Identifying regulators of parental imprinting by CRISPR/Cas9 screening in haploid human embryonic stem cells

Shiran Bar1, Dan Vershkov1, Gal Keshet1, Elyad Lezmi1, Naama Meller1, Atilgan Yilmaz1, Ofra Yanuka1, Malka Nissim-Rafinia2,3, Eran Meshorer2,3, Talia Eldar-Geva4,5 & Nissim Benvenisty1

In mammals, imprinted genes are regulated by differentially methylated regions (DMRs) that are inherited from germ cells, leading to monoallelic expression in accordance with parent-of-origin. Yet, it is largely unknown how imprinted DMRs are maintained in human embryos despite global DNA demethylation following fertilization. Here, we explored the mechanisms involved in imprinting regulation by employing human parthenogenetic embryonic stem cells (hpESCs), which lack paternal alleles. We show that although global loss of DNA methylation in hpESCs affects most imprinted DMRs, many paternally-expressed genes (PEGs) remain repressed. To search for factors regulating PEGs, we performed a genome-wide CRISPR/Cas9 screen in haploid hpESCs. This revealed ATF7IP as an essential repressor of a set of PEGs, which we further show is also required for silencing sperm-specific genes. Our study reinforces an important role for histone modifications in regulating imprinted genes and suggests a link between parental imprinting and germ cell identity.
Parental imprinting is a unique epigenetic phenomenon in mammals, involving a group of genes that are only expressed from one parental allele, while the other allele is silenced. Proper expression of imprinted genes is essential for mammalian development, as uniparental embryos having a maternal-only (parthenogenetic) or paternal-only (androgenetic) genomes die during gestation. This notion established imprinting as being the block for asexual reproduction in mammals. Moreover, inappropriate expression of imprinted genes leads to various developmental disorders and is also involved in cancer progression.

The monoallelic expression of imprinted genes is regulated by differentially methylated regions (DMRs) that are established in the germline. Shortly after fertilization, the mammalian genome is extensively demethylated and by the time it reaches the blastocyst stage, most parental DNA methylation patterns are erased. Nevertheless, imprinted germline DMRs are exceptionally maintained during this time, preserving allelic parental identity. During early embryogenesis, “secondary” DMRs may appear at imprinted genes, while other “transient” DMRs disappear within days after fertilization. Yet, the factors regulating the maintenance of imprinting during human embryogenesis are mostly unknown.

Multi-locus imprinting disturbances (MLID) result in DNA hypomethylation of several imprinted loci (HIL). This phenomenon was identified in patients diagnosed with imprinting diseases which are caused by aberrant DNA methylation, such as Beckwith-Wiedemann and Russell-Silver syndromes. It is hypothesized that MLID patients carry mutations in genes involved in imprinting regulation, however for most cases, the identity of such factors has yet to be identified. MLID occurs in patients diagnosed with Transient-Neonatal Diabetes Mellitus (TNDM) and is driven by homozygous mutations in the zinc finger protein gene ZFP57. These patients exhibit methylation aberrations in several DMRs (e.g., PLAGL1 and GRB10), yet other TNDM patients who do not carry ZFP57 mutations experience hypomethylation in many additional DMRs, which is driven by an unknown mechanism. In mouse embryonic stem cells (ESCs), ZFP57 was shown to regulate multiple imprinted loci and recently, the zinc finger protein ZNF445 was also suggested to be involved in imprinting regulation. Nevertheless, while extensive loss-of-imprinting occurs in mice following knockout of both Zfp57 and Znf445, most imprinted DMRs are still preserved in human ESCs following knockdown of these genes. In addition, PGC7/STELLA and G9a/GLP were shown to be important for the preservation of several DMRs in mouse ESCs. Nevertheless, their effect on imprinting in humans has not been thoroughly examined. Collectively, accumulating evidence suggests that imprinting might be regulated by multiple factors in a species-specific manner.

Here, we use haploid and diploid parthenogenetic human ESCs (hpESCs) to perform a comprehensive analysis of imprinting regulation. We employed a CRISPR/Cas9 loss-of-function screening to identify factors involved in this process. hpESCs serve as a favorable tool to study imprinting in humans for several reasons: (1) hpESCs consist of only the maternal genome. Therefore, imprinted paternally expressed genes (PEGs) are silenced in these cells, markedly simplifying interrogation of imprinting. (2) hESCs are widely used to model human embryonic development because of their pluripotent nature. (3) Using haploid cells enhances the efficiency of loss-of-function screens. Exploiting these advantages, we investigated both known and unknown mechanisms that are required for maintaining parental imprinting.

**Results**

**Global DNA demethylation in hpESCs activates only a subset of imprinted genes.** DNA methylation is considered a central mechanism for the monoallelic silencing of imprinted genes. To explore this mode of imprinting regulation, we analyzed the consequences of global DNA demethylation achieved by both genetic and pharmacological manipulations. To this end, we applied CRISPR/Cas9 to genetically target the DNA methyltransferase 1 (DNMT1) enzyme and used the DNMT inhibitor 5-aza-2′-deoxycytidine (5-aza-dC) to chemically induce hypomethylation in hpESCs. While mouse ESCs were shown to retain proper self-renewal upon a triple knockout (KO) of Dnmt1/Dnmt3α/Dnmt3β, human ESCs are sensitive to the loss of DNMT1. Indeed, several days after the initiation of demethylation (by infection with lentivirus encoding sgRNA targeting Dnmt1, or following treatment with 5-aza-dC), there was an apparent cell death in the hpESC culture. However, we were able to optimize the experimental conditions to yield efficient infection and quick selection, which allowed the collection of viable cells for analysis of gene expression and CpG methylation profiles.

5-aza-dC treatment resulted in a significant reduction of global DNA methylation levels (Fig. 1a, \( P < 2.2 \times 10^{-16} \), one-tailed, paired t-test) and an upregulation of a subset of genes (Fig. 1b). DNMT1 KO also induced a notable overexpression of a group of genes, similar to those that were affected by 5-aza-dC (Fig. 1b and Supplementary Fig. 1a). Combining DNMT1 KO and 5-aza-dC-treated samples for a differential expression analysis of hypomethylated vs. control hpESCs revealed significantly upregulated and downregulated genes following demethylation, which were mostly associated with gene ontology (GO) terms related to methylation silencing and KRAS activation (Supplementary Fig. 1b, c).

While paternal DMRS are usually found in intergenic regions, maternal DMRS are commonly based at gene promoters and are thought to mainly regulate imprinting in cis. Following 5-aza-dC treatment, most maternal DMRS (which are hypermethylated in parthenogenetic cells) were extensively demethylated (Fig. 1c, d), similar to the genome-wide effect observed for hypermethylated promoters (Fig. 1e). Nevertheless, several DMRS (e.g., SNRPN DMRS 1 & 2, GRB10, PLAG1, GNAS-XL) retained relatively high DNA methylation levels (Fig. 1c), suggesting differential sensitivity to global demethylation among maternal DMRS. The evident loss of methylation in the remaining maternal DMRS (Fig. 1c), induced a subsequent expression in some PEGs, but most of them remained silenced (58%, 11/19 PEGs; Fig. 1f and Supplementary Fig. 1d, e). Some PEGs that were activated following demethylation (e.g., NDN, NNAT) were transcribed at similar levels as in bi-parental cells, whereas the expression of others (e.g., NAPIL5) were still much lower than in bi-parental cells (Supplementary Fig. 1f). Together, these results reveal distinct responses to DNMT1 inhibition among PEGs, both in the extent of demethylation, as well as in the transcriptional consequences of hypomethylation. This suggests that the mode by which DNA methylation regulates imprinting in humans is locus specific. Notably, the repressive state of many PEGs is retained, even when their associated DMRS were hypomethylated. This implies the existence of additional regulatory layers which might repress the expression of these PEGs on the maternal allele.

**Genome-wide CRISPR/Cas9 screen to identify factors that maintain imprinting of PEG10.** To search for unknown factors that maintain imprinting, we conducted a CRISPR/Cas9 genome-wide loss-of-function screen in haploid hpESCs (Fig. 2a). We focused on the imprinted gene PEG10, because it is a single-isosform imprinted gene which is silenced in hpESCs and is not significantly activated following demethylation (Fig. 1e, f). Moreover, PEG10 is important for normal placent
Fig. 1 Global DNA demethylation in parthenogenetic human ESCs. 

a Boxplots displaying the median with 25th–75th percentile and range of global DNA methylation levels (β values) across all analyzed CpGs in bi-parental hESCs (yellow), control parthenogenetic hESCs (red) and 5-azadC-treated parthenogenetic hESCs (green). **P < 2.2e-16 (one-tailed, paired t-test). Whiskers represent values within 1.5 X the interquartile range.

b 2D density plot showing the log2 TPM of all genes in DNMT1-KO (left y-axis) or 5-azadC-treated hESCs (right y-axis) vs. control hESCs (empty Cas9 vector or DMSO, x-axis). 

c Heatmap of mean methylation levels (β values) of maternal DMRs in bi-parental, control parthenogenetic and 5-azadC-treated hESCs. d 2D density plot showing mean methylation levels (β values) of regions associated with maternal DMRs in control (x axis) vs. 5-azadC-treated (y axis) hESCs. Only regions with mean methylation >0.5 in control hESCs are included. Dashed line represents identical methylation (slope = 1).

d Same analysis as in (c), performed for genome-wide regions associated with transcription start sites (TSS), 5′-UTR and 1st exon.

e Two-sided bar plot of PEGs that are expressed in bi-parental hESCs. Left bars (orange) illustrate the delta methylation levels (β values) of the associated maternal DMR, between 5-azadC-treated and control hESCs. Wide bars are used for DMRs that control several PEGs. Right bars (blue) represent fold change (FC) of the mean TPM between DNMT1 KO and 5-azadC-treated (referred to as demethylated) vs. control hESCs. Blue line illustrates unchanged expression (FC = 1).
development and is overexpressed in various cancers, promoting tumor proliferation. To isolate cells that activate PEG10, we performed intracellular immunofluorescent staining using anti-PEG10 antibody. FACS analysis of immunostained samples was able to distinguish between androgenetic cells (expressing PEG10) and parthenogenetic cells (in which PEG10 is silenced) (Fig. 2b). Next, we applied this immunostaining to the CRISPR/Cas9 parthenogenetic library that consist ~180,000 sgRNAs targeting 18,166 genes (~150 × 10^6 cells were stained in each replicate, >800 fold of the library size). Immunostained library cells were sorted using FACS to collect the small PEG10-positive (PEG10^+ 4 × 10^5 − 1 × 10^6 cells), as well as a comparable fraction of PEG10-negative cells. The distribution of sgRNAs within the sorted samples were assessed using high-throughput
sequencing and compared to the unsorted library to identify candidate genes whose perturbations resulted in PEG10 activation (Fig. 2a, c, d). We removed genes that were significantly enriched also in the PEG10-negative sorted samples, thereby eliminating multiple tumor suppressors (Supplementary Fig. 2a, b; see “Methods” for more details). Finally, we established a list of 115 significantly enriched genes that are potentially important for maintaining maternal imprinting at the PEG10 locus (referred to as “candidate genes”; Supplementary Data File 1). These candidate genes were enriched in GO terms associated with transcription, DNA binding, and repressive functions (Supplementary Fig. 2a) and a subset of them were functionally related to chromatin, appearing in the EpiFactors database35, identified as zinc finger proteins (ZNF, from www.genenames.org), transcription factors (TF)36, or other chromatin-related roles (Fig. 2e). Genes listed in the EpiFactors database were specifically enriched in the candidate list compared to their representation in the entire library (P = 0.027, Fisher’s exact test). The identification of many chromatin regulators within the candidate genes is consistent with the expected outcomes of the screen to discover factors that repress imprinted genes. Nonetheless, some enriched genes had other functions that are not directly related to chromatin regulation but could still affect imprinting indirectly via various pathways.

Inhibiting the MEK/ERK pathway in hESCs drives loss of imprinting. We first searched for upstream global pathways that might be associated with the candidate genes using the Transcription Factor Enrichment Analysis (TFEA) from the X2K Web tool37 which labels upstream regulatory networks from a supplied list of genes, followed by Kinase Enrichment Analysis (KEA)38. This analysis identified MAPK1 (ERK2) as the most significant regulator of a transcriptional network involving a set of candidate genes. Another member of the MEK/ERK pathway, MAPK3 (ERK1), was also found among the top 20 enriched kinases (Supplementary Table 1). Moreover, the functions of several candidate genes are directly linked with this pathway (e.g., CAMKV, MPAP3L, FAM83A). Inhibiting MEK/ERK is part of the 2i protocol for establishing naive mESCs39, which are characterized by genome-wide DNA hypomethylation40 and often exhibit imprinting aberrations41. Recent efforts established multiple protocols to generate naive human pluripotent cells, all of which involve inhibition of the MEK/ERK pathway42. These cellular methods were also found to exhibit loss-of-imprinting in various regions43. To test the involvement of FGF signaling and the MEK/ERK pathway in human imprinting regulation, we treated hESCs with a specific MEK/ERK inhibitor (PD0325901, 25 μM) for 5 days, in the absence of bFGF. This resulted in significant upregulation of PEG10, as suggested by our screen (Fig. 2f). Notably, several additional PEGs were also activated upon MEK/ERK inhibition, namely INPP5F, SGCE, and BLCAP (Fig. 2f).

ATF7IP regulates PEGs by facilitating the repressive histone modification H3K9me3. Next, we selected for validation five candidate genes that were shown to be involved in transcriptional repression. Using CRISPR/Cas9, we mutated each of these genes in hpESCs individually and showed that knocking out ATF7IP and ZMYM2 led to the activation of PEG10 (Fig. 2f). ATF7IP was shown to bind the histone methyltransferase SETDB1 and is required for catalyzing the heterochromatin mark H3K9me3 in various genomic regions43,44. It also interacts with the methyl CpG binding protein MBD1 to repress transcription45 and maintain X chromosome inactivation46. ZMYM2 associates with the LSD1/Co-REST/HDAC (LCH) repressive complex, which deacetylates histones and demethylates H3K4me346,47. Interestingly, a direct interaction between ATF7IP and ZMYM2 proteins in mESCs has been recently discovered48.

Among the top candidate genes, disruption of ATF7IP in hpESCs induced the highest levels of PEG10 expression (Fig. 3a). Hierarchical clustering analysis based on the expression of all genes revealed similar global expression patterns in ATF7IP-KO and control hpESCs (Supplementary Fig. 3a), whereas applying this clustering to PEGs only segregated ATF7IP-KO from control hpESCs (Supplementary Fig. 3b). This is different than the pattern observed for DNMT1-KO or 5-azadC-treated hpESCs, which clustered separately from control in both the genome-wide and PEG-only analyses (Supplementary Fig. 3a, b). This suggests that while DNA demethylation drives global changes in gene expression, the genome-wide effect of ATF7IP disruption is more restricted, with a major impact on expression of PEGs. Differential expression analysis between ATF7IP-KO and control hpESCs identified three significantly upregulated PEGs: PEG10, FAM50B, and SNRPN (Fig. 3b). Yet, while PEG10 and FAM50B are highly expressed in ATF7IP KO cells, in comparable levels to those observed in bi-parental cells, the activation of SNRPN reaches ~1% of bi-parental cells (Fig. 3c and Supplementary Fig. 3c). A previous study in HeLa cells demonstrated substantial loss of H3K9me3 in several genomic regions following KO of ATF7IP or SETDB144. We reanalyzed this ChIP-Seq data and focused on H3K9me3 peaks surrounding imprinted DMRs. In agreement with the expression patterns observed in hpESCs, H3K9me3 peaks at the PEG10 and FAM50B DMRs were lost following ATF7IP KO (Fig. 3d, e). However, H3K9me3 levels throughout the Prader-Willi syndrome (PWS) locus encompassing SNRPN, were similar between WT and ATF7IP-KO or SETDB1-KO (Supplementary Fig. 3d). Next, we examined
changes in H3K9me3 levels at other maternal DMRs, aiming to point to additional PEGs that might be regulated by ATF7IP. Specifically, the DMRs of MEST, KCNQ1OT1, PEG3, and GRB10, also exhibited loss of H3K9me3 peaks in ATF7IP KO hpESCs, possibly because unlike PEG10 and FAM50B, these genes also consist a non-imprinted isoform 32, or because they are not normally transcribed in hESCs. Other imprinted loci such as DIRAS3 did not exhibit a reduction in H3K9me3 levels following ATF7IP KO (Supplementary Fig. 3h). We also performed ChIP qPCR of H3K9me3 in hpESCs and confirm its reduction at both PEG10 and FAM50B DMRs (Supplementary Fig. 3i). Next, to probe the effect of ATF7IP on DNA methylation, we performed reduced representation bisulfite sequencing (RRBS) and analyzed methylation patterns in ATF7IP KO and control hpESCs. Genome-wide analysis revealed that loss of ATF7IP results in a small, but significant, reduction of global DNA methylation levels (Supplementary Fig. 4a, P < 2.2e-16, one-tailed t-test). Focusing on imprinted regions demonstrated considerable hypomethylation at the DMRs of PEG10 and FAM50B in ATF7IP KO hpESCs and bi-parental hESCs (Fig. 3g and Supplementary Fig. 4b–d). Moreover, significant hypomethylation was also observed at the DMR of PEG13 (Fig. 3h) which was not activated following loss of ATF7IP, possibly because it is mostly not expressed in hESCs.

While loss of ATF7IP induced expression of several imprinted regions, PEG10 was the only imprinted gene that was significantly upregulated in ZMYM2 KO hpESCs. In agreement with these
transcriptional changes, probing DNA methylation in ZMYM2 KO hpESCs using RRBS, confirmed a significant reduction at the PEG10 DMR, but not at the FAM50B DMR (Fig. 3g and Supplementary Fig. 4b, c). We showed previously that ZMYM2 KO in bi-parental hESCs induces global histone H3 hyperacetylation across multiple loci. Analysis of these ChIP-Seq data focusing on imprinted regions demonstrated hyperacetylation within the DMR of PEG10 in ZMYM2 KO bi-parental samples (Supplementary Fig. 4e).

Altogether, our analysis on imprinting regulation favors a model in which different factors affect distinct imprinted loci by maintaining various layers of repressive chromatin modifications (Fig. 3i). While the monoallelic silencing of some PEGs (e.g., NDN) depends on the presence of DNA methylation and DNMT1, suppression of other loci (e.g., PEG10, FAM50B) relies on histone modifications and their associated chromatin-modifying complexes, namely ZMYM2 and ATF7IP. Still, for some PEGs (e.g., SNRPN), the mode of regulation and the proteins required to preserve its imprinting are still obscure and call for further research (Fig. 3i).

ATF7IP represses genes associated with spermatogenesis. Although the broad function of ATF7IP has been studied in many contexts, its role in hESCs has not been characterized before. To outline the global regulatory activity of ATF7IP in hESCs, we performed a genome-wide differential expression analysis between ATF7IP-KO and control hpESCs. This revealed a bias for upregulated genes, consistent with ATF7IP acting mostly as a transcriptional repressor. Analyzing the tissue expression distribution of these upregulated genes using the Genotype-Tissue Expression (GTEx) database revealed a striking large cluster of genes which are specifically expressed in testis (Fig. 4a, b and Supplementary Data File 2). Among these upregulated genes are several epigenetic repressors that are mostly active in sperm (e.g., PRAME, CTCLN, DNMT3L, PIWILL1, MOVI10L1). We also found a significant enrichment of multiple members of the histone gene cluster HIST1 (P < 0.00001, Fisher’s exact test) (Fig. 4b). HIST1 genes were shown to be highly expressed in normal mitotic spermatogonia in mice and to be downregulated in mutant spermatogonia cells. Enriched GO terms associated with the upregulated protein-coding genes (discarding HIST1 genes), identified a marked enrichment for functions related to germ cells, including terms associated with reproduction, male gamete generation and meiosis (Fig. 4c). To ensure that these observations are not restricted to parthenogenic cells, we directed CRISPR/Cas9 to target ATF7IP also in bi-parental hESCs and showed that these cells exhibited similar upregulation of sperm-specific genes (Supplementary Fig. 4a). Additionally, diseases that are associated with ATF7IP mostly involve a testis-related phenotype, and include testicular germ cell tumors, cryptorchidism and male infertility (Supplementary Table 2). Furthermore, inspecting the tissue distribution of the mutations in ATF7IP that are reported in the catalogue of somatic mutations in cancer (COSMIC) database demonstrated a significant bias toward copy number variation (CNV) gains in testes. Herein, more than 46% of samples that are associated with testicular cancer exhibited elevated copy numbers of a region on chromosome 12 containing ATF7IP (Supplementary Fig. 4b). Overall, our data reveals that loss of ATF7IP in hESCs leads to activation of multiple genes that are specifically expressed in the testis and are associated with spermatogenesis.

Discussion

Imprinted genes stand out from the rest of the genome, as they are expressed from either the maternal or the paternal allele, depending on DNA methylation inherited from the parental gametes. Although many genes are differentially methylated between oocyte and sperm, most of these regions are reprogrammed during early development, with the exception of imprinted DMRs. Thus, maintenance of DNA methylation during pre-implantation is the decisive step which dictates imprinting. Yet, the mechanisms enabling this maintenance are mostly unknown, especially in humans. Here, we set out to explore these mechanisms, using hpESCs as a model. We describe how global DNA hypomethylation results in distinct consequences on the expression of imprinted genes, and identify regulators of imprinting using a genome-wide loss-of-function screen.

Previous studies have established the essentiality of DNMT1 and DNA methylation maintenance in human ESCs. Accordingly, we observed extensive cell death after targeting DNMT1 with CRISPR/Cas9 or inhibiting it with 5-azadec, but were able to calibrate a short-term experiment to collect sufficient number of cells for transcriptional and methylation analysis, which also validated the global hypomethylation signature of these cells. Since the initiation of cell death and overall toxicity can affect gene expression and cell state, we do not attempt to infer the global effect of DNMT1 loss on hESCs. Instead, our results are only focused on the direct influences of DNMT1 perturbation on the methylation and activation of PEGs (which are normally hypermethylated and silenced in hpESCs), as they are not expected to be affected by DNA damage or cytotoxicity. Our results showed that PEGs can have diverse responses to global hypomethylation, calling for further investigation on the potential causes and mechanisms for such differences.

The CRISPR/Cas9 screen aimed at identifying loss-of-imprinting of the imprinted gene PEG10, revealed ATF7IP and ZMYM2 as regulators of PEG10 and additional imprinted regions. ATF7IP was shown to bind to SETDB1 and to be essential for SETDB1-dependant H3K9me3 and gene silencing. Proteomic analyses have revealed that ATF7IP directly interacts with the repressive HUSH complex, and that it can bind different partners by distinct protein domains. Interestingly, a recent study has shown that ZMYM2 interacts with the C-terminal fibronectin type-III domain of ATF7IP and that it has a role in transgene silencing by ATF7IP. The fact that our screen also identified both proteins independently, can suggest that they might also cooperate in silencing the maternal allele of PEG10. Nevertheless, exactly how ATF7IP and/or SETDB1 are specifically targeted to their associated genomic loci (including specific imprinted PEGs), remains mostly unknown and is an important open question. In addition to identifying specific regulators, the results of the screen also suggested involvement of the Mek/Erk pathway in regulating imprinting. Inhibiting this pathway (using high concentrations of PD0325901) confirmed upregulation of several PEGs in hpESCs. This is in accordance with previous studies in mESCs, which linked 21 culture conditions and specifically Mek/Erk inhibition or knockout with erosion of DNA methylation, including at imprinted control regions.

Our results demonstrate a direct role for ATF7IP in preserving imprinting in humans. Studies in mice have previously associated histone methylation with imprinting regulation, also specifying the involvement of Setdb1 in this process. Knockdown of Setdb1 was shown to activate the PWS gene cluster on chromosome 15, however our results indicated that there was only a small activation of this locus following ATF7IP-KO in hpESCs. Other studies pointed at the H3K9-methyltransferase G9a to be regulating the PWS locus. Nevertheless, G9a was also reported to directly methylate lysine residues within ATF7IP, thereby affecting its repressive functions. These conflicting reports prompt further investigations to pinpoint the roles of SETDB1/ATF7IP and G9a/GLP complexes in regulating repression of PEGs within the PWS locus, and to examine possible
redundancy between them. ATF7IP and SETDB1 primarily act to repress endogenous retroviruses (ERVs)\(^61-63\), mainly during the global DNA demethylation post fertilization\(^62\). Insertion of ERVs was suggested to drive the evolution of parental imprinting\(^64\) and both PEG10 and FAM50B, which were highly activated in ATF7IP-KO hESCs, are derived from ancestral retrotranspositions\(^67-69\). Moreover, FAM50B is only imprinted in humans, whereas in mice it is biallelically expressed\(^70\). Consequently, our results propose that H3K9me3 set by ATF7IP and SETDB1, has evolved to regulate specific imprinted genes which are associated with retrotransposons, as part of its broader function to silence ERVs.

In *Drosophila*, SETDB1 and WDE (the homolog of ATF7IP) secure the identity of oocytes by repressing testis-specific genes\(^71\). In mice, SETDB1 is essential for meiosis in oocytes\(^72,73\) and for survival of spermatogonial stem cells\(^74\). Together with our findings that ATF7IP KO upregulates many sperm-specific genes in hESCs, SETDB1 and ATF7IP seem to have an evolutionary conserved role in regulating germ cell genes. Interestingly, aberrations in imprinted genes have been previously associated with sperm abnormalities\(^75\). Particularly, FAM50B, which we found to be regulated by ATF7IP, is highly expressed in normal spermatogenic cells\(^69\), whereas it features decreased methylation in asthenozoospermia\(^73\) and exhibit loss-of-imprinting in many

---

**Fig. 4** ATF7IP KO induces expression of genes involved in spermatogenesis. **a** Heatmap of expression levels (Z score TPM) across tissues (data from the GTEx study) of all the upregulated genes in ATF7IP KO cells (158 genes as identified in (b), log FC \(> 2\) and \(P \text{ value} < 0.05\)). **b** Volcano plot of differential expression (calculated by edgeR) showing the log2 fold change (FC) between ATF7IP KO and control (empty Cas9 vector) hESCs for all genes (x axis, \(n = 3\) replicates from each cell type), y axis represents −log10 \(P\) value. Marked in blue are genes with \(P \text{ value} < 0.05\). Upregulated genes from the HIST1 gene cluster are highlighted in purple and representative epigenetic repressors that are mostly active in sperm are shown in red. **c** Top 20 enriched gene ontology (GO) terms (analyzed by GSEA, FDR \(q\) values < 0.05) among the upregulated protein-coding genes following ATF7IP KO (158 genes that were identified in (b), log FC \(> 2\) and \(P \text{ value} < 0.05\), excluding HIST1 genes). x axis represents −log10 \(P\) value (hypergeometric distribution, as calculated by GSEA). Hierarchical clustering for the GO terms was applied based on a matrix of genes associated with each term (“Manhattan” distance measure) and colored accordingly. **d** An illustration summarizing our model for ATF7IP activity in hESCs. Sperm-specific and imprinted paternally expressed genes (PEGs) are differentially expressed between oocyte and sperm and are silenced by ATF7IP in hESCs. PEGs are only repressed from the maternal allele, whereas sperm-specific genes are completely silenced. Following ATF7IP KO, these genes are activated, resulting in upregulation of a group of sperm-specific and paternally expressed genes.
semionatous testicular germ cell tumors. These indications prompt for further investigation regarding the involvement of ATF7IP in male infertility and testicular cancer via regulation of imprinting and non-imprinted genes that are important for spermatogenesis.

Collectively, our study uncovers a coupled role for ATF7IP in silencing sperm-specific genes, as well as a subset of maternally imprinted genes. Interestingly, both groups of genes are differentially regulated between oocyte and sperm, resulting in sex-specific germ cell development and formation of parental imprinting (Fig. 4d). Oocyte and sperm are known to possess eminent epigenetic asymmetries, which later persist in the maternal and paternal pronuclei following fertilization. Notably, spermatozoa chromatin is largely devoid of histones, thereby H3K9 methylation is absent throughout most of the paternal pronucleus, allowing its rapid demethylation following fertilization.

In hESCs, ATF7IP maintains silencing of PEGs on the maternal allele, yet sperm-specific genes are repressed on both alleles (Fig. 4d). These differences call for further research to understand the potential role of the paternal allele for 24 h after ATF7IP-mediated repression, specifically at maternally imprinted genes. Finally, knocking out ATF7IP in hESCs results in the epigenetic erasure of several maternally imprinted genes along with upregulation of sperm-specific genes (Fig. 4d).

**Methods**

**Cell culture.** Throughout the study we used the following cell lines: haploid hPESCs - hPES10; diploid hPESCs - SwapS47; for RNA-Seq and RRBS, pES6, pES2, pES7; for DNA methylation array; bi-parental hESCs - WA09 (H9), CSE6 (for RNA-Seq), NYSFC1, NYSFC2, HuES53, HuES64 (for DNA methylation array). hESCs were cultured on mouse embryonic fibroblast (MEF) treated with mitomycin-C. hESC growth medium was changed every 1–2 days, containing KnockOut Dulbecco’s modified Eagle’s medium (Gibco-Invitrogen, CA) supplemented with 15% KnockOut Serum Replacement (Gibco-Invitrogen, CA), 1 mM glutamine, 0.1 mM β-mercaptoethanol (Sigma-Aldrich, MO), 1% nonessential amino acids stock (Gibco-Invitrogen, CA), penicillin (50 U/mL), streptomycin (50 μg/mL), and 8 ng/mL FGF2 (Gibco-Invitrogen, CA). Then, 10 μM ROCK inhibitor (Y27632, Stemgent) was supplemented to the medium in the first 24 h after thawing and passing cells. Cells were passaged using Trypsin-EDTA (Biological Industries, Beit Haemek, Israel).

**CRISPR/Cas9 library culture.** Our recently established mutant library on haploid hPESCs (based on hPES10 cell line)28 was maintained for 4 weeks and frozen. Briefly, two-hundred clones were obtained using the CRISPR/Cas9 library containing 181,131 sgRNAs at a multiplicity of infection of 0.3. Infected cells were thawed 2 days, containing KnockOut Dulbecco’s modified Eagle’s medium (Gibco-Invitrogen, CA) supplemented with 15% KnockOut Serum Replacement (Gibco-Invitrogen, CA), 1 mM glutamine, 0.1 mM β-mercaptoethanol (Sigma-Aldrich, MO), 1% nonessential amino acids stock (Gibco-Invitrogen, CA), penicillin (50 U/mL), streptomycin (50 μg/mL), and 8 ng/mL FGF2 (Gibco-Invitrogen, CA). Then, 10 μM ROCK inhibitor (Y27632, Stemgent) was supplemented to the medium in the first 24 h after thawing and passing cells. Cells were passaged using Trypsin-EDTA (Biological Industries, Beit Haemek, Israel).

**Immunostaining and FACs sorting.** To distinguish PEG10-expressing cells (PEG10+), we performed intracellular antibody staining. Cells were washed with phosphate buffered saline (PBS) and discarded Trypsin-EDTA. Following centrifugation, cells were resuspended in PBS and fixed by slow dropwise addition of MeOH until reaching 90% MeOH solution. The tubes were incubated for 30 min on ice and then cells were washed twice with PBS supplemented with 0.5% bovine serum albumin (BSA) and stained with anti-PEG10 antibody (1:500, Abcam ab215035) in 100% heat-inactivated fetal bovine serum (FBS) overnight at 4 °C. Cells were washed twice with PBS/BSA and stained with goat anti-Rabbit Alexa594-conjugated secondary antibody (1:800, Abcam ab150580) in 100% PBS for one hour on ice and washed twice again with PBS/BSA. Finally, the cells were filtered through a 70 μm cell strainer (Corning) and sorted using FACs. Two-hundred clones were twice transfected with Trypsin-EDTA, centrifuged, and resuspended in PBS/BSA. Then, 3–4 million cells were harvested without sorting to serve as unsorted control.

**DNA sequencing and sgRNA enrichment analysis.** Genomic DNA was extracted using gSYNC DNA extraction kit (Geneaid). A region containing the sgRNA integration site was amplified with Taq polymerase and was compatible for Nextera DNA library preparations (Illumina), as detailed in Supplementary Table 3 and as previously described86. After purification of the 180-base-pair (bp) product, a second PCR reaction was performed using Nextera adapter primers to generate a Nextera DNA library according to the manufacturer’s instructions (Illumina). DNA libraries containing sgRNAs were constructed from two replicates, and experiments were sequenced using NextSeq 500 (Illumina). For the unsorted controls, between 20 and 40 million cells were analyzed, corresponding to a coverage of ~100–200 fold of the library size. For the PEG10-positive and PEG10-negative conditions, between 400,000 and 1,000,000 were sorted and analyzed. The numbers of reads obtained from sequencing were between 10–20 million for the unsorted controls, 3–7 million for the PEG10-positive and 2–10 million for the PEG10-negative cells. More than 150,000 sgRNAs were represented in the reads for unsorted controls, while the reads for the PEG10-positive and PEG10-negative conditions had a representation of 60,000–120,000 sgRNAs. edger87 was used to calculate the log fold change (Log, FC) of sgRNA counts between unsorted control (n = 3 replicates) and PEG10-positive (n = 4 replicates) or PEG10-negative cells (n = 5 replicates). The differential representation analysis by edger was performed separately for every immunostaining experiment (each containing two replicates). Therefore, every gene incorporated 20 sgRNAs (~10 sgRNAs from the library X 2 experiments). The final Log, FC for a gene was calculated as the median of these sgRNAs after removing outliers (using the R command boxplot.stats(s)x). As expected from a positive selection, most genes had a negative Log, FC value. To normalize this bias, we centered the values around 0 by subtracting the median Log, FC value of all the genes from the Log, FC of each gene. The differential Fold change was calculated by dividing the median Log, FC values of the PEG10-negative sort over that of the PEG10-positive sort. Genes that passed the following filters were considered enriched and included in the “candidate gene” list (Supplementary Data File 1): (1) Log, FC > 1.4 (equivalent to >0.5 before normalization); (2) P value < 0.05; (3) TPM > 1 (calculated from RNA-Seq performed previously on the CRISPR/Cas9 haploid hPESC library88); and (4) Genes having Log, FC > 0.5 and P value < 0.05 in the PEG10-negative sort were discarded.

**RNA extraction and sequencing.** Total RNA was isolated using RNasy Mini Kit (QiAGEN) and mRNA was enriched by pull down of poly(A)-RNA. RNA sequencing libraries were generated using KAPA RNA Prep Kit according to the manufacturer’s protocol and sequenced using Illumina NextSeq 500 with 75 bp single-end reads. For analyses of gene expression, reads were mapped to the GRCh38 reference genome using STAR82. Normalization of the read counts, differential expression (DE) and statistical analyses were performed using edgeR83. The result of the RNA-Seq DE analysis can be found in Supplementary Data File 3.

**MEK/ERK inhibition.** To block FGF2 signaling, SwapS4 hPESCs were cultured in standard hESC growth medium until reaching ~40% confluence. Then bFGF (FGF2) was removed from the medium with addition of 25 μM PD0325901 (Biomeda #911091) or 0.1% DMSO as control for 5 days. Medium was changed daily.

**Generation of individual gene knockouts.** CRISPR/Cas9 with a specific sgRNA was used to target DNM1T and five candidate genes. Sequences of these sgRNAs (Supplementary Table 2) were transfected into SwapS4 cells (a lentiviral vector (a gift from Feng Zhang, Addgene cat. no. 52961). For control, we used lentiCRISPR v2 vector without any sgRNA. For viral production, 293T cells cultured in 10 cm culture plates with around 70–80% confluence were transfected with sgRNA-containing lentiCRISPR v2 (5.7 μg), pCMV-VSV-G (2.8 μg) and pMPA-X2 (4 μg) plasmids, in the presence of polyethylene glycol “Max” (Poly-science). Transfection medium was exchanged with hESC medium after 24 h, and lentiviral particle-containing culture supernatant was harvested 60–65 h after transfection. Culture supernatant was spun down at 3000 c.p.m. for 10 min at 4 °C and then filtered through 0.45 μm cellulose acetate filters (Millipore). hPES10 or SwapS4 were transduced with Trypsin-EDTA, centrifuged, and resuspended in hESC growth medium supplemented with 10 μM ROCK inhibitor (Y27632) and 8 μg/ml polybrene (Sigma) and plated on feeder layer MEFs. Then, ~2 ml from the lentivirus-containing medium were then added to the six-well plates. At 24 h after transduction, virus-containing medium was replaced with standard hESC growth medium. At 48 h after transduction, the medium was exchanged to a new medium that contains puromycin (0.3 μg/mL, Sigma). Cells were kept under antibiotic selection for 7–14 days, followed by extraction of RNA or DNA.
ChiP-Seq analysis. For the analysis of H3K9me3, we reanalyzed previously published ChiP-Seq data performed in WT, SET1B KO, and ATFFFFP KO HeLa cells. Fastq files were downloaded to Galaxy48 via ftp link provided by European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/view/PRJNA342602) and aligned to hg38 using Bowtie2. BAM files were then converted to bigwig by bamCoverage (version 3.3.2) with bin size of 30 bp and visualized in the Integrated Genomics Viewer (IGV).

For the analysis of H3Ac, we used our recently published ChiP-Seq data performed in control and ZMYM2−/− hESCs49.

ChiP-qPCR. ChiP assay was performed as described previously44,45 with slight modifications: In short, cells were crosslinked with 1% formaldehyde at RT for 10 min, quenched with 0.125 M glycine and scraped from plate. Samples were lysed, homogenized, and sonicated for 20 cycles (cycle = 30 s on and 30 s off). For immunoprecipitation, the samples were incubated with an antibody against H3K9me3 (Sigma-Aldrich, Cat#: 07-442) overnight and mixed with protein A beads the next day. Following washes, samples were eluted with TE and reverse-crosslinked at 65 °C. DNA was recovered with DNA cleanup kit (Qiagen) and 1 μl was used for qPCR with primers listed in Supplementary Table 3. For each gene (PEG10 and FAM50B), FCs of H3K9me3 levels in the DMV vs. gene body regions were calculated based on the qPCR results and normalized to control cells.

DNA methylation analysis by methylation array. DNA methylation analysis of bi-parental and parthenogenetic hESCs treated with either DMSO (control) or 5-azaC for 5 days, was performed using Infinium 450K Methylation beadChips (Illumina) following the Infinium HD methylations protocol. Data was processed and normalized by using subset-quantile within array normalization (SWAN) and adjusted for batch effects using the R package ChAMP (version 1.4.0), as previously described86. Imprinted DMRs which were analyzed in this study and their association with CpG probes are listed in Supplementary Data File 4.

DNA methylation analysis by reduced representation bisulfite sequencing. Genomic DNA was extracted from control, ATFFFFP or ZMYM2 KO hESCs using gSYNC DNA extraction kit (Geneaid). From each sample, 1 μg of genomic DNA was sent to CD Genomics (Shirley, NY) for library preparation and sequencing. The DNA was digested with MspI (NEB), followed by ends preparation, adaptor ligation using Premium RRBS kit (Duagenode). Size selection was performed using AMPure XP beads (Beckman Coulter, Inc.) to obtain DNA fractions of MspI-digested products enriching for the most CpG-rich regions in the range of 150–350 bp. Subsequently, bisulfite treatment was conducted using the ZYM0 EZ DNA Methylation-Gold Kit. The converted DNAs are then amplified by 12 cycles of PCR, using 25 μl KAPA HiFi HotStart Ucaril ReadyMix (2X) and 8 bp index primers with a final concentration of 1 μM each and clean-up using AMPure XP beads. The constructed RRBS libraries were quantified using a Qubit fluorometer with Quant-iT dsDNA HS Assay Kit (Invitrogen), and sequenced on Illumina Hiseq DNA Methylation-Gold Kit. The converted DNAs are then amplified by 12 cycles of PCR, using 25 μl KAPA HiFi HotStart Ucaril ReadyMix (2X) and 8 bp index primers with a final concentration of 1 μM each and clean-up using AMPure XP beads. The constructed RRBS libraries were quantified using a Qubit fluorometer with Quant-iT dsDNA HS Assay Kit (Invitrogen), and sequenced on Illumina Hiseq.

Analysis of tissue expression for genes upregulated in ATFFFFP KO. The Expression Atlas website (https://www.ebi.ac.uk/gxa/home) was used to interrogate protein-coding genes that were significantly upregulated in ATFFFFP-KO hESCs (having log FC >2 and P-value <0.05), by downloading their TPM values across tissues taken from The Genotype-Tissue Expression (GTEs) database (https://gtexportal.org/home/).

Reporting further. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The CRISPR/Cas9 library sequencing, RNA-Seq and RRBS data generated in this study have been deposited in the ArrayExpress database under accession codes: E-MTAB-11012, E-MTAB-11014, and E-MTAB-11015. Previously published H3K9me3 ChiP-seq data that was analyzed in this study is available under the following accession code: E-MTAB-8170.

Received: 21 September 2020; Accepted: 28 October 2021; Published online: 18 November 2021

References

1. Plasschaert, R. N. & Bartolomei, M. S. Genomic imprinting in development, growth, behavior and stem cells. Development 141, 1805–1813 (2014).
2. Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. Nat. Rev. Genet. 12, 565–575 (2011).
3. Tucci, V. et al. Genomic imprinting and physiological processes in mammals. Cell 176, 952–965 (2019).
4. Butler, M. G. Genomic imprinting disorders in humans: a mini-review. J. Assist. Reprod. Genet. 26, 477–486 (2009).
5. Plass, C. & Sokowy, P. D. DNA methylation, imprinting and cancer. Eur. J. Hum. Genet. 10, 6–16 (2002).
6. Hanna, C. W. & Kelsey, G. The specification of imprints in mammals. Heredity (Edinb.). 113, 176–183 (2014).
7. Monk, D. Germ-line-derived DNA methylation and early embryo epigenetic reprogramming: the selected survival of imprints. Int. J. Biochem. Cell Biol. 67, 128–138 (2015).
8. Sanchez-Delgado, M. et al. Causes and consequences of multi-locus imprinting disturbances in humans. Trends Genet. 32, 444–455 (2016).
9. Aziz, S. et al. Multilocus methylation analysis in a large cohort of 11p15-related foetal growth disorders (Russell Silver and Beckwith Wiedemann syndromes) reveals simultaneous loss of methylation at paternal and maternal imprinted loci. Hum. Mol. Genet. 18, 4742–4753 (2009).
10. Bleik, J. et al. Hypomethylation at multiple maternally methylated imprinting regions including PLAG1 and GNAS loci in Beckwith–Wiedemann syndrome. Eur. J. Hum. Genet. 17, 611–619 (2009).
11. Valente, F. M. et al. Transcription alterations of KCNQ1 associated with imprinted methylation defects in the Beckwith–Wiedemann locus. Genet. Med. 21, 1808–1820 (2019).
12. Fontana, L. et al. Characterization of multi-locus imprinting disturbances and underlying genetic defects in patients with chromosome 11p15.5 related imprinting disorders. Epigenetics 13, 897–909 (2018).
13. Calliave, A. et al. A founder disorder of altered DNA methylation. J. Med. Genet. 51, 407–412 (2014).
14. Mackay, D. J. G. et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat. Genet. 40, 949–951 (2008).
15. Court, F. et al. Genome-wide allelic methylation analysis reveals disease-specific susceptibility to multiple methylation defects in imprinting syndromes. Hum. Mutat. 34, 595–602 (2013).
16. Docherty, L. E. et al. Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes. J. Med. Genet. 51, 229–238 (2014).
17. Shi, H. et al. ZFP57 regulation of transposable elements and gene expression related and beyond imprinted domains. Epigenetics Chromatin 12, 49 (2019).
18. Riso, V. et al. ZFP57 maintains the parent-of-origin-specific expression of the imprinted genes and differentially affects non-imprinted targets in mouse embryonic stem cells. Nucleic Acids Res. 44, 8165–8178 (2016).
19. Anvar, Z. et al. ZFP57 recognizes multiple and closely spaced sequence motif syndromes. Hum. Mutat. 34, 1118–1132 (2016).
20. Quenneville, S. et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol. Cell 44, 361–372 (2011).
21. Messerschmidt, D. Should I stay or should I go: protection and maintenance of DNA methylation at imprinted genes. Epigenetics 7, 969–975 (2012).
22. Takahashi, N. et al. ZNF445 is a primary regulator of genomic imprinting. Genes Dev. 33, 49–54 (2019).
23. Nakamura, T. et al. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat. Cell Biol. 9, 64–71 (2007).
24. Nakamura, T. et al. PGC7 binds histone H3K9me2 to protect against conversion of 5mc to 5hmC in early embryos. Nature 486, 415–419 (2012).
25. Zhang, T. et al. G9a/GLP Complex Maintains Imprinted DNA Methylation in Embryonic Stem Cells. Cell Rep. 15, 77–85 (2016).
26. Sagi, I. et al. Derivation and differentiation of haploid human embryonic stem cells. Nature 532, 107–111 (2016).
27. Dvash, T., Ben-Voyet, D. & Eiges, R. Human embryonic stem cells as a powerful tool for studying human embryogenesis. Pediatr. Res. 60, 111–117 (2006).
28. Yilmaz, A., Perez, M., Aharony, A., Sagi, I. & Benvenisty, N. Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. Nat. Cell Biol. 20, 610–619 (2018).
29. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
30. Tsunmura, A. et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
31. Liao, J. et al. Targeted disruption of Dnmt1, Dnmt3a and Dnmt3b in human embryonic stem cells. *Nat. Genet.* **47**, 469–478 (2015).
32. Smithies, O. et al. Dnmt3a and Dnmt3b are essential for multiple-tissue- and isospecific-imprinted transcribed imprinting. *Cell Rep.* **11**, 308–320 (2015).
33. Ono, R. et al. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **38**, 101–106 (2006).
34. Xie, T. et al. PEG10 as an oncogene: expression regulatory mechanisms and contribution to tumor progression. *Cancer Cell Int.* **18**, 112 (2018).
35. Medvedeva, Y. A. et al. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database* **2015**, bav067 (2015).
36. Lambert, S. A. et al. The human transcription factors. *Nucleic Acids Res.* **46**, W171–W179 (2018).
37. Clarke, D. J. B. et al. Expression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Res.* **44**, 359–41 (2016).
38. Timms, R. T., Tchasovnikarova, I. A., Antrobus, R., Dougan, G. & Lehner, P. J. Essential for heterochromatin formation by the HUSH complex. *ATF7IP-mediated stabilization of the histone methyltransferase SETDB1 is essential for meiosis and mitosis in mouse oocytes and early embryos. Development** **143**, 2767–2779 (2016).
39. Ware, C. B. Concise review: lessons from naïve human pluripotent cells. *Stem Cell Rep.* **5**, 35–41 (2015).
40. Ficz, G. et al. FGF signaling inhibition in ESCs drives rapid genome-wide silencing mediated by the SETDB1 complex. *PLoS Genet.* **10**, e1004562 (2014).
41. Kim, Y. et al. Targeting the histone methyltransferase G9a activates imprinted genes and improves survival of a mouse model of Prader-Willi syndrome. *Nat. Genet.* **42**, 23–23 (2010).
42. Leung, D. et al. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc. Natl Acad. Sci. USA* **111**, 6690–6695 (2014).
43. Carvajal, E. et al. Reactivation of maternal SNORD116 cluster via SETDB1 knockdown in Prader-Willi syndrome iPSCs. *Hum. Mol. Genet.* **23**, 4674–4685 (2014).
44. Basak, C. et al. The human histone methyltransferase ESET is required for the survival of spermatogonial stem/progenitor cells in mice. *Cell Death Dis.* **5**, e1196–e1196 (2014).
45. Alajem, A. et al. Differential association of chromatin proteins identifies novel targets of apparently idiosyncratic male infertility. *Int. J. Mol. Sci.* **21**, 1728 (2020).
46. Okada, Y. & Yamaguchi, K. Epigenetic modifications and reprogramming in paternal pronucleus: sperm, preimplantation embryo, and beyond. *Cell. Mol. Life Sci.* **74**, 1957–1967 (2017).
47. Afgan, E. et al. Galaxy: a one-stop shop for functional genomics. *Nucleic Acids Res.* **47**, W559–W570 (2019).
48. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
49. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
50. Hezroni, H. et al. H3K9 histone acetylation predicts pluripotency and reprogramming capacity of ES cells. *Nucleus* **2**, 300–301 (2011).
51. Weissbein, U., Plotnik, O., Vershkov, D. & Benvenisty, N. Culture-induced recurrent epigenetic aberrations in human pluripotent stem cells. *PLoS Genet.* **13**, e1006977 (2017).
52. Court, F. et al. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germ-line-methylation-independent mechanism of establishment. *Genome Res.* **24**, 554–569 (2014).
53. Smolko, A. et al. Human and mouse XAP-5 and XAP-5-like (X5L) genes: identification of an ancient functional retrotransposon differentially expressed in testis. *Genomics* **61**, 125–132 (1999).
54. Zhang, X. et al. Novel retrotransposed imprinted loci identified at human 6p25. *Nucleic Acids Res.* **39**, 5388–5400 (2011).
55. Smithies, O. et al. The human transcription factors. *Nucleic Acids Res.* **46**, W171–W179 (2018).
56. Dong, H. et al. The histone methyltransferase SETDB1 maintains female identity in Drosophila germ cells. *Nat. Commun.* **9**, 4155 (2018).
57. Kim, J. et al. Maternal Setdb1 Is required for meiotic progression and preimplantation development in mouse. *PLoS Genet.* **12**, e1005970 (2016).
58. Okada, Y. & Yamaguchi, K. Epigenetic modifications and reprogramming in paternal pronucleus: sperm, preimplantation embryo, and beyond. *Cell. Mol. Life Sci.* **74**, 1957–1967 (2017).
59. Alajem, A. et al. Differential association of chromatin proteins identifies novel targets of apparently idiosyncratic male infertility. *Int. J. Mol. Sci.* **21**, 1728 (2020).
60. Okada, Y. & Yamaguchi, K. Epigenetic modifications and reprogramming in paternal pronucleus: sperm, preimplantation embryo, and beyond. *Cell. Mol. Life Sci.* **74**, 1957–1967 (2017).
61. Afgan, E. et al. The Galaxy platform for accessible, reproducible and friendly scientific computing. *NATURE COMMUNICATIONS* | https://doi.org/10.1038/s41467-021-26949-7 | www.nature.com/naturecommunications
88. Neph, S. et al. BEDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920 (2012).

Acknowledgements
We thank all members of The Azrieli Center for Stem Cells and Genetic Research for critical reading of the manuscript. This work was partially supported by the Israel Science Foundation (494/17), the Israel Precision Medicine Program, the Rosetrees Trust, and Azrieli Foundation. S.B. is a Clore fellow. N.B. is the Herbert Cohn Chair in Cancer Research.

Author contributions
S.B. and N.B. conceived the study, designed the experiments, interpreted the data, and wrote the paper with input from all authors. S.B. performed the experiments with a significant contribution from O.Y., N.M., D.V., G.K., A.Y., T. E-G., M.N-R., and E.M.; S.B. analyzed the data with the assistance of G.C., D.V., and E.L.; N.B. supervised the study.

Competing interests
N.B. is CSO of NewStem Ltd. The other authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-26949-7.

Correspondence and requests for materials should be addressed to Nissim Benvenisty.

Peer review information Nature Communications thanks Andrea Riccio and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

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

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

© The Author(s) 2021