

RESEARCH

Open Access



Epigenetic reshaping through damage: promoting cell fate transition by BrdU and IdU incorporation

Chuang Li^{1,2,3}, Xiaoduo Xu^{1,2,3†}, Shuyan Chen^{2,3,5†}, Anchun Xu^{1,2,3}, Tongxing Guan^{1,2,3}, Haokaifeng Wu^{3,4,6*}, Duanqing Pei^{1*} and Jing Liu^{2,3*} 

Abstract

Background Thymidine analogs have long been recognized for their ability to randomly incorporate into DNA. However, the precise mechanisms through which thymidine analogs facilitate cell fate transition remains unclear.

Results Here, we discovered a strong correlation between the dosage dependence of thymidine analogs and their ability to overcome reprogramming barrier. The extraembryonic endoderm (XEN) state seems to be a cell's selective response to DNA damage repair (DDR), offering a shortcut to overcome reprogramming barriers. Meanwhile, we found that homologous recombination repair (HRR) pathway causes an overall epigenetic reshaping of cells and enabling them to overcome greater barriers. This response leads to the creation of a hypomethylated environment, which facilitates the transition of cell fate in various reprogramming systems. We term this mechanism as Epigenetic Reshaping through Damage (ERD).

Conclusion Overall, our study finds that BrdU/IdU can activate the DNA damage repair pathway (HRR), leading to increased histone acetylation and genome-wide DNA demethylation, regulating somatic cell reprogramming. This offers valuable insights into mechanisms underlying cell fate transition.

[†]Xiaoduo Xu and Shuyan Chen contributed equally to this work.

*Correspondence:

Haokaifeng Wu

wu_haokaifeng@gibh.ac.cn

Duanqing Pei

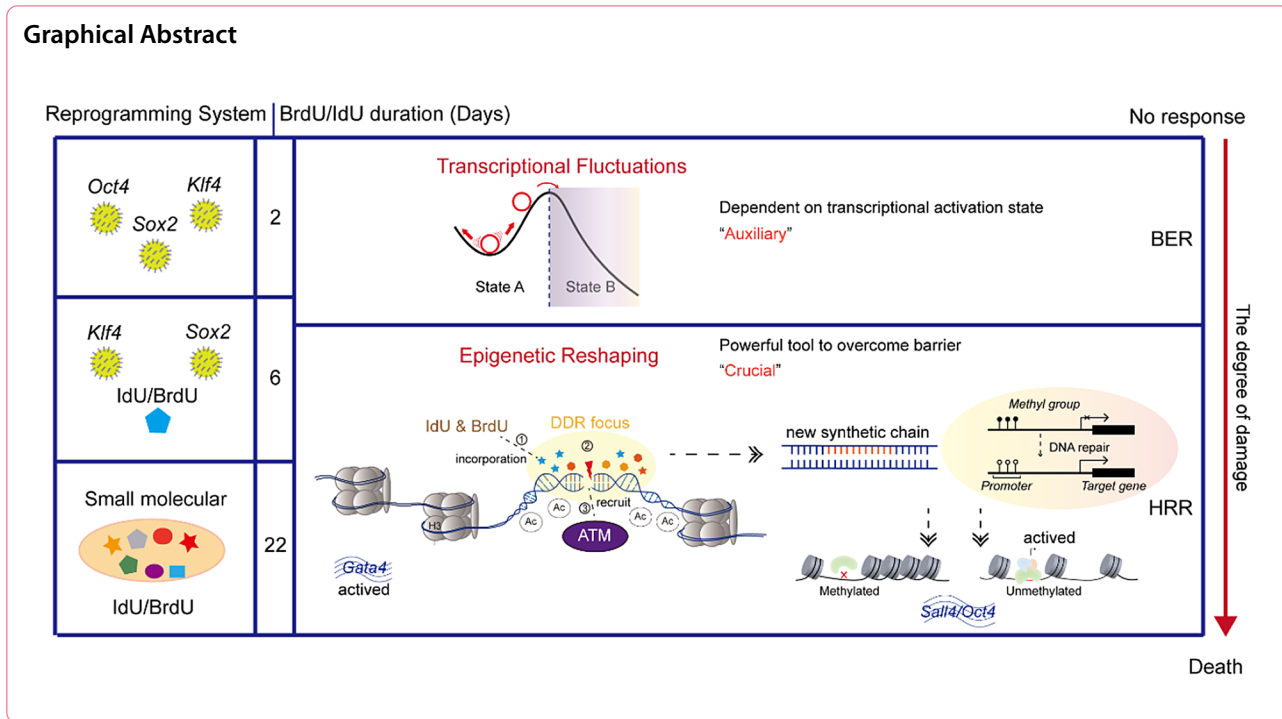
peiduanqing@westlake.edu.cn

Jing Liu

liu_jing@gibh.ac.cn

Full list of author information is available at the end of the article





Introduction

Thymidine analogs are often considered to incorporate randomly into DNA sequences, introducing inherent unpredictability and randomness in their mechanisms of action [1]. Despite this inherent nature, they play a critical role in the precise regulation of cell fate [2–7]. In the study conducted by Xie et al., the utilization of BrdU, widely employed to label proliferating cells in vivo, has demonstrated its significant potential in facilitating chemical induction of pluripotency [3]. The study by Cao et al., sheds light on how BrdU can impact the reorganization of nuclear architecture and influence the cell fate decisions [5]. IdU, which is also a thymidine analog was reported to cause stochastic fluctuations in gene expression to facilitate cellular reprogramming [8].

However, the precise mechanisms through which BrdU/IdU facilitates cell reprogramming are not yet fully understood. BrdU/IdU may be involved in cell reprogramming through mechanisms beyond transcriptional fluctuations. This could include the impacts of thymidine analogs on DNA structure, repair mechanisms, or other cellular processes, providing a potential shortcut for cells to overcome reprogramming barriers. As a result, thymidine analogs may play a more crucial and multifaceted role in regulating cell fate, extending beyond the induction of transcriptional fluctuations alone. This study links DNA damage repair with genome, epigenetics and cell fate regulation, providing us with new understanding

and helping to advance the field of cell fate decision and regenerative medicine.

Results

BrdU and IdU: essential in overcoming two barriers during chemical induction of pluripotency

To determine the optimal duration of BrdU treatment for chemical induction of pluripotency (CIP), a series of time windows were optimized and observed for their impact on CIP (Fig. 1A). Two main barriers were found in the CIP reprogramming process, and BrdU was found to play a crucial role in overcoming these barriers. BrdU dropout experiments showed that dropout of BrdU (removing BrdU while keeping all other culture conditions unchanged) at barrier I resulted in a failure to form colonies at Day 22, while dropout at barrier II between Day 12–22 resulted in a failure to form Oct4-GFP⁺ colonies at Day 40 (Fig. 1B). Meanwhile, we have established a time gradient for each barrier (Additional file 1: Fig. S1A) in increments of two days to investigate the minimum duration of BrdU action and the effective time window of its action. Breaking through the barrier requires at least 6 days of BrdU incorporation, which must be initiated within the first 4 days (Additional file 1: Fig. S1B). Increasing the duration of BrdU treatment led to an increase in the number of Oct4-GFP⁺ colonies, with a 0–22 days treatment resulting in over 240 Oct4-GFP⁺ colonies at Day 40 from a starting population of

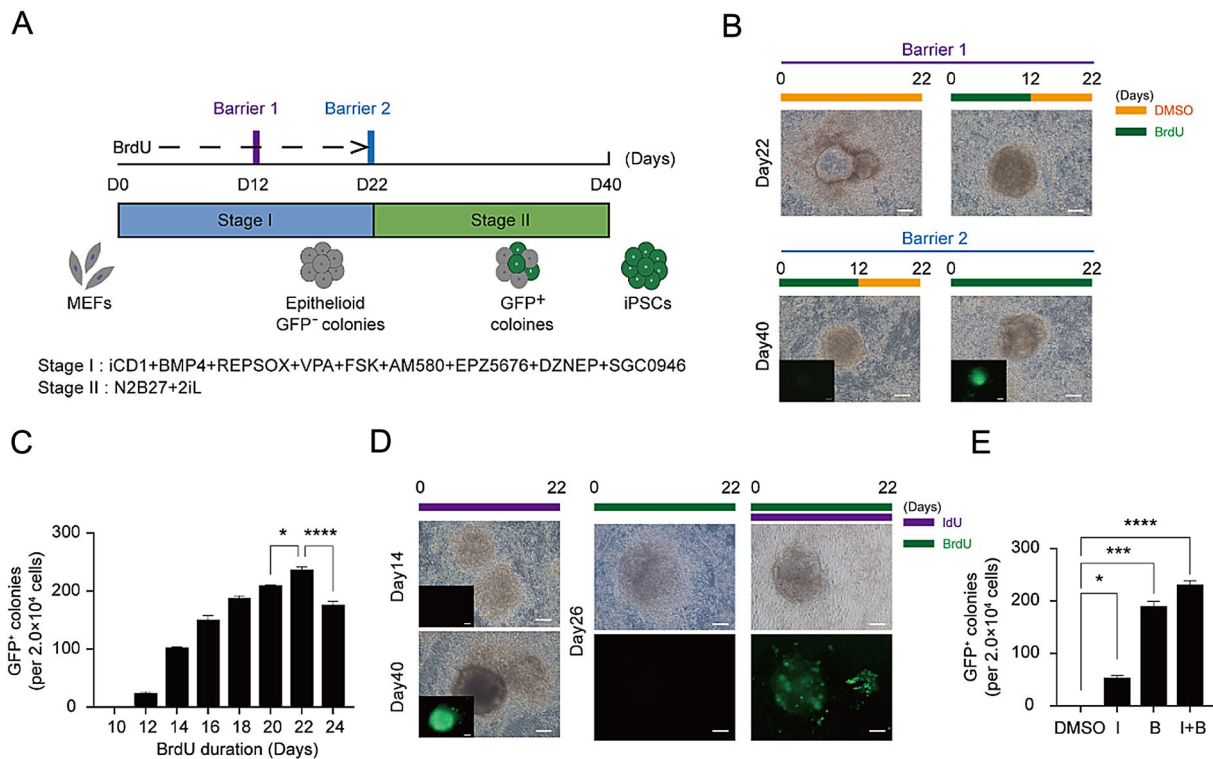


Fig. 1 BrdU and IdU: Essential in overcoming two barriers during chemical induction of pluripotency. **A** Schematic diagram of the induction of CiPSCs from MEFs. **B** Morphological changes at two distinct barriers during induction of CiPSCs. Scale bar 100 μm . **C** Number of Oct4-GFP + CiPSC colonies generated under indicated conditions. $n=3$, * $P < 0.05$, **** $P < 0.0001$. **D** Morphological changes at distinct time points during induction of CiPSCs treated with BrdU or IdU. Scale bar 100 μm . **E** Number of Oct4-GFP + CiPSC colonies generated under different treatment conditions. I, IdU; B, BrdU; I+B, IdU + BrdU, $n=3$, * $P < 0.05$, *** $P < 0.001$ and **** $P < 0.0001$

20,000 cells, achieving an efficiency of 1.2%. In contrast, a 0–12 days treatment only resulted in about 20 Oct4-GFP⁺ colonies (Fig. 1C and Additional file 1: Fig. S1C). Furthermore, we attempted to use three base analogs, namely IdU, EdU, and 5Aza, but only IdU was found to be capable of replacing the function of BrdU during CIP reprogramming (Fig. 1D and Additional file 1: Fig. S1D). The combination of both thymidine analogs can increase the selectivity of cells, so that only cells that have been correctly reprogrammed can survive; hence accelerated the reprogramming process (Fig. 1D,E and Additional file 1: Fig. S1E). Thus, it can be concluded

that BrdU and IdU are essential in overcoming two main barriers during CIP.

Gene expression dynamics and chromatin accessibility dynamics during CIP with or without BrdU/IdU

To identify the two barriers in which BrdU and IdU are involved during CIP reprogramming, RNA-seq and ATAC-seq was performed on MEFs undergoing CIP with or without BrdU, IdU and I+B (IdU + BrdU) at D8, 14, 20, 26 and 40 (Fig. 2A). It was found that loss of BrdU and IdU at barrier I resulted in incomplete reprogramming of MEFs to XEN-like cells (Fig. 2B). The expression of

(See figure on next page.)

Fig. 2 Gene expression dynamics and chromatin accessibility dynamics during CIP with or without BrdU/IdU. **A** Schematic diagram of the induction of CiPSCs from MEFs. **B** Heatmap of RNA-seq for CIP at D8, 14, 20 and 26 with or without BrdU or IdU. **C** The global chromatin status of closed in MEFs but open in ESCs (CO)/ open in MEFs but closed in ESCs (OC) and permanent open (PO) for CIP at D8, 14, 20, 26 and 40 with or without BrdU or IdU. **D** Representative fibroblast genes, XEN genes and pluripotency genes peaks from ATAC-seq for CIP at D8, 14, 20 and 26 with or without BrdU or IdU. **E** The global chromatin status of CO/OC and PO arranged as groups for CIP at D8, 14, 20, 26 and 40 with or without BrdU or IdU. **F** TF motifs enriched at least twofold for CO/OC loci defined by ATAC-seq peaks with or without BrdU or IdU. TF families are indicated on the right of the heatmap. **G** Statistical analysis of number of total peaks in CO1-6/OC1-6. **H** Venn diagrams of OC1-6/CO1-6 peaks between samples with or without BrdU or IdU

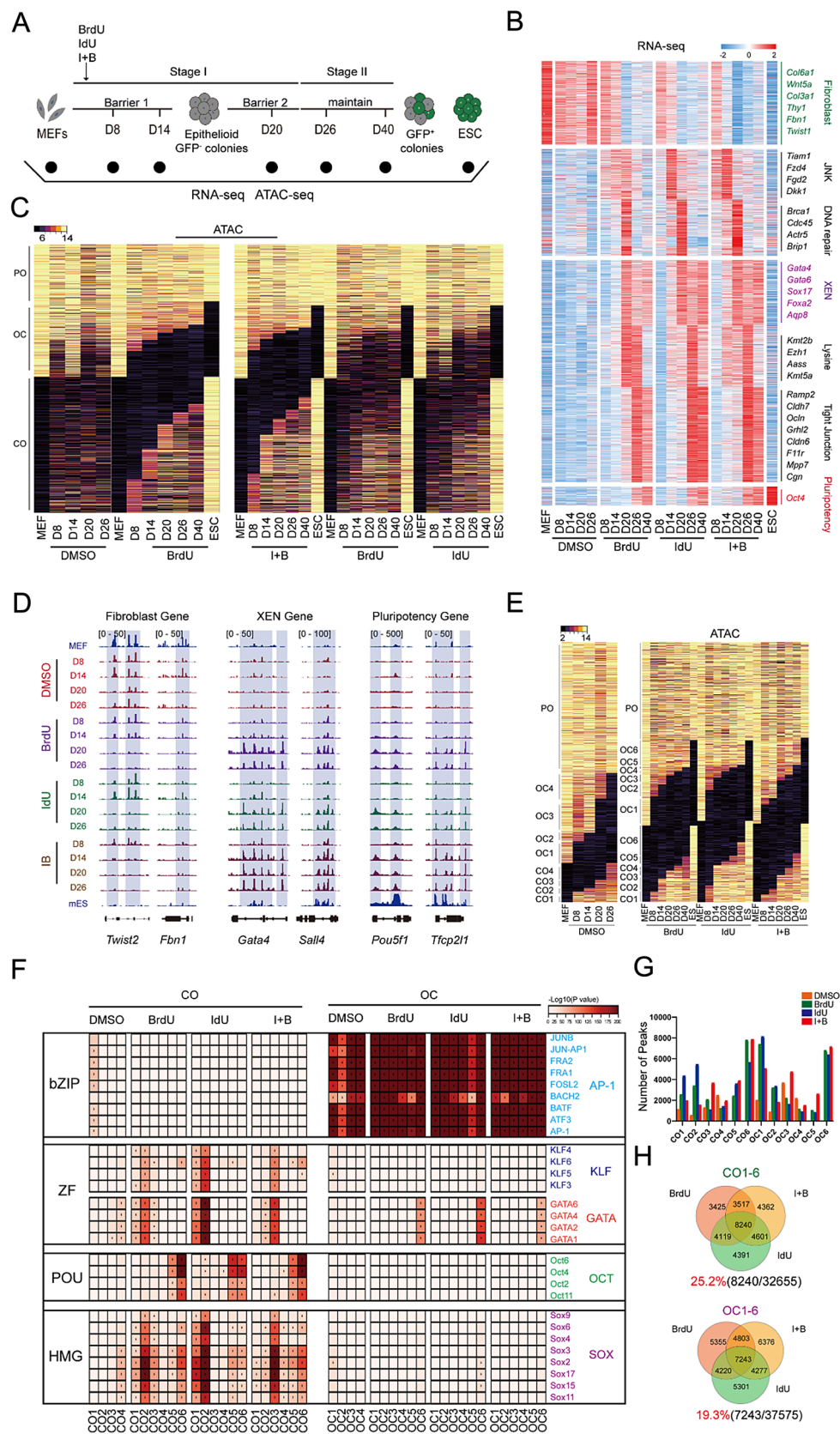


Fig. 2 (See legend on previous page.)

XEN-like genes such as *Gata4*, *Sox17* and *Aqp8* increased with increasing duration of thymidine analogs treatment and reached a peak at D20. Conversely, dropout of BrdU resulted in a failure to express XEN-like genes (Additional file 1: Fig. S2A, C), suggesting that MEFs undergoing reprogramming without BrdU /IdU follow a different fate path. Previous studies have reported that CIP goes through an XEN-like intermediate stage [2]. Additionally, loss of BrdU and IdU at barrier II resulted in incomplete reprogramming of XEN-like cells to chemical induction pluripotent stem cell (CiPSCs). Pluripotency genes such as *Oct4*, *Esrrb*, *Tfcp2l1*, *Nanog* and *Sox2* were highly activated in a BrdU/IdU dependent manner during stage 2. In addition, the incorporation of I+B showed a faster transition towards to iPSCs compared to BrdU/IdU treatment at the RNA level (Additional file 1: Fig. S2B, D).

The function state of a cell is determined by its genome architecture. To investigate the role of BrdU/IdU in CIP, we mapped chromatin accessibility dynamics (CADs) during CIP with or without BrdU/IdU and found that many loci opened during CIP failed to open without BrdU/IdU. Moreover, detailed analysis of the ATAC-seq datasets revealed many similarities between BrdU and IdU treatments. Chromatin accessibility increased when MEFs undergoing CIP were treated with I+B compared to only BrdU or IdU treatment (Fig. 2C). In Fig. 2D, we analyzed the loci near fibroblast marker gene *Twist2* and *Fbn1* found that they were more open in DMSO treatment compared to BrdU/IdU treatment. Conversely, loci near XEN marker genes such as *Gata4* and *Sall4* were only open in BrdU/IdU treatment. Loci near the pluripotency marker gene *Oct4* and *Tfcp2l1* were also only open in BrdU/IdU treatment, and the chromatin accessibility was much higher when samples were treated with both I+B.

We compared the peaks at each locus between these samples and classified them into three categories: closed in MEFs but open in ESCs (CO), open in MEFs but closed in ESCs (OC), and permanent open (PO). Further, we divided the CO and OC peaks into subgroups based on the day of reprogramming to illustrate the progression of cellular reprogramming associated CADs, as shown in Fig. 2E. To further investigate the molecular mechanism of BrdU/IdU during CIP, we performed motif analysis, as

illustrated in Fig. 2F. Loci containing motifs for OCT2, 4, 6, and 11 gradually opened, peaking at CO6 when treated with BrdU/IdU. Loci with motifs for GATA1, 2, 4, 6, KLF3-6, and SOX2-4, 6, 9, 10, 15, 17 opened from CO1-6 when treated with BrdU/IdU. However, the TF motifs profile was similar with or without BrdU/IdU in OC loci, indicating that BrdU/IdU treatment tended to open up loci that are enriched with motifs binding to TFs from the GATA, KLF, and SOX families.

We then statistically analyzed the number of total peaks in CO1-6/OC1-6, respectively, and presented the results in the form of a Venn diagram, as shown in Fig. 2G, H. The diagram showed that BrdU/IdU/I+B shared 25.2% (8240/32655) in CO peaks, but only 19.3% (7243/37575) in OC peaks. Additional file 1: Fig. S2E illustrates that a larger fraction of peaks in CO are located in the promoter regions. In conclusion, the function of BrdU/IdU is more biased towards opening chromatin and activating gene expression.

Incorporation of BrdU/IdU leads to DNA damage repair

According to the RNA-seq analysis we obtained in Fig. 2, we then performed Venn diagrams and enriched GO functions (Fig. 3A, B and Additional file 1: Fig. S3A, B). Interestingly, Venn diagrams for overlapping genes in the upregulated groups between BrdU, IdU, and I+B versus the control (without thymidine analogs) showed that DNA repair-related GO functions were highly enriched when MEFs were treated with BrdU/IdU during CIP (Fig. 3A, B). Specifically, DNA repair-related gene expression was highly upregulated in all three groups. The main repair pathways are base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination repair (HRR) and non-homologous end joining repair (NHEJ). In our study, we observed that *Apex1*, a gene related to BER pathway, was significantly upregulated in the early stages of reprogramming after treatment with BrdU/IdU, but gradually decreased with longer exposure. The HRR related genes *Brca1* and *Brca2* were upregulated as *Apex1* expression decreased (Additional file 1: Fig. S3C). Additionally, a combination of both thymidine analogs (I+B) accelerated the DNA repair process, as shown by the early upregulation of DNA repair genes expression at D8 in

(See figure on next page.)

Fig. 3 Incorporation of BrdU/IdU leads to DNA damage repair. **A** Venn diagrams for overlapping genes in the upregulated groups between BrdU, IdU, and BrdU + IdU versus the control. **B** Enriched GO functions in upregulated groups. **C** Heatmap of DNA-repair genes related RNA-seq for CIP at D8, 14, 20 and 26 with or without BrdU or IdU. **D** Alkaline comet assays of samples with or without BrdU or IdU. Scale bar 50 μ m. **E** Statistical analysis of olive tailmomer from **D**. **F** Average Olive Tail Moment from **E**. **G** Immunofluorescent analysis to observe the spatial colocalization of BrdU (in red) and γ -H2AX (in green). Scale bar 40 μ m. **H** Representative peaks from ATAC-seq aligned with γ -H2AX Cut & tag signals for *Gata4* and *Sox17* with or without BrdU or IdU

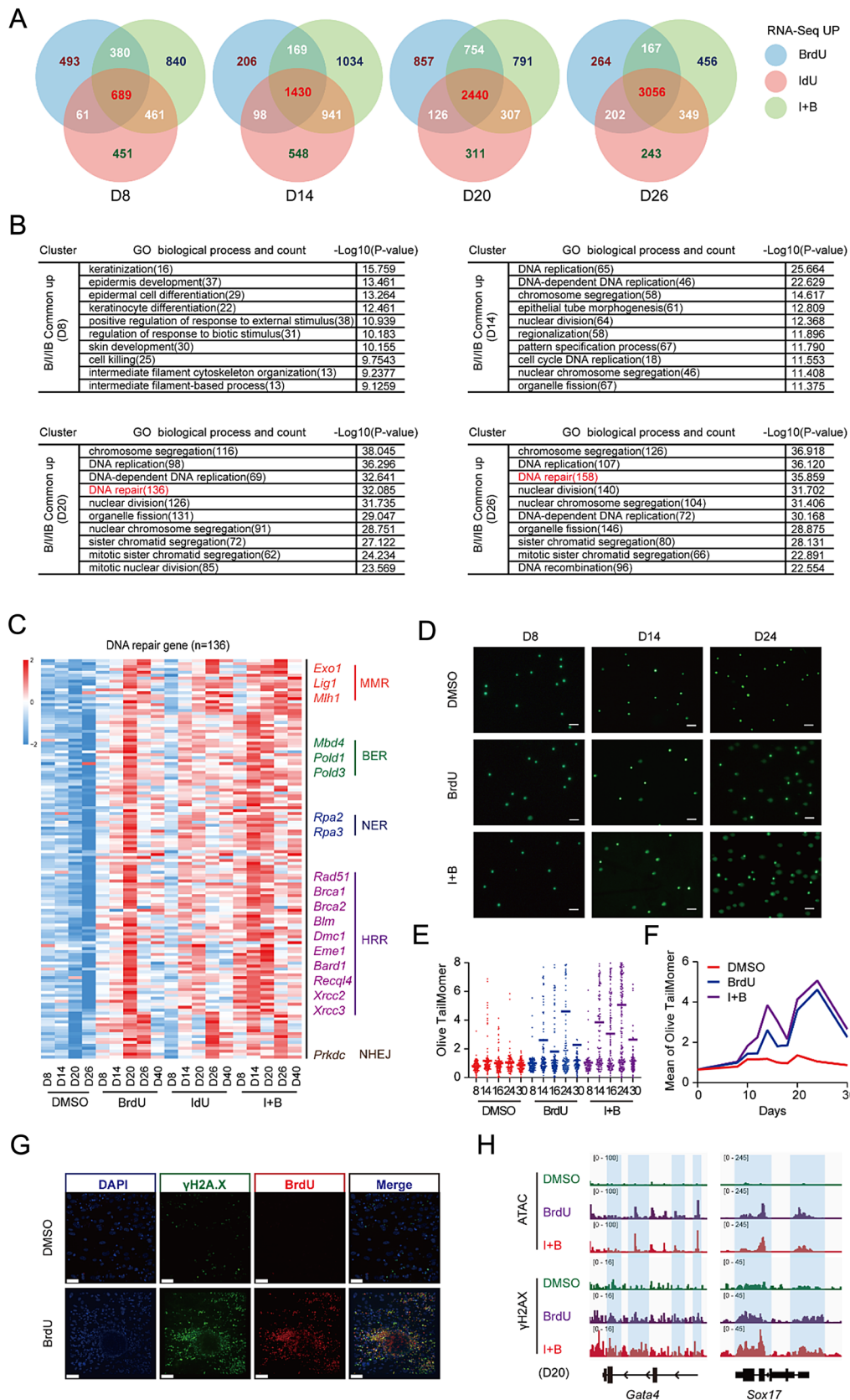


Fig. 3 (See legend on previous page.)

this group compared to the BrdU or IdU only groups (Fig. 3C).

To verify the relationship between thymidine analogs and DNA damage repair, alkaline comet assays were performed to compare the olive tail moment difference between samples with or without thymidine analogs treatment (Fig. 3D, E). Interestingly, the BrdU treatment showed an increased olive tail moment indicating significant DNA damage. Furthermore, the degree of DNA damage was further promoted by a combination of BrdU and IdU. Figure 3F shows two DNA damage peaks during the CIP process at D14 and D24, which is coherent with the time window of the two main barriers discussed in the previous sections. This evidence illustrates a close association between DNA damage/repair and thymidine analogues assisting CIP through two barriers.

To verify whether BrdU treatment directly causes DNA damage, immunofluorescent analysis was performed to observe the spatial colocalization of BrdU (in red) and γ -H2AX (in green), a novel biomarker for DNA double-strand breaks (Fig. 3G). The dropout of BrdU showed a much weaker γ -H2AX signal, leading us to conclude that the accumulation of DNA damage during the CIP process is directly caused by BrdU binding to the DNA. Using CHIP-seq analysis targeting the DNA double-strand damage marker γ -H2AX, a correlation was found between its level and the ATAC-seq for *Gata4* and *Sox17* (Fig. 3H). This suggests that the activation of the *Gata4* and *Sox17* genes is mediated through the DNA damage repair pathway.

Mechanism of DDR in promoting somatic cell reprogramming

ATM (Ataxia Telangiectasia Mutated) is a protein kinase that plays a critical role in the cellular response to DNA damage, particularly double-strand breaks (DSBs). When DSBs occur, the MRN complex (Mre11-Rad50-Nbs1) recognizes and binds to the site of break, recruiting ATM. Subsequently, ATM undergoes autophosphorylation, resulting in its full activation. Once activated, ATM phosphorylates several downstream targets involved in cell cycle checkpoint control, DNA repair, and apoptosis. Additionally, ATM activation is sustained by a phosphorylation-acetylation cascade. This cascade helps to maintain ATM activity and promote efficient DNA repair. Overall, ATM serves as a critical player in the cellular response to DNA damage, and its activation and downstream signaling are tightly regulated to ensure proper DNA repair and maintenance of genomic integrity [9].

In order to understand how DDR is involved in regulating somatic cell reprogramming, we hypothesize that DDR can upregulate the acetylation levels of genes at the site of DNA damage, thereby regulating gene expression,

as shown in Fig. 4A. Firstly, we found that inhibiting the ATM signaling pathway with an ATM inhibitor KU-55933 prevented BrdU from functioning (Fig. 4B, C). Furthermore, we revealed that *Gata4*, *Sox17* and *Sall4*'s H3K27ac and H3K9ac were significantly upregulated under the treatment of BrdU and I+B (Fig. 4D). To further verify this hypothesis, we increased the concentration of VPA (a histone deacetylase inhibitor) to further promote acetylation accumulation. We found that this significantly accelerated the process of chemical reprogramming, and its acceleration effect depended on the addition of BrdU (Fig. 4E, F).

Therefore, we believe that BrdU/IdU induces DNA damage by incorporating into the genome, and recruits ATM to the site of damage. Subsequently, the acetylation levels of gene loci at the site of DNA damage is upregulated through a series of phosphorylation and acetylation cascades, thereby altering chromatin accessibility and activating gene expression.

Combined with the results of RNA-seq, ATAC-seq, motif, γ -H2AX, H3K9ac, and H3K27ac analysis, we found that *Gata4* has a high degree of specificity under the treatment of BrdU/IdU (Figs. 2D,E, 3H, 4D, Additional file 1: Fig. S2A). Therefore, GATA4 may be one of the specific downstream factors of BrdU/IdU involved in somatic cell reprogramming. Further Cut & tag analysis showed that the binding sites of GATA4 were significantly increased under the treatment of BrdU/IdU, and the sites enriched with GATA4 were accompanied by higher chromatin accessibility (Fig. 4G), including pluripotency gene loci such as *Oct4*, *Sall4* as illustrated in Fig. 4H. This indicates that GATA4 can respond to DDR, thereby promoting somatic cell reprogramming.

BrdU/IdU creates a more open hypomethylated environment for the transformation of cell fate

BrdU and IdU have been demonstrated to play a crucial role in chemical reprogramming, exhibiting two distinct stages of action. Hence we postulate that BrdU not only activates intermediate XEN genes such as *Gata4*, but also potentially induces DNA demethylation, where DNA methylation is a major regulatory factor that limits GATA4 function [10]. To confirm this, we measured the methylation levels of GATA4 downstream gene *Aqp8* and XEN gene *Pth1r*. We observed a significant decrease in DNA methylation levels with BrdU/IdU treatment (Fig. 5A), indicating that BrdU/IdU not only activates gene expression but also induces DNA demethylation, thus creating a conducive environment for pluripotency network activation.

We also integrated the aspect of DNA damage repair and hypothesized that BrdU and IdU act through double-stranded break excision and new chain synthesis

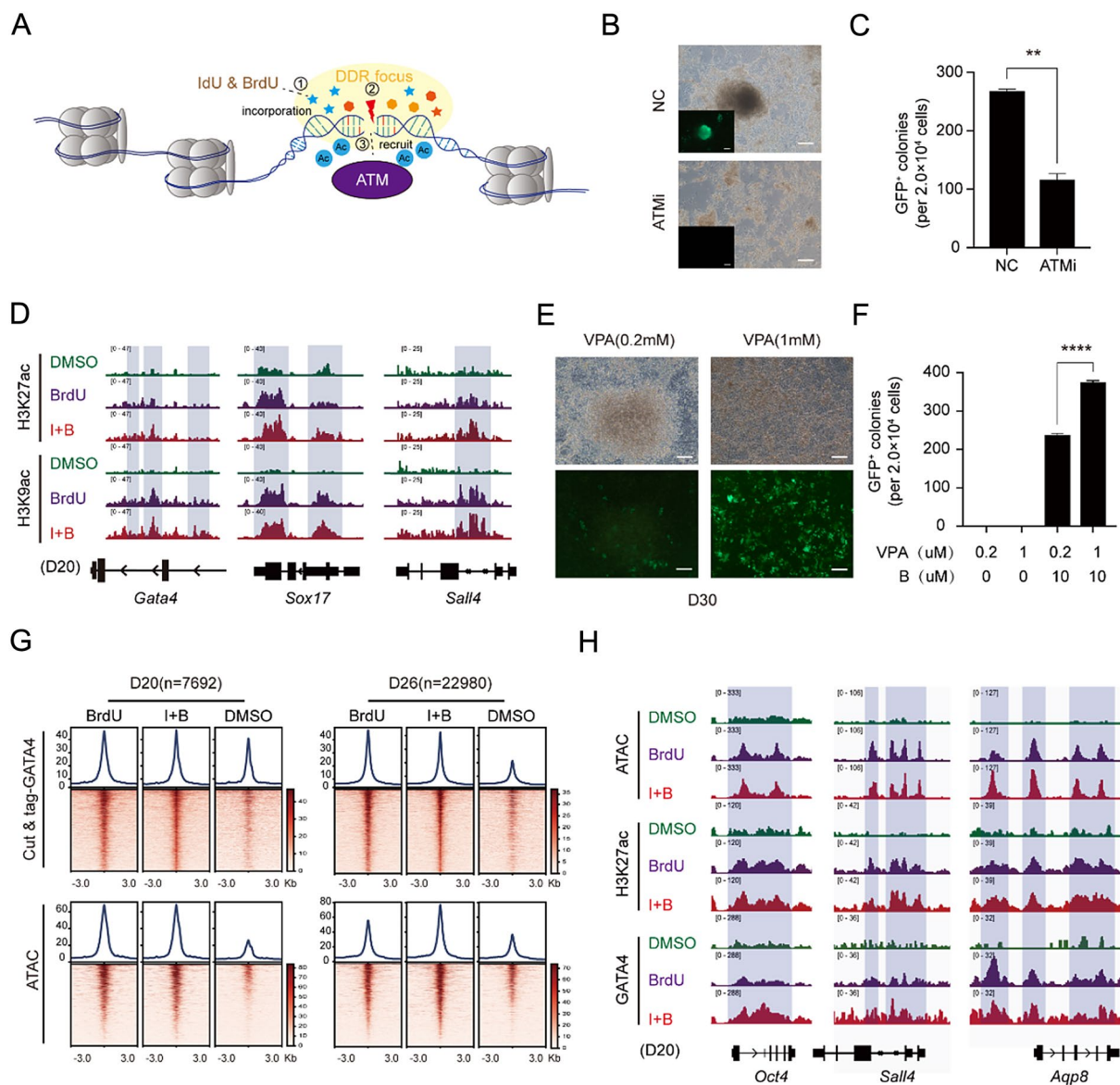


Fig. 4 Mechanism of DDR in promoting somatic cell reprogramming. **A** A model for DNA damage repair (DDR) mechanism during BrdU/IdU incorporation. **B** Morphological changes during induction of CIPSCs treated with or without ATM inhibitor (KU-55933). Scale bar 100 μ m. **C** Number of Oct4-GFP + CIPSC colonies generated under different treatment conditions. n = 3, **P < 0.01. **D** Representative *Gata4*, *Sox17* and *Sall4* peaks from H3K27ac Cut & tag aligned with H3K9ac Cut & tag signals for CIP at D20 with or without BrdU/IdU. **E** Morphological changes during induction of CIPSCs treated with different concentration of VPA. Scale bar 100 μ m. **F** Number of Oct4-GFP + CIPSC colonies generated under different treatment conditions. n = 3, ****P < 0.0001. **G** GATA4 Cut & tag analysis and ATAC-seq analysis for CIP at D20 and D26 under different treatment conditions. **H** Representative *Oct4*, *Sall4* and *Aqp8* peaks from ATAC-seq aligned with H3K27ac Cut & tag and GATA4 Cut & tag for CIP at D20 with or without BrdU/IdU

during the process of DNA damage repair. Since new chain synthesis occurs without DNA methylation marks, it leads to DNA demethylation. To validate this hypothesis, we inhibited DNMT using CM272 and SGI1027 to prevent the re-methylation of new chains. The results indicated that this approach significantly accelerated the process of chemical reprogramming

(Fig. 5B and C), and this effect was dependent on the presence of BrdU. To summarize, as the degree of BrdU damage increases, the frequency of double-stranded breakage and synthesis also increases, which leads to DNA demethylation, thereby creating a more open hypomethylated environment for the transformation of cell fate.

In the global genome methylation analysis using Gm-seq, we found that the methylation levels of CpG sites after treatment with BrdU, IdU, and I+B were mostly concentrated below 50%, while the distribution was more even in the untreated (DMSO) group, and in MEF it was biased towards above 50% (Fig. 5D). The average percentage of CpG sites with a methylated C base was below 30% after treatment with BrdU, IdU, and IB, 50% in the DMSO group, and 60% in MEF (Additional file 1: Fig. S4A). Furthermore, we used a circos plot to show the distribution of methylation density on chromosomes. The size of each bin was 1,000,000, and the number of methylated Cytosine in each bin with different sequence environments (CpG, CHG, CHH) was counted. The distribution density of CpG sites in different sequence environments throughout the genome showed that after treatment with BrdU/IdU, a significant global hypomethylation state was observed (Additional file 1: Fig. S4B).

We performed methylation differential analysis between treated and untreated cells with thymidine analogs. Interestingly, we found that the pluripotency related gene *Sall4*, *Oct4*, and *Prdm14* loci exhibited significant hypomethylation after thymidine analog treatment (Fig. 5E, G). This further confirms that BrdU/IdU creates a more open hypomethylated environment for the transformation of cell fate. In addition, we found that IdU exhibited a more widespread differential methylation distribution, followed by I+B, and BrdU exhibited the least, which may be related to its atomic size, where the atomic size of IdU > I+B > BrdU (Fig. 5F). Differences in atomic size lead to variations in the degree of DNA damage. Using IdU alone would cause severe damage, while using BrdU alone wouldn't have a significant impact. Therefore, I+B might be able to compensate for each other and generate a synergistic effect,

BrdU/IdU can participate in cell fate regulation with negligible mutations

As the substitute of thymidine, the safety of BrdU has always been a concern. Therefore, we performed mutation analysis on the samples treated with or without thymidine analogs and MEFs in our system, including SNP (Single Nucleotide Polymorphism), InDel

(insertion–deletion), CNV (Copy Number Variant), and SV (Structural variation). We found that the mutation level of BrdU/IdU-treated samples was almost the same as that of MEFs (Fig. 6A–E), while no treatment with thymidine analogs led to more mutations. In addition, we found that IdU had some insertion mutations. Overall, this suggests that the rational use of BrdU/IdU can participate in cell fate regulation with negligible mutations.

BrdU and IdU have functions in various reprogramming systems

To investigate whether the effects of BrdU and IdU are applicable in other reprogramming systems, we found that BrdU significantly improves reprogramming efficiency in the OKS system (Additional file 1: Fig. S5A–D, Fig. 7A, B). Additionally, we found that BrdU can substitute for OCT4 and play a critical role in the KS system (Additional file 1: Fig. S5E–G, Fig. 7A, B). Incorporation of BrdU not only improves reprogramming efficiency, but also accelerates reprogramming speed (Fig. 7C). Interestingly, by comparing the effects of BrdU on OKS, KS, and CIP (Fig. 7A, B), we found that their dependence on BrdU treatment time gradually increases. Specially, in OKS system, Oct4-GFP⁺ colonies reach their peak when BrdU is treated for 2 days. In KS system, Oct4-GFP⁺ colonies reach their peak when BrdU is treated for 6 days. Furthermore, in CIP system, 22 days of BrdU treatment is required to reach maximum number of Oct4-GFP⁺ colonies, as illustrated in Fig. 7B. This suggests that different degrees of DNA damage may be required to overcome various barriers to reprogramming.

We discovered that the addition of BrdU in the KS reprogramming system can specifically activate *Gata4*, *Sox17*, and certain developmental stem cell pluripotency regulating genes, as shown in Fig. 7D. Additionally, we observed that the most significant difference occurred at D12, which was precisely the time point when genes such as *Gata4* were highly expressed. We found that inhibiting the ATM signaling pathway with an ATM inhibitor also prevented BrdU from functioning in KS system (Fig. 7E). By using GATA4 Cut & tag, we discovered that GATA4-enriched peaks were accompanied by ATAC opening (Fig. 7F). By overexpressing *Gata4* in the KS system

(See figure on next page.)

Fig. 5 BrdU/IdU creates a more open hypomethylated environment for the transformation of cell fate. **A** The methylation patterns of *Aqp8* and *Pthr1* when treated with or without BrdU or IdU. **B** Morphological changes at D30 during induction of CiPSCs treated with DNMT inhibitor CM272 and SGI1027. **C** FSC analysis of Oct4-GFP colonies generated under different treatment conditions. **D** Histogram of percentage CpG methylation generated under different treatment conditions. **E** Methylation differential analysis between treated and untreated cells with IdU/BrdU. **F** Genomic panorama differential methylation regional distribution. From outermost to innermost: Outer circle represents chromosome ideogram; Second circle displays the methylation levels in the treatment group; Third circle shows gene density, where darker red indicates higher gene density; Fourth circle exhibits methylation differences between the treatment and control groups, with darker blue indicating greater differences; Innermost circle illustrates the methylation levels in the control group. **G** Differential DNA methylation gene and annotations

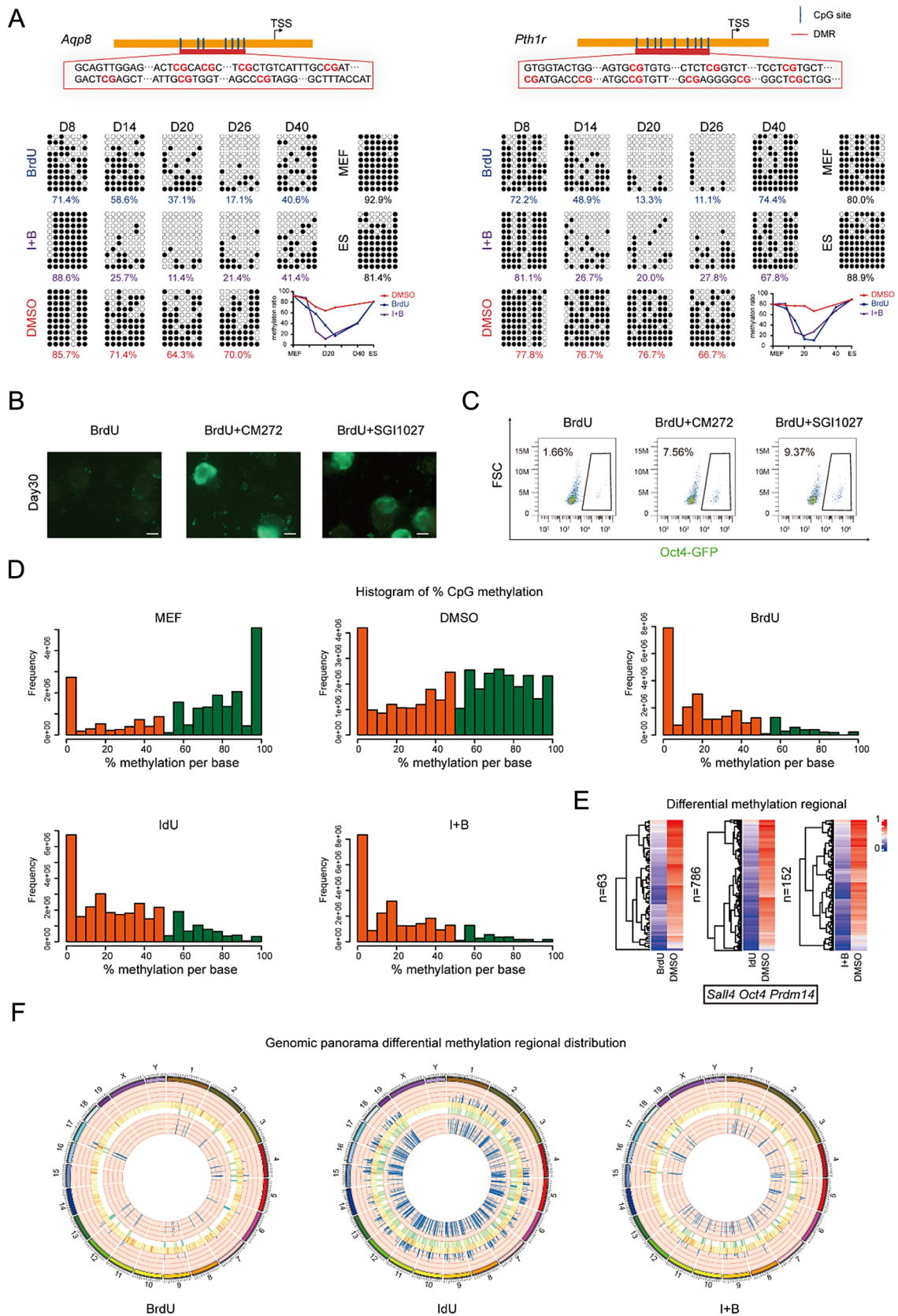


Fig. 5 (See legend on previous page.)

G

Differential DNA methylation gene and annotations

gene	chr	start	stop	CpGs	BrdU	DMSO	mean.meth.diff	prom	exon	intron	q.value	refseqID
Sall4	chr2	168757158	168758568	23	0.1583	0.6221	-0.4638	0	0	1	0.03896	XM_006500486
Pou5f1	chr17	35506192	35508189	94	0.2535	0.5506	-0.2970	1	1	1	0.00010	NM_013633
Pou5f1	chr17	35510080	35511929	51	0.2569	0.5920	-0.3350	0	1	1	0.00010	NM_001252452
gene	chr	start	stop	CpGs	IdU	DMSO	mean.meth.diff	prom	exon	intron	q.value	refseqID
Prdm14	chr1	13141710	13143325	35	0.2552	0.7100	-0.4548	0	1	1	0.02730	NM_001081209
Pou5f1	chr17	35507156	35507939	36	0.1439	0.4151	-0.2711	1	0	1	0.03732	NM_001252452
gene	chr	start	stop	CpGs	IdU+BrdU	DMSO	mean.meth.diff	prom	exon	intron	q.value	refseqID
Prdm14	chr1	13142149	13142690	27	0.0555	0.7530	-0.6975	0	1	0	0.00191	NM_001081209
Sall4	chr2	168757312	168758715	36	0.1767	0.7383	-0.5615	0	0	1	0.00212	XM_006500486
Pou5f1	chr17	35507156	35508189	48	0.1555	0.4330	-0.2775	1	0	1	0.00896	NM_001252452
Pou5f1	chr17	35510080	35513280	77	0.2341	0.5652	-0.3310	0	1	1	0.00001	NM_001252452

Fig. 5 continued

under the presence of BrdU, the reprogramming process can be accelerated and the cell state can be improved (Fig. 7G). Remarkably, our results bear a resemblance to the phenomenon observed in chemical reprogramming.

Discussion

In summary, we found that the use of the thymidine analogs BrdU/IdU causes epigenetic reshaping. This potentiates cells to enter a plastic state and accelerates the fate transition of cells by activating DNA damage repair and causing significant H3K27ac, H3K9ac and DNA demethylation. Additionally, the study discovered that the XEN state can be specifically activated after BrdU/IdU treatment. This suggests that XEN state may be a selective response behavior of cells to DNA damage repair, providing a shortcut for cells to overcome reprogramming barriers, and that different degree of DNA damage may be required to overcome various barriers to reprogramming. Furthermore, rational dosage of BrdU/IdU can participate in cell fate transition with negligible mutations.

Chemical small molecules interact in a complex and intricate manner during somatic cell reprogramming, forming a complex network of cell reprogramming. These molecules may exert synergistic or antagonistic effects on each other, with their influence varying across different stages of reprogramming, modulated by factors such as timing, concentration, and cell type. Investigating these interactions, and their comprehensive impact on the reprogramming process, is vital for advancing

our understanding of cellular reprogramming mechanisms, enhancing pluripotency induction efficiency, and developing therapeutic applications. Our research on thymidine analogs sheds light on their synergistic effects with molecules targeting histone acetylation and DNA methylation.

In our study, through gene expression patterns, immunofluorescence of gH2AX (a marker for DNA double-strand breaks), ChIP-seq, and Comet assays, we confirmed that prolonged incorporation of BrdU/IdU causes a more severe form of DNA damage. Importantly, we did not observe any significant mutations, ruling out a role for non-homologous end joining (NHEJ). Therefore, we concluded that the homologous recombination repair (HRR) pathway becomes more dominant with prolonged incorporation of BrdU/IdU. This emphasizes the importance of genomic stability in the reprogramming process and its potential impact on the overall efficacy and safety of iPSC generation. DNA single-strand break causes “Discordant Transcription through Repair (DiThR)” [8], while DNA double-strand break results in “Epigenetic Reshaping through Damage (ERD)”. However, it is worth to investigate the roles of other DNA repair pathways, such as MMR, NHRJ and NER, in the process of cellular fate transition. Additionally, not all of thymidine analogs are equally effective in somatic reprogramming, and their specific mechanisms of action remain to be fully explored. Interestingly, although BrdU and IdU are structurally very similar, with only minor differences

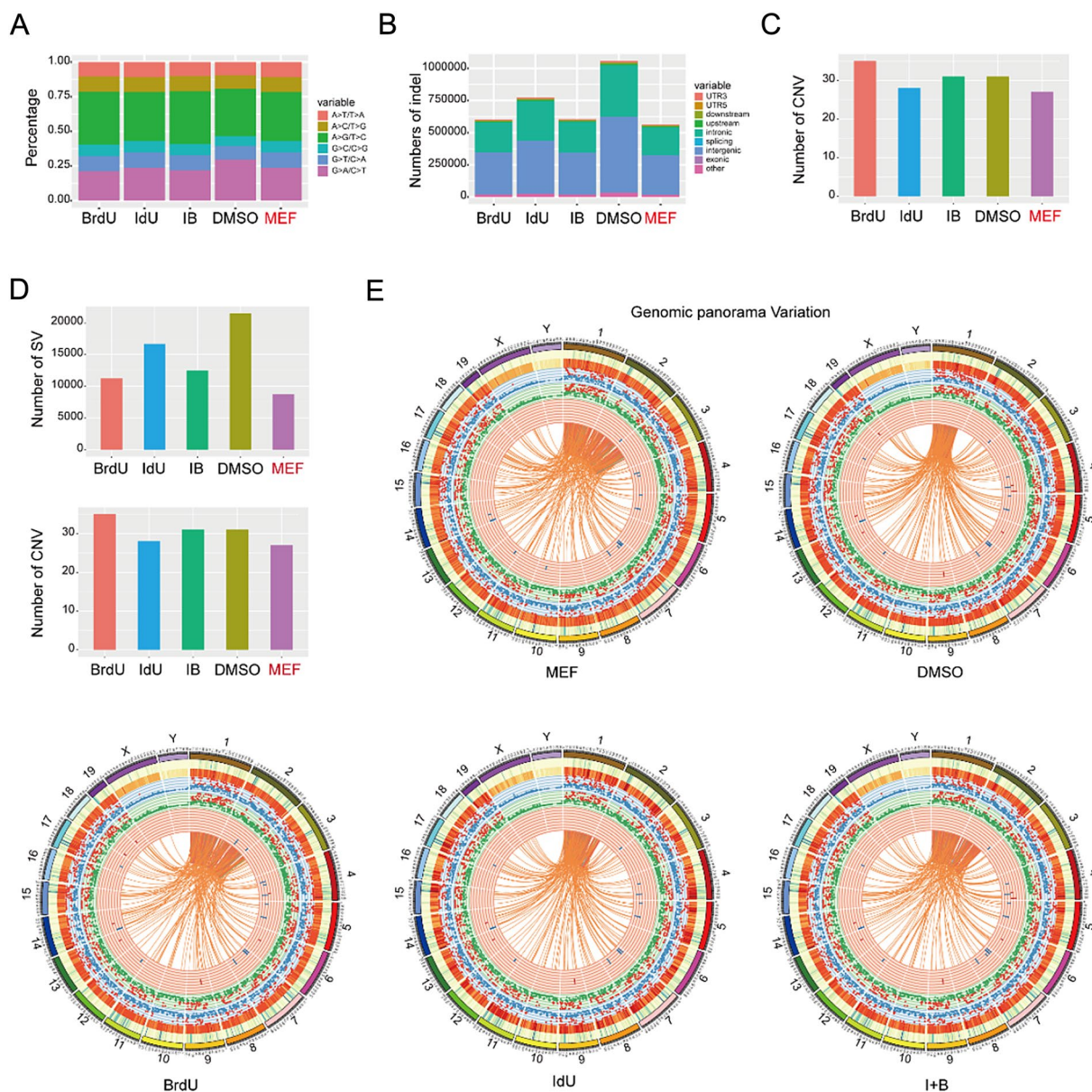


Fig. 6 BrdU/IdU can participate in cell fate regulation with negligible mutations. **A** Mutation analysis-SNP (Single Nucleotide Polymorphism) on the samples treated with or without BrdU/IdU and MEFs. **B** InDel (insertion-deletion) analysis. **C** CNV (Copy Number Variant) analysis. **D** SV (Structural variant) analysis. **E** Genomic panorama variation analysis. The outermost circle displays chromosomal information: lengths and karyotypes of each chromosome. The second circle shows gene density as a blue heatmap, with darker shades indicating higher gene numbers in that chromosomal area. The third circle exhibits a red heatmap of sequencing coverage, with deeper colors showing higher average sequencing coverage within the chromosomal region. The fourth circle is a scatter plot of SNP density against a light blue background: red dots above the average density, and blue dots below. The fifth circle represents InDel density with a similar scatter plot: red dots above the average density, and green dots below. The sixth circle displays CNV distribution as a line plot: upward lines indicate amplifications (red), and downward lines indicate deletions (blue). The seventh circle signifies SV (Structural Variations) in different colors: green for insertions, red for deletions, blue for duplications, purple for inversions, and orange for translocations

in their chemical structure, the difference in atomic size may lead to a more pronounced impact of IdU on the entire genome. This difference in impact may also explain why other thymidine analogs are not effective.

It is worth to consider that the XEN-like intermediate state is not unique to chemical reprogramming, as cells may require multiple intermediate states to reach their final fate during the process of fate transition.

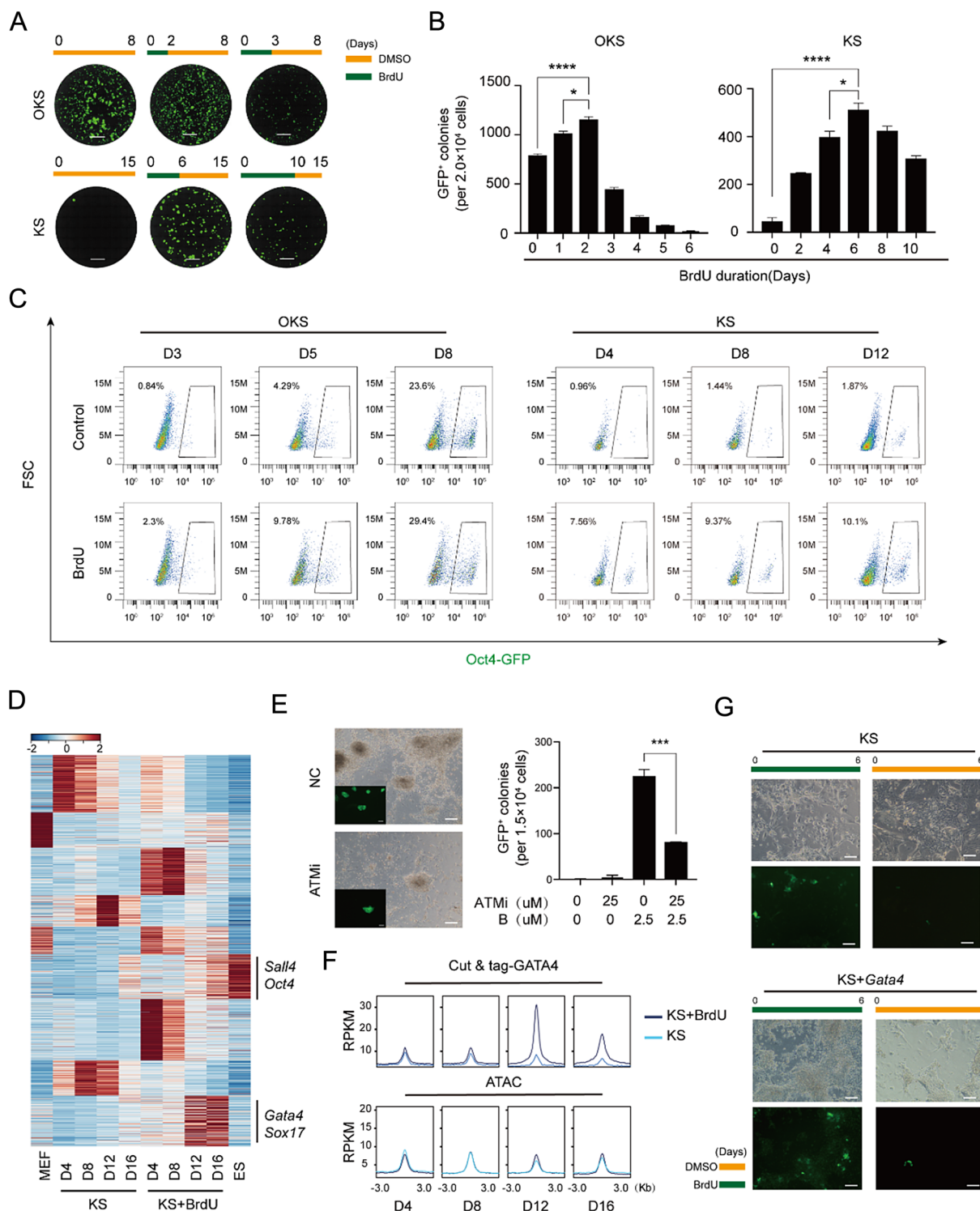


Fig. 7 BrdU and IdU have functions in various reprogramming systems. **A** Images of GFP + colonies taken by fluorescence microscope in situ. Scale bar 5 mm. **B** Number of Oct4-GFP colonies generated under indicated conditions. n=3, *P<0.05, ****P<0.0001. **C** FSC analysis of Oct4-GFP colonies generated under different treatment conditions. **D** Heatmap of RNA-seq for KS at D4, 8, 12 and 16 with or without BrdU or IdU. **E** Morphological changes during induction of KS treated with ATM inhibitor and corresponding statistical analysis. Scale bar 50 μm. n=3, ***P<0.001. **F** GATA4 Cut & tag analysis and ATAC-seq analysis for CIP at D4, 8, 12 and 16 under different treatment conditions. **G** Morphological changes during induction of KS when overexpressing Gata4 in the system. Scale bar 50 μm

Reprogramming barriers may limit the process of fate transition, and thus cells may choose different intermediate states as shortcuts. *Gata4* and *Sox17*, among others, are representative of XEN states that have been selected as transitional states that facilitate cell fate transition because they promote both dedifferentiation and redifferentiation of cells, and play essential roles in embryonic development and cell regeneration. Moreover, DNA damage repair may also drive cells into a plastic state. When DNA damage occurs, cells need to repair the damage to maintain genome integrity. DNA repair processes may lead to reprogramming events, resulting in cells entering a plastic state and providing an alternative pathway for cell fate transition.

The long-term use of thymidine analogs can have diverse effects on cells and organisms, contingent on the specific analog used, the dosage, and the exposure duration. For example, in our study, we observed that excessive application of BrdU can lead to cell death. However, this effect can be avoided by maintaining the dosage within a controlled range, thus effectively facilitating the reprogramming process.

There is a significant issue with current induced Pluripotent Stem Cells (iPSCs) is epigenetic variation, including abnormal DNA methylation, chromatin conformation changes, or imbalances in epigenetic modification patterns. These variations can impact the quality and stability of iPSCs. Therefore, understanding epigenetic variation and reshaping in iPSC induction is crucial for ensuring their quality, stability, and safety. In our research, we discovered a new phenomenon of epigenetic reshaping through DNA Damage Response (DDR) and interactions between genetics and epigenetics, which can optimize iPSC generation and enhance their potential in regenerative medicine and disease modeling.

Materials and methods

Mice

Beijing Vital River Laboratory supplied the 129 Sv/Jae and ICR mice, while the Jackson Laboratory provided the Oct4-GFP(OG2) transgenic allele-carrying mice (CBA/CaJ X C57BL/6J). All animal experiments were carried out in compliance with the Animal Protection Guidelines of Guangzhou Institutes of Biomedicine and Health, located in Guangzhou, China and affiliated with the Chinese Academy of Sciences.

Cell culture

Mouse embryonic fibroblasts (MEFs)

To isolate Mouse Embryonic Fibroblasts (MEFs), E13.5 embryos were dissected and all internal organs, head, limbs, and tails were removed and discarded. The remaining tissue was sliced into small pieces and dissociated

with a digestive solution (consisting of 0.25% trypsin and 0.05% trypsin in a 1:1 ratio from GIBCO) for 15 min at 37 °C to obtain a single-cell suspension. The cells from each embryo were then plated onto a 6-cm culture dish coated with 0.1% gelatin and cultured in DMEM (Hyclone) supplemented with 10% FBS (GIBCO), 1% GlutaMAX (GIBCO), and 1% NEAA (GIBCO), referred to as fibroblast medium.

mESCs

The feeder-free and serum-free ESC medium used for maintaining mESCs and CiPSCs consisted of DMEM supplemented with 1%N2(GIBCO), 2%B27 (GIBCO), LIF, 1% GlutMax, 1% NEAA, 0.1 mM beta-mercaptoethanol, and 3 μM CHIR99021 and 1 μM PD0325901. Additionally, the cell lines were tested negative for mycoplasma using Lonza's Kit (LT07-318).

Culture medium preparation

Chemical reprogramming stage 1 iCD1 medium (Chen et al., 2011a) comprises vitamin C (50 mg/ml), bFGF (10 ng/ml), and CHIR99021 (3 mM), along with several small molecules, including Brdu (10 μM), RepSox (5 mM), FSK (10 mM), VPA (0.1 mM), AM580 (0.05 mM), EPZ5676 (5 mM), DZNeP (0.05 mM), SGC0946 (5 mM), BMP4 (10 ng/ml), and Capmatinib (5 μM).

Chemical reprogramming stage 2 DMEM (high glucose, Hyclone), 1% N2 (GIBCO), 2% B27 (GIBCO), 1% GlutMax (GIBCO), 1% NEAA (GIBCO), 1% sodium pyruvate (GIBCO) 3 mM CHIR99021, 1 mM PD0325901 and 1000 U LIF.

KS reprogramming stage 1 iCD1 medium, Repsox(5 μM), FSK(10 μM), SGC0946(1 μM), GSK-LSD1(1 μM) and BrdU(2.5 μM).

KS reprogramming stage 2 iCD1 medium, Repsox(5 μM), FSK(10 μM), SGC0946(1 μM), GSK-LSD1(1 μM).

KS reprogramming stage 3 DMEM (high glucose, Hyclone), 1% N2 (GIBCO), 2% B27 (GIBCO), 1% GlutMax (GIBCO), 1% NEAA (GIBCO), 1% sodium pyruvate (GIBCO) 3 mM CHIR99021, 1 mM PD0325901 and 1000 U LIF.

Chemical induction of iPSCs from mouse fibroblasts

The 12-well plates were pre-coated with 0.1% gelatin and seeded with MEFs at a density of 20,000 cells per well in 10% fibroblast medium. The following day, Stage 1 chemical reprogramming medium was added and refreshed daily. After day 22, the medium was changed to Stage 2 chemical reprogramming medium. On day 40, the

Oct4-GFP colonies were either counted or detected via FACS.

***Klf4* and *Sox2* mediated MEF reprogramming**

Plasmids carrying murine *Klf4*, *Sox2* cDNA were purchased from Addgene, and the fragments were then ligated to the PMX vector to obtain a recombinant plasmid. Plat-E cells were transfected with individual plasmids with the Ca₃(PO₄)₂-method, a modified calcium phosphate transfection method was conducted as follows: for each factor, 1068 ml ddH₂O, 25 mg plasmid, 156.25 ml 2 M CaCl₂, 1.25 ml 2.3 HBS (total 2.5 ml) were added to a 15 ml tube. Mix it vigorously after adding 2.3 HBS, and incubate for 5 min at room temperature. Then, gently transfer the 2.5 ml mixture into the Plat-E cell dish with 7.5 ml fresh medium. After 48 h of transfection, the virus supernatant was collected and used to infect cells. Approximately 15,000 cells were seeded per well in 24-well plates containing DMEM medium supplemented with 10% FBS, and were cultured for 24–12 h prior to MEF infection. The *Klf4* and *Sox2* virus solution was filtered using a 0.45 μm filter, and then 8 μg/ml of polybrene was added for cell infection. The cells were subjected to a second infection after 24 h. After 48 h of MEF cell infection, the cells were transferred to induction medium, with the day of medium change being designated as day 0. KS reprogramming stage 1 medium was used for the first 6 days, stage 2 medium was used from day 6 to day 10, and stage 3 medium was used from day 10 onwards. The medium was changed every 24 h.

Immunofluorescence staining

The coverslip-grown cells were washed twice with PBS and fixed using 4% paraformaldehyde (PFA) at room temperature for 30 min. Next, the samples were treated with 1N HCl on ice for 10 min, followed by 2N HCl at room temperature for 10 min and then 2N HCl at 37 °C for 20 min. After washing three times with PBS for 5 min each, the cells were permeabilized with 0.1% Triton X-100 for 30 min. Subsequently, the cells were blocked with 3% BSA for an hour and washed three times with PBS for 5 min each while shaking. The cells were then incubated overnight at 4 °C with primary antibodies, diluted in 3% BSA (1:250) in PBS. After four washes with PBS for 5 min each while shaking, the cells were incubated for an hour at room temperature with secondary antibodies, diluted in 3% BSA (1:200) in PBS. Later, the cells were incubated in DAPI for 1 min, washed twice with PBS, and finally, the coverslips were mounted on the slides for observation under a confocal microscope (Andor Dragonfly 200). The following antibodies were used: anti-BrdU (Sigma), anti-γ-H2AX (Abcam), secondary antibodies Alexa Fluor 568

goat anti-mouse IgG (Invitrogen) and Alexa Fluor 488 goat anti-mouse IgG (Invitrogen).

Comet assay

During the chemical reprogramming process, the control and experimental group cells at D14, D16, D24, and D30 were digested with 0.25% trypsin into single cells, and resuspended in PBS at a concentration of 10,000 cells/ml. The Oxiselect Comet Assay Kit (Cell-Biolabs) was used as a reference for the experiment. The cell suspension was mixed with agarose, and then incubated in the dark at 4 °C for 15 min. The mixture was placed on a slide and immersed in lysis buffer, then incubated in the dark at 4 °C for 60 min. The slide was transferred to an alkaline solution, and incubated in the dark at 4 °C for 30 min. The slide was then placed into an electrophoresis chamber, and alkaline electrophoresis buffer was added under the conditions of 300 mA and 15 min. After electrophoresis, the slide was transferred to water and soaked for 2 min. Then it was transferred to 70% ethanol and soaked for 5 min. The slide was placed in a 37 °C oven overnight. Vista Green DNA dye was added for 10 min, and the cells were observed and recorded using an inverted fluorescence microscope (ZEISS). The software comet score was used to analyze the tail moment of the cells.

FACS analysis

The cells were rinsed with PBS and then digested with 0.25% trypsin at 37 °C for 5 min. The digestion was stopped by adding 10% FBS DMEM medium. After filtering with a sieve and centrifuging at 250g for 5 min, the supernatant was discarded and the cells were resuspended in PBS at a concentration of 100,000 to 1,000,000 cells/ml. The percentage of positive cells was detected using a BD Accuri™ C6 Plus flow cytometer. The flow cytometry data was analyzed using FlowJo 7.6.1.

Bisulfite genomic sequencing

Extract DNA template. After centrifuging the cells, add 600 μl of Nuclei Lysis Solution (Promega) and mix by pipetting. Add 200 μl of Protein Precipitation Solution (Promega) and mix by pipetting. Let it sit on ice for 10 min. Centrifuge at 12,000 rpm for 10 min. Take the supernatant, add 600 μl of isopropanol, and let it sit on ice for 10 min. Centrifuge at 12,000 rpm for 5 min. Discard the supernatant, add 600 μl of 70% ethanol. Heat in a metal bath at 60 °C for 5 min. Add RNAase-free water and heat in a metal bath at 60 °C for 30 min. Take 2 μg of template DNA and bisulfite treat according to the EpiTect Bisulfite Kit (QIAGEN) manual. Amplify the *Aqp8* and *Pth1r* promoter regions by PCR. Clone and sequence using the pMD18-T vector (TaKara).

RNA-seq and data analysis

Cells were treated with TRIzol to extract total RNA, which was then converted into cDNA using ReverTra Ace (Toyobo) and oligo-dT (Takara). The resulting cDNAs were analyzed using Premix Ex Taq (Takara) in qPCR experiments. For library construction, the TruSeq RNA Sample Prep Kit (RS-122-2001, Illumina) was employed, and RNA-seq was performed using the Miseq Reagent Kit V2 (MS-102-2001, Illumina).

The original sequencing data was quality-controlled using FASTQC, and low-quality bases and sequencing adapters were removed using trim_galore. HISAT2 [11] was used to align the filtered clean reads with the mouse mm10 reference genome. Samtools [12] was used to filter out unaligned or unpaired sequencing fragments and obtain the bam file. featureCounts [13] software was used to quantify gene expression levels. EdgeR was used to identify differentially expressed genes. ClusterProfiler [14] was used to perform GO or KEGG functional enrichment analysis to determine the molecular functions of the differentially expressed genes.

ATAC-seq

The ATAC-seq procedure was carried out following the previously published protocols [15, 16] and TruePrep DNA Library Prep Kit V2 for Illumina. Briefly, 50,000 cells were washed with 50 ml of cold PBS and suspended in 50 ml of lysis buffer containing 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, and 0.2% (v/v) IGEPAL CA-630. The cell suspension was centrifuged at 500 g for 10 min at 4 °C, followed by the addition of 50 ml of transposition reaction mix from the TruePrep DNA Library Prep Kit V2 for Illumina. The resulting samples were subjected to PCR amplification and incubated at 37 °C for 30 min. VAHTS DNA Clean Beads (Vazyme #N411) were used for purification and recovery of the DNA fragments. The purified product was amplified. VAHTS DNA Clean Beads were used for further purify and separate 200–700 bp fragments. The ATAC library was finally sequenced on a NextSeq 500, using a NextSeq 500 High Output Kit v2 (150 cycles) (FC-404-2002, Illumina), following the manufacturer's instructions.

Cut&tag

According to the protocol of the hyperactive universal Cut&Tag assay Kit for Illumina, library construction was performed. 100,000 cells were collected and incubated with processed ConA beads. The sample was incubated with primary antibody overnight at 4 °C, followed by incubation with secondary antibody at room temperature for 1 h. The sample was then incubated with transposase pA/G-Tnp at room temperature for 1 h, and the DNA was fragmented for 1 h before extraction. After

amplifying DNA fragments according to concentration, VAHTS DNA Clean Beads (Vazyme #N411) were used for purification and recovery of the DNA fragments.

ATAC-seq and Cut&tag bioinformatics analysis

The original sequencing data was quality-controlled using FASTQC, and low-quality bases and sequencing adapter sequences were removed using trim_galore. Bowtie2 [17] was used to align high-quality sequencing fragments with the mouse mm10 reference genome. Samtools was used to retain only paired and uniquely aligned sequences, exclude mitochondrial sequence fragments, and obtain the bam file. Picard (<http://broadinstitute.github.io/picard/>) was used to remove duplicate fragments caused by PCR amplification during library preparation. DeepTools [18] was used to convert the bam file to a bw file, and the results were visualized using IGV (<https://igv.org/>). The peak calling software MACS2 [19] was used to detect enrichment regions of open chromatin or DNA fragments. HOMER was used to identify transcription factor binding motifs enriched in chromatin regions. CHIPseeker [20] was used to annotate genomic features (such as genes, promoters, enhancers, and transcription factor binding sites) of peaks intervals. Bedtools [21] was used to analyze differential binding sites, and deeptools was used to generate signal matrix files and visualize the results. Other analyses were performed using glbase [22].

GM-seq

The Gm-seq procedure was carried out following the previously published protocols in cooperation with Gene plus company [23]. We utilized Hieff NGS[®] Ultima Pro DNA Library Prep Kit for Illumina (Yeason, cat. on. 12201ES96) for end repair, A-tailing, and adaptor ligation. After magnetic beads were purified, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) were oxidized to 5-acylcytosine (5fC) or 5-carboxyl cytosine (5caC) by TET enzyme, then 5fC or 5caC was treated with pyridine borane and reduced to dihydro-uracil (DHU). Finally, PCR amplification is performed and barcode is introduced to obtain the sequencing library. After going through this series of processes the methylated cytosine would be identified as thymine (T) by MGI T7 sequencing platform. The samples for whole genome sequencing directly were also going through the same procedures that were used for GM-seq but without oxidation and pyridine borane reduction preparation. Whole genome analyses were performed by the sequencing library we constructed above.

GM-seq bioinformatics analysis

The raw fastq and bam files were evaluated using BWA and SAMtools software. Alignment and analysis of DNA

sequencing data were performed utilizing asTair tools in automated mode from raw fastq files to finalized genotyping data. This pipeline utilized GATK tools and ExomeDepth tools to perform base quality score, genotype calling and variant annotation, including SNP and Indel. And ichorCNA were employed to analyze genome-wide CNV and calculate insert size. Sequencing data were analyzed by three computational pipelines in order to understand the reliability of the detection of fungi, bacteria, and viruses: Kraken, Kaiju and SNAP. BLAST was used to compare the sequencing results with the high-risk microorganisms. For methylation analyses, diagrams for CpG coverage were generated using methylKit 1.4.0. The GC bias plot was generated using Picard's CollectGCBiasMetrics. Correlation analysis was performed using methylKit version 1.4.0 with default settings except for a minimum coverage threshold of one read.

Quantification and statistical analysis

Sample size was not predetermined using statistical methods, and the experiments were not randomized. The investigators were not blinded during the experiment or outcome assessment. The data was presented as mean \pm s.d. Statistical analysis was conducted using either two-tailed unpaired student's t-test or one-way ANOVA with GraphPad Prism 6, and a P-value of <0.05 was considered statistically significant. The level of significance was indicated as $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$. The relevant figure legends provided information on the statistical test, precise P-Value, exact sample sizes, and independent experiments.

Abbreviations

XEN	Extraembryonic endoderm
DDR	DNA damage repair
ERD	Epigenetic reshaping through damage
CiP	Chemical induction of pluripotency
CiPSC	Chemical induction pluripotent stem cell
CADs	Chromatin accessibility dynamics
BER	Base excision repair
NER	Nucleotide excision repair
HRR	Homologous recombination repair
MMR	Mismatch repair
NHEJ	Non-homologous end joining repair
SNP	Single Nucleotide Polymorphism
InDel	Insertion–deletion
CNV	Copy Number Variant
SV	Structural variation
DiThR	Discordant Transcription through Repair

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13578-024-01192-x>.

Additional file 1: Figure S1. A. Schematic diagram of the induction of stage1. B. Morphological changes at barriers1 under indicated conditions. Scale bar 100 μ m. C. Images of GFP+ colonies taken by fluorescence

microscope in situ. Scale bar, 5 mm. D. Morphological changes at barriers1 treated with BrdU, EdU or 5Aza. E. Number of *Oct4*-GFP + CiPSC colonies generated under BrdU, EdU or 5Aza. $n = 3$, $** * P < 0.001$. F. Morphological changes at distinct time points during induction of CiPSCs treated with BrdU, 1 + B (IdU + BrdU) or B + E (BrdU + EdU). Scale bar 100 μ m. G. Number of *Oct4*-GFP + CiPSC colonies generated under BrdU, 1 + B or B + E. $n = 3$, $* P < 0.05$, $** P < 0.01$. **Figure S2.** A. Representative fibroblast genes, XEN genes and pluripotency genes peaks from RNA-seq for CiP at D8, 14, 20 and 26 with or without BrdU or IdU. B. PCA analysis for RNA-seq data from CiP under indicated conditions. C. Heatmap of XEN genes related RNA-seq for CiP at D8, 14, 20 and 26 with or without BrdU or IdU. D. The R2 correlation coefficient matrix of all versus all RNA-seq datasets as indicated. **Figure S3.** A. Venn diagrams for overlapping genes in the downregulated groups between BrdU, IdU, and BrdU + IdU versus the control. B. Enriched GO functions in upregulated groups. C. Representative DNA repair genes from RNA-seq for CiP. D. qPCR analysis of the representative DNA repair genes. **Figure S4.** A. M-bias plot analysis. B. Genomic panorama DNA methylation. Outermost circle: Chromosome karyotype. Second circle: Methylation levels at various genomic positions. The horizontal axis represents genome positions, and the vertical axis represents methylation levels (ranging from 0 to 1, inward to outward direction). Red indicates CpG methylation, blue represents CHG methylation, and green signifies CHH methylation. Third circle: Gene density, where darker shades of blue indicate higher density. Innermost circle: Number of methylation sites at different genomic positions. The horizontal axis represents genome positions, and the vertical axis represents the count of methylation sites (increasing from inward to outward). Red indicates CpG sites, blue represents CHG sites, and green represents CHH sites. **Figure S5.** A. Schematic diagram of the induction of iPS from MEFs with OKS. B. Images of GFP + colonies taken by fluorescence microscope in situ. Scale bar, 5 mm. C. Number of *Oct4*-GFP + CiPSC colonies generated under indicated conditions. D. Morphological changes at OKS treated with different concentrations BrdU. $n = 3$, $* P < 0.05$, $*** P < 0.001$. E. Schematic diagram of the induction of iPS from MEFs with KS. F. Morphological changes at KS treated with or without BrdU. G. Number of *Oct4*-GFP + CiPSC colonies generated under indicated conditions. R, Repsox, FFSK, S, SGC0946, G, GSK-LSD1. $n = 3$, $** P < 0.01$, $*** P < 0.001$ and $**** P < 0.0001$.

Additional file 2: Table S1. Key resources table. Including: antibodies, chemicals, critical commercial assays, deposited data, experimental models, software and algorithms, primer.

Acknowledgements

We would like to thank the support from the Guangzhou Regional Center of Bioscience Instrument of the Chinese Academy of Sciences.

Author contributions

JL and CL initiated the project and designed the experiments. CL performed the chemical reprogramming experiments. CL conducted and performed ATAC-seq, RNA-seq, Cut & tag library construction with XX. SC and CL performed bioinformatics analysis. XX performed OKS and KS reprogramming experiments with CL. XX performed bisulfite sequencing PCR. CL performed GM-seq with Gene plus company. JL, DP and HW supervised the whole study. HW supervised the whole study and wrote the manuscript. JL conceived the whole study and approved the final version.

Funding

This research was supported by grants from the National Key R&D Program of China (2018YFE0204800 [J.L.]), National Natural Science Foundation of China (32370791 [J.L.], 32022019 [J.L.], U20A2013 [C.Z.], 31970681 [J.L.], 32100594 [C.Z.], 32200695 [S.Y.]), Guangdong Basic and Applied Basic Research Foundation (2020A1515110122 [L.W.], 2021A1515111156 [S.Y.], 2022A1515012267 [S.Y.]), Science and Technology Planning Project of Guangdong Province (2020B1212060052 [J.L.], 2022B1212010010 [Y.L.]), Health@InnoHK Program launched by the Innovation Technology Commission of the Hong Kong SAR, P.R. China, and the China Postdoctoral Science Foundation (2022M713172 [S.Y.]).

Availability of data and materials

The datasets generated during the current study are available in the National Center for Biotechnology Information under the accession number PRJCA019276. The key resources are available in Additional file 2: Table S1.

Declarations**Ethics approval and consent to participate**

Not applicable for that section.

Consent for publication

Not applicable for that section.

Competing interests

The authors declare no competing financial interests.

Author details

¹School of Life Sciences, University of Science and Technology of China, Hefei 230026, China. ²Center for Cell Lineage and Development, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China. ³CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China. ⁴Centre for Regenerative Medicine and Health, Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong SAR, People's Republic of China. ⁵University of Chinese Academy of Science, Beijing 100049, People's Republic of China. ⁶Key Laboratory of Biological Targeting Diagnosis, Therapy and Rehabilitation of Guangdong Higher Education Institutes, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou 510000, China.

Received: 3 November 2023 Accepted: 4 January 2024

Published online: 16 January 2024

References

- Goz B. The effects of incorporation of 5-halogenated deoxyuridines into the DNA of eukaryotic cells. *Pharmacol Rev.* 1977;29:249–72.
- Zhao Y, et al. A XEN-like state bridges somatic cells to pluripotency during chemical reprogramming. *Cell.* 2015;163:1678–91.
- Long Y, et al. Bromodeoxyuridine promotes full-chemical induction of mouse pluripotent stem cells. *Cell Res.* 2015;25:1171–4.
- Ping W, et al. Genome-wide DNA methylation analysis reveals that mouse chemical iPSCs have closer epigenetic features to mESCs than OSKM-integrated iPSCs. *Cell Death Dis.* 2018;9:187.
- Cao S, et al. Chromatin accessibility dynamics during chemical induction of pluripotency. *Cell Stem Cell.* 2018;22:529–42.
- Fu H, et al. Dynamics of telomere rejuvenation during chemical induction to pluripotent stem cells. *Stem Cell Rep.* 2018;11:70–87.
- Ye J, et al. Pluripotent stem cells induced from mouse neural stem cells and small intestinal epithelial cells by small molecule compounds. *Cell Res.* 2016;26:34–45.
- Desai RV, et al. A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. *Science.* 2021;373: eabc6506.
- Blackford AN, Jackson SP. ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Mol Cell.* 2017;66:801–17.
- Oda M, et al. DNA methylation restricts lineage-specific functions of transcription factor Gata4 during embryonic stem cell differentiation. *PLoS Genet.* 2013;9: e1003574.
- Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12:357–60.
- Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
- Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30:923–30.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284–7.
- Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol.* 2015;109:21–9.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods.* 2013;10:1213–8.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
- Ramirez F, Dunder F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 2014;42:W187–191.
- Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9:R137.
- Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* 2015;31:2382–3.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
- Hutchins AP, Jauch R, Dyla M, Miranda-Saavedra D. glbase: a framework for combining, analyzing and displaying heterogeneous genomic and high-throughput sequencing data. *Cell Regen.* 2014;3:1.
- Chen X, et al. Identification of DNA methylation and genetic alteration simultaneously from a single blood biopsy. *Genes Genomics.* 2023;45:627–35.

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

