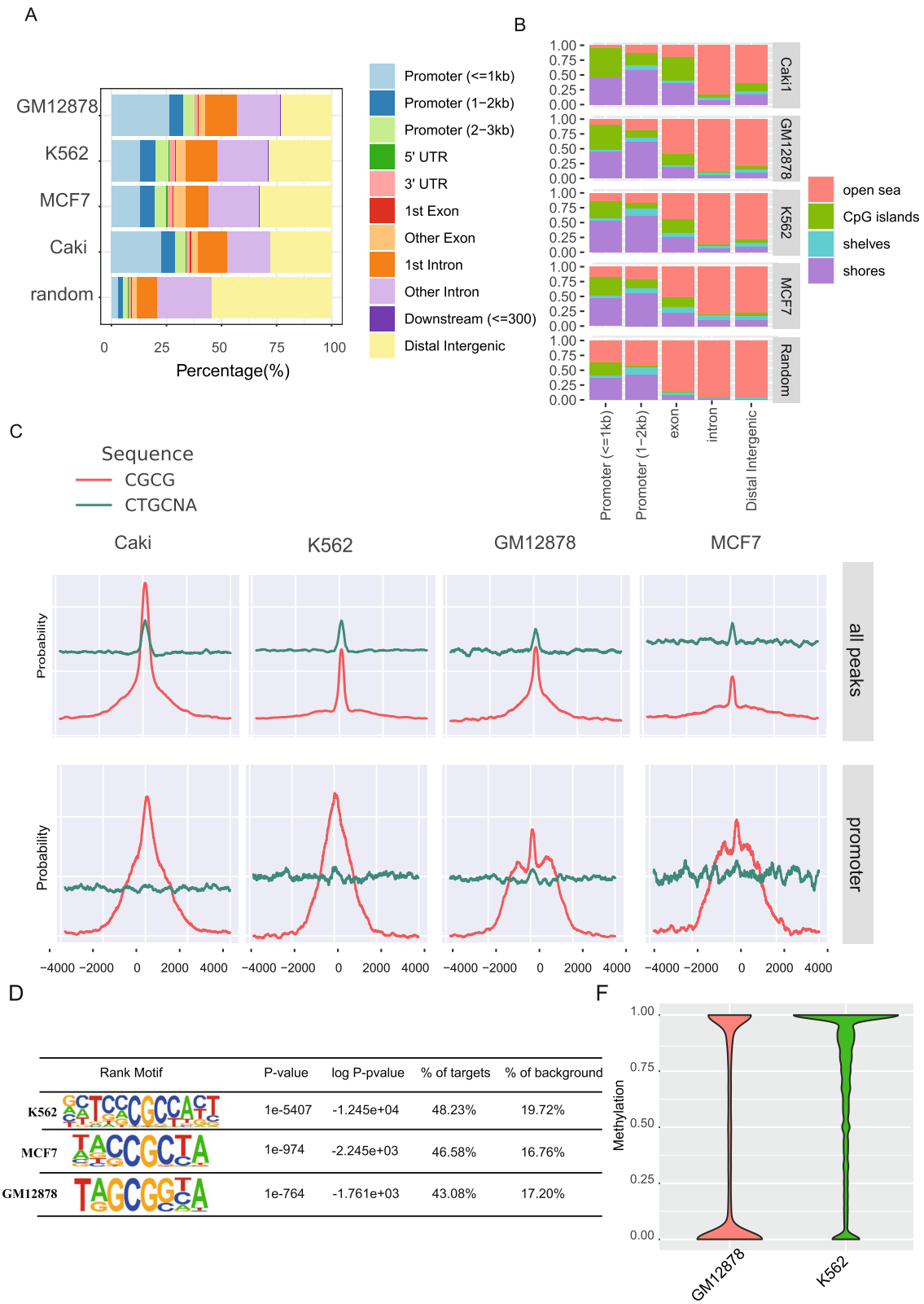


Fig. 8 Kaiso binding landscape in K567, GM12878, MCF7 cell lines, **A** distribution of Kaiso binding regions across the genome; **B** the overlay between Kaiso binding sites and CpG islands, shores, shelves, and open sea; **C** enrichment of indicated top motif in Kaiso binding regions for all (top panel) and promoter (bottom) peaks; **D** de novo Homer analysis of sequences in Kaiso binding regions in indicated cell lines; **F** DNA methylation within Kaiso binding regions detected in K562 and GM12878

non-cancerous GM12878 cells was found to be at an intermediate level (Supplementary Fig. 14a, c). The Kaiso binding regions in the cancerous K562 cells exhibited hypermethylation, despite the overall hypomethylated state of the genome (Fig. 8f, Supplementary Fig. 14a). The methylation pattern of Kaiso binding sites was remarkably similar to those observed in Caki-1 cells. Hypermethylation of binding regions was detected across nearly all genomic categories. The only exception was found in the promoter regions, where a bimodal methylation pattern was evident, similar to that observed in Caki-1 cells (Fig. 3a, Supplementary Fig. 14a). For the non-cancerous GM12878 cells, we detected a bimodal distribution of methylation in Kaiso binding regions, with the lowest levels of methylation observed in promoter regions and intron–exon boundaries (Fig. 8f, Supplementary Fig. 14c). Next, we were interested in exploring whether wave-like methylation patterns observed in Caki-1 cells exist in other cell types. Since these CpG islands exhibited increased length, we conducted a methylation analysis of CpG islands longer than 2000 base pairs that overlapped with Kaiso peaks. Indeed, we were able to detect this type of methylation pattern for both K562 and GM12878 cells (Supplementary Fig. 15) specific to Kaiso peaks. The methylation pattern was very similar between the cell lines; however, it is noteworthy that there were extended regions that were fully methylated in some cells while partially methylated in others (Supplementary Fig. 15). This suggests that this may be a consequence of either cell heterogeneity or the presence of monoallelic methylation.

Discussion

Multiple studies have demonstrated that Kaiso effectively binds both methylated and unmethylated DNA in vitro [26, 27, 47]. However, the role of DNA methylation in regulating Kaiso binding in vivo remains controversial. Specifically, Kaiso preferentially binds to nonmethylated DNA [28]. This discrepancy may have been due to the low depth of sequence data coverage or the use of poorly selected antibodies for ChIP. To address these issues, we generated a Kaiso-deficient renal clear cancer cell line, Caki-1, and performed ChIP-seq analysis in Caki-1 cells using Kaiso-KO cells as a negative control to minimize the detection of false positive binding sites. By using Kaiso-deficient cells as a control, we aimed to identify Kaiso binding sites more accurately and elucidate the role of DNA methylation in regulating Kaiso binding in vivo. Our findings regarding the distribution of Kaiso in the genome were generalized based on ChIP-seq data from cell lines K562 and GM12878.

In our study, we demonstrated that Kaiso predominantly binds to methylated regions, showing broader binding specificity than MBD proteins do by interacting with both CpG-rich and CpG-poor regions. The Kaiso binding sequence (KBS) motif (CTGCNA) was identified in 77% of the Kaiso binding regions, which is consistent with previous in vitro findings [48], although KBS is widely distributed throughout the genome. Enrichment of KBS was primarily observed in Kaiso-binding regions located in introns and intergenic regions. By integrating data from Kaiso binding profiles and whole-genome bisulfite sequencing, we tested the association between Kaiso recruitment and the density of methylated CpG sites. Our findings revealed a positive moderate correlation between the enrichment of Kaiso and the density of methylated CpG sites. This relationship mirrors the behavior observed in methyl-DNA binding proteins such as MBD1, MBD2, MBD4, and MeCP2, where their enrichment shows a linear relationship with local methylation density [49].

Despite the similarities in binding to highly methylated DNA between the Kaiso and MBD proteins, the impact of Kaiso and MBD2 inactivation on DNA methylation patterns differs. The knockdown of MBD2 in embryonic stem cells results in a slight decrease in genome-wide methylation, whereas Kaiso-deficient Caki-1 cells exhibit a slight increase in methylation across the genome, accompanied by hypomethylation at enhancers and gene bodies [8, 50]. The hypermethylation observed in Kaiso-deficient cells throughout the genome, irrespective of the presence of the Kaiso binding site, may be linked to reduced Tet1 dioxygenase transcription [8]. DMRs associated with Kaiso binding regions are predominantly located in promoters, demonstrating both hypo- and hypermethylation. It has previously been shown that Kaiso can compete for DNA binding with KLF4, which, by recruiting TET dioxygenases, can influence methylation [8]. The hypermethylation of DMRs in Kaiso-deficient cells suggests that Kaiso in wild-type cells competes with other proteins for DNA binding that attract the methyltransferase complex. These findings collectively suggest that Kaiso plays a role in regulating the DNA methylation balance of specific genomic targets. One of the identified targets is the SQSTM1 gene. Kaiso binds to a region located downstream of its TSS. Depletion of Kaiso results in a reduction of methylation at this locus, accompanied by an increase in gene transcription. SQSTM1 (p62) functions as an adaptor protein that facilitates the transport of ubiquitinated proteins to the autophagosome and protein degradation. Notably, SQSTM1 acts as a cell survival signal for acute myeloid leukemia (AML) cells undergoing granulocyte differentiation[51]. On the other hand, Kaiso has been discovered as a factor involved in the pathogenesis of acute myeloid leukemia, having been identified among genes with

somatic mutations that serve as positive selective markers and drivers of clonal hematopoiesis [22]. It would be interesting to further investigate whether Kaiso maintains its regulatory capacity over SQSTM1 in myeloid leukemia cells.

One of the closest homologs of Kaiso is ZBTB38. Despite their structural similarity, repression properties, and shared ability to bind methylated DNA, these factors appear to have distinct functional roles and nonoverlapping tissue-specific expression patterns [52]. Deficiencies in Kaiso and ZBTB38 lead to different consequences [18, 53]. Both Kaiso and ZBTB38 exhibit similar localization in upstream regions from transcription start sites (TSSs) in CpG islands and/or shores [53]. However, in contrast to ZBTB38, which is predominantly found at active transcription sites, Kaiso binds to promoters regardless of their activity status. Interestingly, in inactive genes, Kaiso is located not upstream but directly at the TSS itself. These two factors may have partial functional interchangeability, particularly in terms of binding to upstream regions in the promoters of active genes and CpG islands/shores. However, they clearly exhibit differences in their DNA binding profiles. This fact suggests that they are not entirely redundant and may play unique roles in gene regulation and cellular processes.

Under normal cellular conditions, the majority of CpG islands remain unmethylated, with only a small fraction exhibiting hypermethylation. However, in cancer cells, there is a notable shift in CpG island methylation patterns. A significant portion of CpG islands in cancer cells undergo hypermethylation, and a phenomenon known as eroded methylation occurs, where methylation is observed at the 5' or 3' ends of CpG islands or where hypermethylation is present at both ends, resulting in a central unmethylated region within the CpG islands [46]. The boundaries of eroded CpG islands in cancer cells are marked by H3K4me1 modification, which is also observed in normal cells and plays a crucial role in the development of eroded methylation in CpG islands during cancer cell transformation. Our research revealed that the Kaiso protein is enriched at both methylated and eroded CpG islands, indicating its preference for binding to methylated DNA. Here, we identified a novel class of CpG islands characterized by interesting methylation patterns: (i) wave-like or (ii) those with methylation in the central region of the CpG island. We demonstrated that Kaiso is associated with these types of CpG islands. Although these wave-like regions were relatively rare, Kaiso bound to approximately two-thirds of them. We did not observe significant alterations in the methylation status of Kaiso knockout cells. Therefore, we speculate that Kaiso may be involved in establishing such DNA methylation profiles but not in their maintenance. Notably, the CpG islands with wave-like methylation patterns differ from all others in that they are quite extensive, reaching lengths of up to 14,000 base pairs. These islands may encompass entire genes, including promoters, exons, enhancers and multiple TSSs. The mechanism that establishes and maintains methylation for only particular parts of large CpG islands remains unknown. The role of Kaiso as an initiator or interpreter of DNA methylation for such CpG islands has yet to be discovered.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13072-024-00565-3.

Additional file 1.

Acknowledgements

We thank Albert Reynolds (1956--2022) for providing anti-Kaiso antibodies. The calculations were performed via computational resources of the federal collective usage center MCC NRC "Kurchatov Institute", http://computing.nrcki.ru/ (accessed on 1 May 2021). Sanger sequencing was carried out on the equipment of the Center of Collective Use "Bioengineering" of the Federal Research Center of Biotechnology RAS.

Author contributions

A.S., performed data analysis of ChIP-seq, RNA-seq, WGBS. P.A. performed data analysis of Chip-seq, DMRs analysis. Y.L. performed ChIP. F.S. performed primary analysis of deep sequencing data and methylome analysis. G.F. prepared figures, genome editting. Dm.K. performed enrichment analysis of predicted binding sites. D.K. performed cell culture work, material collection,western. I.D. -plotting graphs for genes with wave-like methylation. A.M. performed library preparation and deep gene sequencing. E.P. resources, funding acquisition,co-wrote the manuscript, interpreted the data. S.Z. performed genome editting, analyzed and interpreted the data, envisioned project, co-wrote manuscript, prepared figures. All authors reviewed the manuscript.

Funding

This research was funded by the Russian Science Foundation, project no. 19–74-30026 (methylome analysis), and by the Ministry of Science and Higher Education of the Russian Federation, project AAAA-A20-120121790092–6 (EMSA experiments).

Data availability

The RNA-seq, ChIP-seq, and WGBS raw data have been submitted to the Sequence Read Archive under the accession number PRJNA734133. BigWig files viewable in genome browsers are publicly accessible on Figshare (https://doi.org/https://doi.org/10.6084/m9.figshare.26116978).

Declarations

Competing interests

The authors declare no competing interests.

Author details

[1]Federal Research Centre, Fundamentals of Biotechnology», Russian Academy of Sciences, 119071 Moscow, Russia. [2]Institute of Gene Biology, Russian Academy of Sciences, 119334 Moscow, Russia. [3]Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991 Moscow, Russia. [4]Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117997 Moscow, Russia.

Received: 10 September 2024 Accepted: 3 December 2024
Published: 19 December 2024

References

1. Wan J, Oliver VF, Wang G, Zhu H, Zack DJ, Merbs SL, et al. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BMC Genomics. 2015;16:49.
2. Kaplun DS, Kaluzhny DN, Prokhortchouk EB, Zhenilo SV. DNA methylation: genomewide distribution, regulatory mechanism and therapy target. Acta Naturae. 2022;14:4–19.
3. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128:683–92.
4. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009;41:178–86.
5. Heyn H, Vidal E, Ferreira HJ, Vizoso M, Sayols S, Gomez A, et al. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. Genome Biol. 2016;17:11.
6. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nat Rev Cancer. 2013;13:497–510.
7. Damaschke NA, Gawdzik J, Avilla M, Yang B, Svaren J, Roopra A, et al. CTCF loss mediates unique DNA hypermethylation landscapes in human cancers. Clin Epigenetics. 2020;12:80.
8. Kaplun D, Starshin A, Sharko F, Gainova K, Filonova G, Zhigalova N, et al. Kaiso regulates DNA methylation homeostasis. Int J Mol Sci. 2021. https://doi.org/10.3390/ijms22147587.
9. Salmon M. Transcriptional and Epigenetic Regulation of Krüppel-Like Transcription Factors. Gene expression and phenotypic traits [Internet]. 2020; Available from: https://books.google.com/books?hl=en&lr=&id=NEv9DwAAQBAJ&oi=fnd&pg=PA55&dq=Transcriptional+and+Epigenetic+Regulation+of+Kr%C3%BCppel-Like+Transcription+Factors&ots=IzqCwR-0TG&sig=MMcxlVudhCjsYJw_iZD4RDctup8
10. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet. 2016;17:551–65.
11. Zhigalova NA, Sokolov AS, Prokhortchouk EB, Zhenilo SV. S100A3 is a new target gene of Kaiso in mouse skin. Mol Biol. 2015;49:362–5.
12. Zhenilo SV, Musharova OS, Pokhorchuk EB. Transcription factor Kaiso does not interact with hydroxymethylated DNA within CTGCNA sequence context. Mol Biol. 2013;47:522–5.
13. Huynh KD, Bardwell VJ. The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. Oncogene. 1998;17:2473–84.
14. Yoon H-G, Chan DW, Reynolds AB, Qin J, Wong J. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell. 2003;12:723–34.
15. Raghav SK, Waszak SM, Krier I, Gubelmann C, Isakova A, Mikkelsen TS, et al. Integrative genomics identifies the corepressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBPβ and KAISO. Mol Cell. 2012;46:335–50.
16. Rodova M, Kelly KF, VanSaun M, Daniel JM, Werle MJ. Regulation of the rapsyn promoter by kaiso and delta-catenin. Mol Cell Biol. 2004;24:7188–96.
17. Zhenilo S, Deyev I, Litvinova E, Zhigalova N, Kaplun D, Sokolov A, et al. DeSUMOylation switches Kaiso from activator to repressor upon hyperosmotic stress. Cell Death Differ. 2018;25:1938–51.
18. Kaplun DS, Fok RE, Korostina VS, Prokhortchouk EB, Zhenilo SV. Kaiso gene knockout promotes somatic cell reprogramming. Biochemistry. 2019;84:283–90.
19. Bohne F, Langer D, Martiné U, Eider CS, Cencic R, Begemann M, et al. Kaiso mediates human ICR1 methylation maintenance and H19 transcriptional fine regulation. Clin Epigen. 2016. https://doi.org/10.1186/s13148-016-0215-4
20. Kulikov AV, Korostina VS, Kulikova EA, Fursenko DV, Akulov AE, Moshkin MP, et al. Knockout Zbtb33 gene results in an increased locomotion, exploration and pre-pulse inhibition in mice. Behav Brain Res. 2016;297:76–83.
21. Robinson SC, Klobucar K, Pierre CC, Ansari A, Zhenilo S, Prokhortchouk E, et al. Kaiso differentially regulates components of the Notch signaling pathway in intestinal cells. Cell Commun Signal. 2017;15:24.
22. Bernstein N, Spencer Chapman M, Nyamondo K, Chen Z, Williams N, Mitchell E, et al. Analysis of somatic mutations in whole blood from 200,618 individuals identifies pervasive positive selection and novel drivers of clonal hematopoiesis. Nat Genet. 2024. https://doi.org/10.1038/s41588-024-01755-1.
23. Beauchamp EM, Leventhal M, Bernard E, Hoppe ER, Todisco G, Creignou M, et al. ZBTB33 is mutated in clonal hematopoiesis and myelodysplastic syndromes and impacts RNA splicing. Blood Cancer Discov. 2021;2:500–17.
24. Davidson G, Helleux A, Vano YA, Lindner V, Fattori A, Cerciat M, et al. Mesenchymal-like tumor cells and myofibroblastic cancer-associated fibroblasts are associated with progression and immunotherapy response of clear cell renal cell carcinoma. Can Res. 2023;83:2952–69.
25. Bai D, Cheng Y, Lu X, Namasivayam GP, Sugiyama H. DNA methylation modification patterns identify distinct prognosis and responses to immunotherapy and targeted therapy in renal cell carcinoma. Front Biosci (Landmark Ed). 2023;28:224.
26. Prokhortchouk A. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor [Internet]. Genes & Development. 2001. p. 1613–8. https://doi.org/10.1101/gad.198501
27. Daniel JM. The p120ctn-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides [Internet]. Nucleic Acids Res. 2002. https://doi.org/10.1093/nar/gkf398.
28. Blattler A, Yao L, Wang Y, Ye Z, Jin VX, Farnham PJ. ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes. Epigenet Chromatin. 2013;6:13.
29. Nikolova EN, Stanfield RL, Dyson HJ, Wright PE. CH···O hydrogen bonds mediate highly specific recognition of methylated CpG sites by the zinc finger protein Kaiso. Biochemistry. 2018;57:2109–20.
30. Nikolova EN, Stanfield RL, Dyson HJ, Wright PE. A conformational switch in the zinc finger protein kaiso mediates differential readout of specific and methylated DNA sequences. Biochemistry. 2020;59:1909–26.
31. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9:R137.
32. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
33. Artemov AV, Zhigalova N, Zhenilo S, Mazur AM, Prokhortchouk EB. VHL inactivation without hypoxia is sufficient to achieve genome hypermethylation. Sci Rep. 2018;8:10667.
34. Artemov AV, Zhenilo S, Kaplun D, Starshin A, Sokolov A, Mazur AM, et al. An IDH-independent mechanism of DNA hypermethylation upon VHL inactivation in cancer. Epigenetics. 2022;17:894–905.
35. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27:1571–2.
36. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS ONE. 2013;8: e81148.
37. Isagulieva AK, Kaluzhny DN, Beniaminov AD, Soshnikova NV, Shtil AA. Differential Impact of Random GC Tetrad Binding and Chromatin Events on Transcriptional Inhibition by Olivomycin A. Int J Mol Sci. 2022. https://doi.org/10.3390/ijms23168871.
38. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576–89.
39. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160–5.
40. Altuna Akalin [aut, cre], Vedran Franke [aut, cre], KatarzynaWreczycka [aut], Alexander Gosdschan [ctb], Liz Ing-Simmons[ctb], Bozena Mika-Gospodorz [ctb]. genomation [Internet]. Bioconductor; 2017. Available from: https://bioconductor.org/packages/genomation
41. Balaramane D, Spill YG, Weber M, Bardet AF. MethyLasso: a segmentation approach to analyze DNA methylation patterns and identify differentially methylated regions from whole-genome datasets. Nucleic Acids Res. 2024. https://doi.org/10.1093/nar/gkae880.

42. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36:2628–9.
43. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49:D545–51.
44. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8:2281–308.
45. Qin S, Zhang B, Tian W, Gu L, Lu Z, Deng D. Kaiso mainly locates in the nucleus in vivo and binds to methylated, but not hydroxymethylated DNA. Chin J Cancer Res. 2015;27:148–55.
46. Skvortsova K, Masle-Farquhar E, Luu P-L, Song JZ, Qu W, Zotenko E, et al. DNA hypermethylation encroachment at CpG island borders in cancer is predisposed by H3K4 monomethylation patterns. Cancer Cell. 2019;35:297-314.e8.
47. Prokhortchouk AV, Aitkhozhina DS, Sablina AA, Ruzov AS, Prokhortchouk EB. Kaiso, a new protein of the BTB/POZ family, specifically binds to methylated DNA sequences. Russ J Genet. 2001;37:603–9.
48. Donaldson NS, Pierre CC, Anstey MI, Robinson SC, Weerawardane SM, Daniel JM. Kaiso represses the cell cycle gene cyclin D1 via sequence-specific and methyl-CpG-dependent mechanisms. PLoS ONE. 2012;7:e50398.
49. Baubec T, Ivánek R, Lienert F, Schübeler D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell. 2013;153:480–92.
50. Hainer SJ, McCannell KN, Yu J, Ee L-S, Zhu LJ, Rando OJ, et al. DNA methylation directs genomic localization of Mbd2 and Mbd3 in embryonic stem cells. Elife. 2016;5:1. https://doi.org/10.7554/eLife.21964.
51. Trocoli A, Bensadoun P, Richard E, Labrunie G, Merhi F, Schläfli AM, et al. p62/SQSTM1 upregulation constitutes a survival mechanism that occurs during granulocytic differentiation of acute myeloid leukemia cells. Cell Death Differ. 2014;21:1852–61.
52. Filion GJP, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez P-A. A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol. 2006;26:169–81.
53. Marchal C, Defossez P-A, Miotto B. Context-dependent CpG methylation directs cell-specific binding of transcription factor ZBTB38. Epigenetics. 2022;1:1–22.
54. Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 2001;15:710–23.
55. Cooper SJ, Trinklein ND, Anton ED, Nguyen L, Myers RM. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. Genome Res. 2006;16:1–10.

Publisher's Note