

Genome-wide screening for genes involved in the epigenetic basis of fragile X syndrome

Dan Vershkov,¹ Atilgan Yilmaz,¹ Ofra Yanuka,¹ Anders Lade Nielsen,² and Nissim Benvenisty^{1,*}

¹The Azrieli Center for Stem Cells and Genetic Research, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem 91904, Israel

²Department of Biomedicine, Aarhus University, Aarhus, Denmark

*Correspondence: nissimb@mail.huji.ac.il

<https://doi.org/10.1016/j.stemcr.2022.03.011>

SUMMARY

Fragile X syndrome (FXS), the most prevalent heritable form of intellectual disability, is caused by the transcriptional silencing of the *FMR1* gene. The epigenetic factors responsible for *FMR1* inactivation are largely unknown. Here, we initially demonstrated the feasibility of *FMR1* reactivation by targeting a single epigenetic factor, *DNMT1*. Next, we established a model system for *FMR1* silencing using a construct containing the FXS-related mutation upstream to a reporter gene. This construct was methylated *in vitro* and introduced into a genome-wide loss-of-function (LOF) library established in haploid human pluripotent stem cells (PSCs), allowing the identification of genes whose functional loss reversed the methylation-induced silencing of the *FMR1* reporter. Selected candidate genes were further analyzed in haploid and FXS-patient-derived PSCs, highlighting the epigenetic and metabolic pathways involved in *FMR1* regulation. Our work sheds light on the mechanisms responsible for CGG-expansion-mediated *FMR1* inactivation and offers novel targets for therapeutic *FMR1* reactivation.

INTRODUCTION

Since the identification of the CGG repeat expansion in the fragile X mental retardation 1 (*FMR1*) gene as the causative mutation in fragile X syndrome (FXS), many efforts have been invested in deciphering the epigenetic processes that disrupt *FMR1* expression in patients' cells. Although several studies have characterized the heterochromatic configuration of full mutation alleles (>200 CGG repeats) (Epsztejn-Litman and Eiges, 2019; Kumari and Usdin, 2010), the factors involved in causing and maintaining *FMR1* heterochromatinization remain elusive. It is assumed that the recruitment of repressive DNA-binding factors by the expanded CGG repeats mediates the DNA hypermethylation and inactivation of the *FMR1* locus, similar to other disease-associated repeat expansions that are characterized by the acquisition of abnormal DNA hypermethylation (Colak et al., 2014; Yanovsky-Dagan et al., 2015).

As the FXS-causing mutation is located in the non-coding region of *FMR1*, understanding and targeting the mechanisms responsible for *FMR1* inactivation might have a therapeutic value. The rare existence of individuals with apparent normal intelligence who carry an unmethylated CGG expansion indicates that the expression of full-mutation alleles produces functional protein and can prevent the neurocognitive manifestations of FXS (Smeets et al., 1995).

Human pluripotent stem cell (hPSC)-based models of FXS allow us to study the mechanisms responsible for *FMR1* silencing and to explore novel treatments capable of reactivating *FMR1* expression (Vershkov and Benvenisty, 2017; Zhou et al., 2016). Compound screening using FXS-patient-derived induced pluripotent stem cells (iPSCs) that harbor a

completely silenced *FMR1* locus highlighted the importance of DNA methylation in the maintenance of *FMR1* silencing and identified several candidate compounds that are able to target *FMR1* heterochromatinization (Vershkov et al., 2019).

In this study, we aimed to perform a comprehensive analysis of the genes and pathways regulating *FMR1* silencing and to identify novel targets for *FMR1* reactivation. Since our compound-screening study identified DNA methylation as a central mechanism for repressing *FMR1* expression, we first analyzed the consequences of the targeted perturbation of DNA methyltransferase 1 (*DNMT1*) in FXS-iPSCs and further characterized its role in the maintenance of *FMR1* inactivation. Next, we aimed to utilize the recent advances in CRISPR-Cas9-based genetic-screening technology to search for novel regulators involved in maintaining *FMR1* inactivation. While establishing complete loss-of-function (LOF) phenotypes is somewhat challenging in diploid cells that require the establishment of homozygous mutations, the use of haploid cells, which harbor a single set of chromosomes, increases the chances of conducting a comprehensive functional genetic screen (Yilmaz et al., 2016). Therefore, we generated a tractable model system to study methylation-induced *FMR1* silencing in haploid human embryonic stem cells (ESCs) and used it to screen for genes involved in *FMR1* inactivation.

RESULTS

CRISPR-Cas9-based disruption of *DNMT1* in FXS-iPSCs

In order to test the possibility of *FMR1* reactivation using gene targeting of a single epigenetic factor and to further

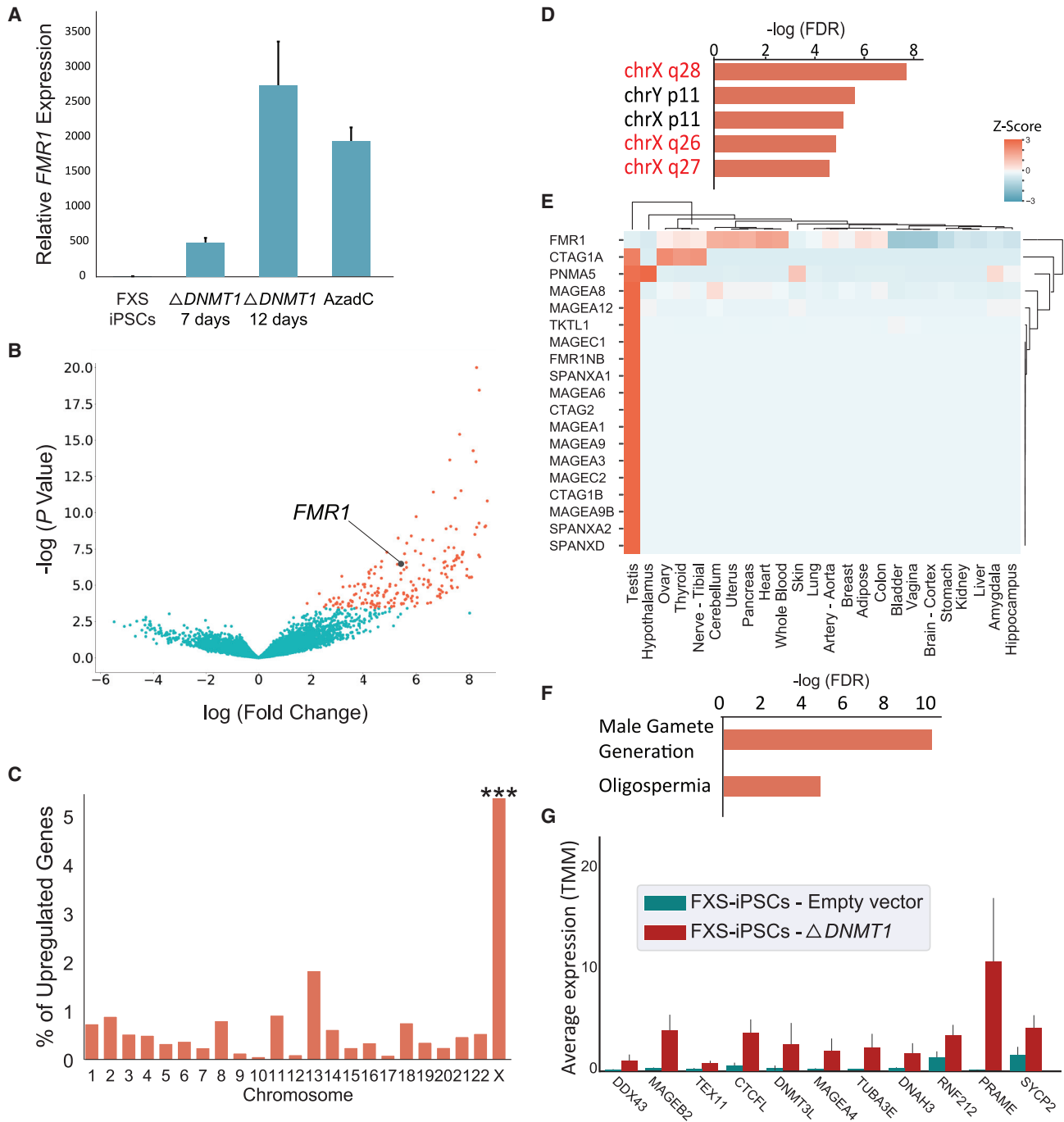


Figure 1. Analysis of *DNMT1* disruption in FXS-iPSCs

(A) RT-PCR analysis of *FMR1* expression in FXS-iPSCs 7 and 12 days after the delivery of sgRNA targeting *DNMT1* and Cas9, compared with treatment with 5-azadC (5 μ M, 4 days). Values represent the average \pm SEM of 2 independent experiments with 3 technical replicates. (B) Volcano plot showing the median log-fold expression change (x axis) and $-\log(\text{FDR})$ (y axis) for each gene in FXS-iPSCs following *DNMT1* perturbation (3 independent experiments), compared with empty vector controls (4 independent experiments). (C) Percentage of upregulated genes (FDR < 0.05, $\log(\text{FC}) > 1$) per chromosome following *DNMT1* disruption in FXS-iPSCs. ***hypergeometric $p < 0.001$. (D) Positional enrichment analysis in gene set enrichment analysis (GSEA) for upregulated genes upon *DNMT1* perturbation.

(legend continued on next page)



explore the maintenance of DNA methylation in the *FMR1* locus, we sought to analyze the consequences of targeted *DNMT1* perturbation in FXS-iPSCs. To overcome the sensitivity of hPSCs to the loss of DNMT1 (Liao et al., 2015), we transduced FXS-iPSCs with a lentiviral vector containing Cas9 and single guide RNA (sgRNA) targeting *DNMT1* and collected the culture following a short selection period. This way, despite the apparent cell death following lentiviral transduction, we were able to collect viable cultures for gene expression analysis. RT-PCR analysis of the mutated samples demonstrated the reactivation of *FMR1* expression following *DNMT1* disruption to levels comparable with 5'-aza-2'-deoxycytidine (5-azadC) treatment (Figure 1A). Analysis of an ESC line with a normal range of CGG repeats did not show any effect of *DNMT1* disruption on *FMR1* expression, validating the specific effect of *DNMT1* targeting on full-mutation alleles (Figure S1A). The transcriptional activity of *FMR1* in FXS-iPSCs upon *DNMT1* perturbation was accompanied by a significant decrease in DNA-methylation levels in the *FMR1* promoter (Figure S1B).

Next, we performed a global transcriptional analysis of the *DNMT1* mutants and the control FXS-iPSC samples, which identified *FMR1* as one of the top significantly upregulated genes following *DNMT1* mutagenesis (Figure 1B, false discovery rate [FDR] <0.001; Table S1). Analysis of the global transcriptional response following *DNMT1* perturbation revealed a striking enrichment of genes located in the regions adjacent to the *FMR1* locus and around the distal end of the long (q) arm of the X chromosome (Xq27-28), as well as several other loci on the X chromosome and a single region on the Y chromosome (Figures 1C and 1D). The enrichment of X-chromosome genes in the transcriptional response to *DNMT1* loss was not explained by the relative abundance of CpG islands in this chromosome, as analyzed by the CpG-island annotation from the UCSC genome browser (*CpgIslandExt*) (Figure S1C).

Tissue expression analysis of the top upregulated Xq27-28 genes (Figures 1E and 1F, FDR <0.05), as well as of all genes in the fragile-X-adjacent region (Figure S1D, X chromosome 140–148 Mb), revealed a cluster of testis-specific expressed genes that are mostly silenced in normal hPSCs. Analysis of the genome-wide transcriptional response following *DNMT1* disruption also revealed a significant enrichment of testis-specific expressed genes, which were

associated with Gene Ontology (GO) terms such as gonadal development, male gamete generation, and oligospermia, without an associated induction of the other three embryonic germ-layer marker genes (Figures 1F, 1G, S2A, and S2B; Table S2). The activation of germ cell genes upon DNA demethylation is in line with the global erasure of DNA methylation during normal development of primordial germ cells (Guo et al., 2015), suggesting a role for the DNA-demethylation process in the activation of germ cell differentiation. The association between Xq27-28 testis-specific gene induction and *FMR1* activation might have specific implications on *FMR1* regulation, which was previously linked with testicular differentiation (Reyniers et al., 1993; Bakker et al., 2000).

Establishment of a screening protocol for the identification of genes involved in *FMR1* silencing

To search for novel regulators involved in maintaining *FMR1* inactivation, we next aimed to conduct a LOF genetic screen using a genome-wide CRISPR-Cas9 library established in haploid hESCs. Since haploid hESCs harbor a standard range of CGG repeats and actively express *FMR1*, we sought to generate a tractable model system to analyze *FMR1* regulation in hPSCs without an endogenous CGG expansion. For this aim, we used a reporter plasmid in which the enhanced green fluorescent protein (EGFP) was placed under the control of a human *FMR1* minimal promoter (*pFMR1*), continued by the *FMR1* 5' UTR sequence, starting from 112 bp upstream to the CGG repeats to 68 bp downstream the repetitive sequence (Sølvsten and Nielsen, 2011). The continuous *FMR1* promoter and 5' UTR fragment with 240 CGG repeats were positioned upstream to the rabbit β -globin intron II followed by the EGFP reporter in a similar position concerning the mutant CGG repetitive tract as the coding sequence of the endogenous *FMR1* gene in FXS patients (Figure 2A).

To avoid the effect of transgene-integration-site variability, we tested the expression of the *pFMR1*-CGG-EGFP construct using transient transfection in haploid hESCs (Figure 2A). As expected, the occurrence of a full-mutation length ($n = 240$) CGG repeat tract *per se* was not sufficient for the transcriptional inactivation of EGFP expression (Figure S3A). This observation is in line with previous studies showing that unmethylated CGG expansions are expressed in FXS ESCs without acquiring *de novo* DNA hypermethylation (Vershkov and Benvenisty, 2017).

(E) Heatmap of expression levels (Z score transcript per million [TPM]) across tissues (data from the GTEx study) of the significantly upregulated Xq27-28 genes (FDR < 0.05) in FXS-iPSCs following *DNMT1* disruption.

(F) Enriched GO and Human Phenotype Ontology terms (analyzed using GSEA, FDR $q < 0.05$) among the upregulated genes in FXS-iPSCs following *DNMT1* disruption.

(G) Activation of testis-specific marker genes involved in transcriptional regulation upon *DNMT1* disruption (mean \pm SEM values, $n = 3$ *DNMT1* perturbation, $n = 4$ control FXS-iPSCs).

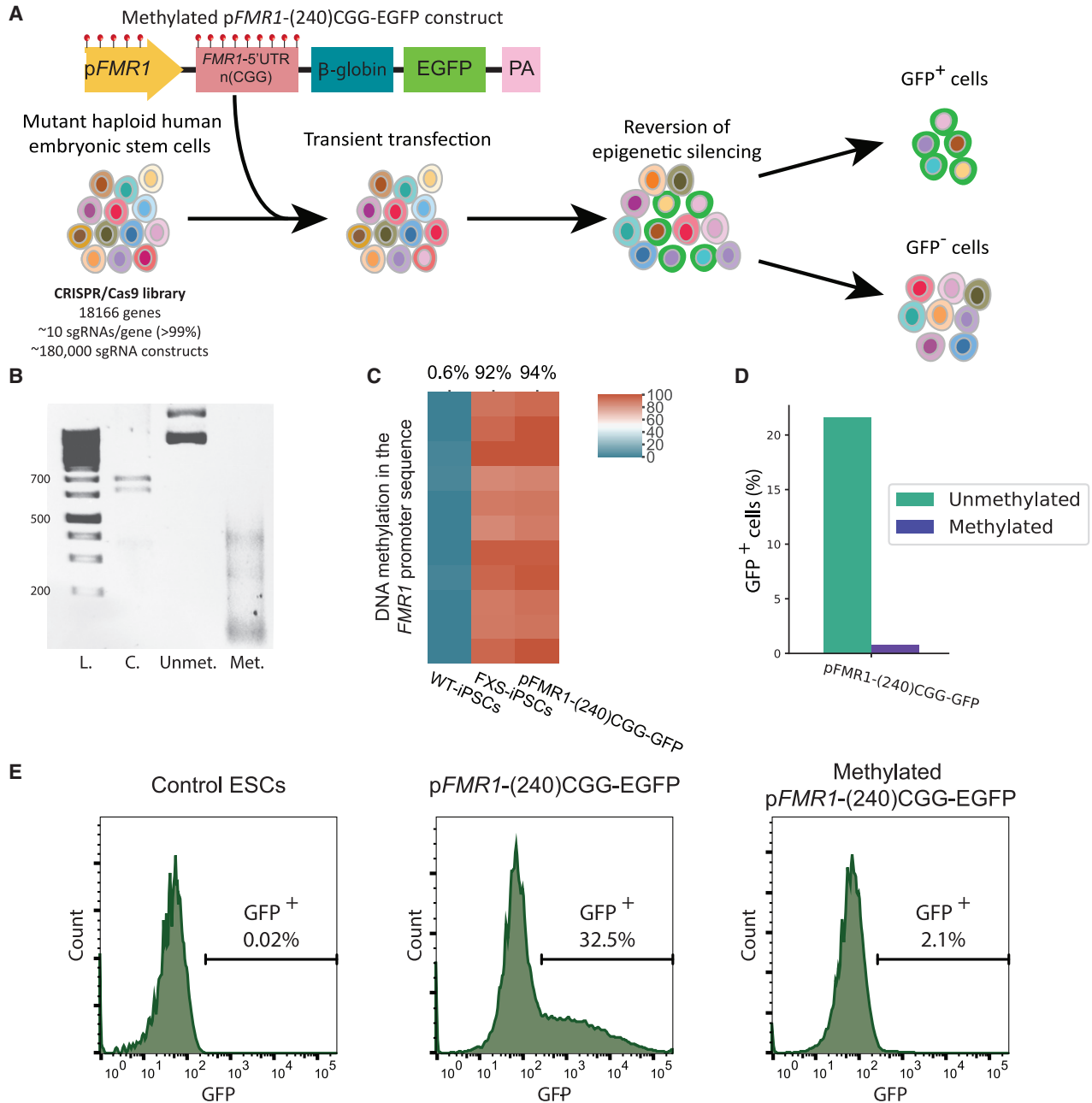


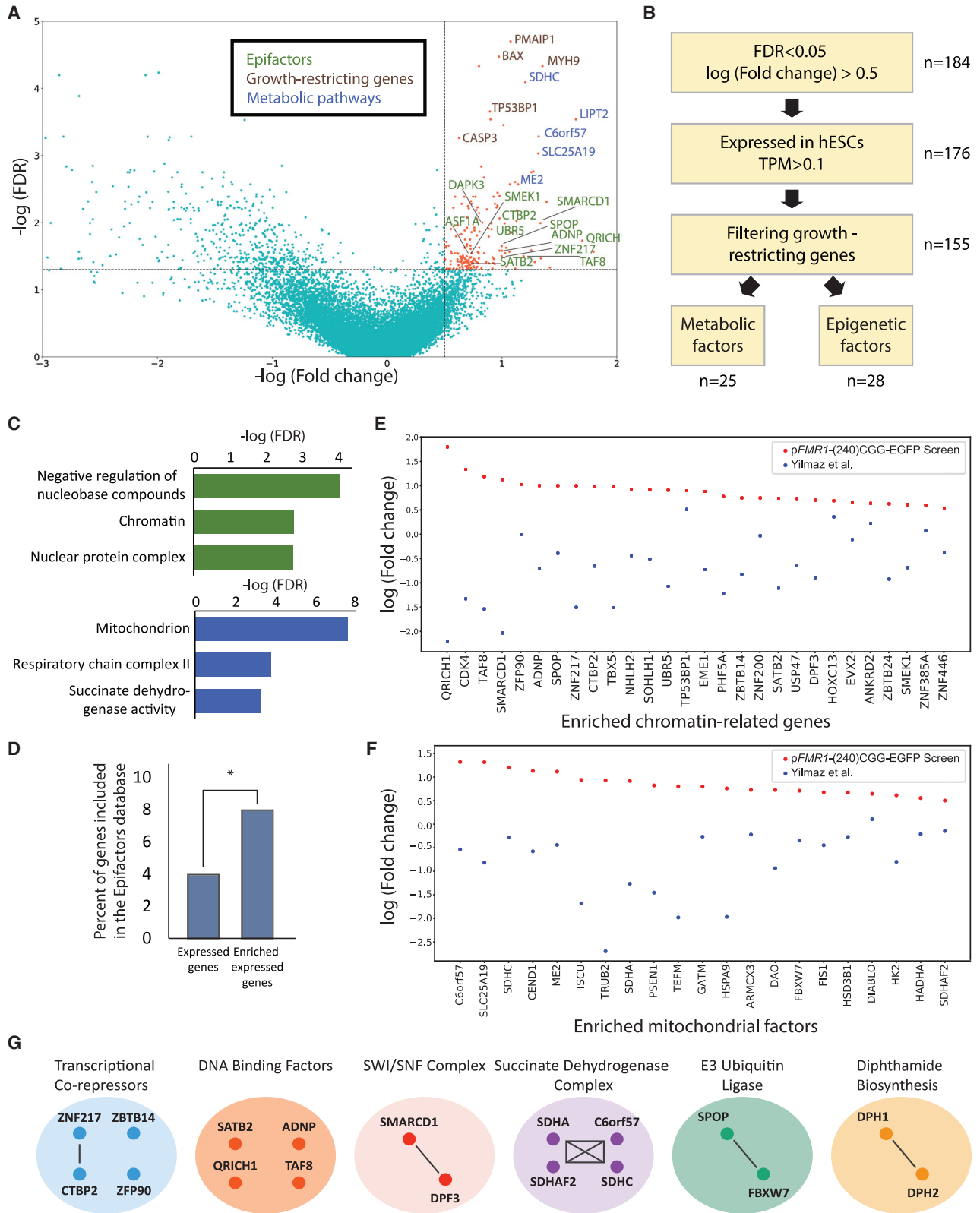
Figure 2. Establishment of a screening protocol for the identification of genes involved in *FMR1* silencing

(A) Schematic illustration depicting the LOF genetic screening experimental setup. Haploid hESCs transduced with the lentiviral CRISPR-Cas9 sgRNA library were transfected with the methylated pFMR1-(240)CGG-EGFP construct in 4 independent replicates. The least (~30%) and most (3%–4%) GFP-fluorescent cells were sorted 48 h following transfection, and DNA sequencing of the sgRNA segment was used to analyze the distribution of sgRNAs within the GFP(+) and GFP(-) populations.

(B) DNA methylation analysis of the pFMR1-(240)CGG-EGFP construct using the McrBC restriction enzyme, which cleaves DNA containing methylcytosine on one or both strands. From left to right: 1 – ladder, 2 – control plasmid with one McrBC site, 3 – unmethylated construct, and 4 – methylated construct.

(C) Pyrosequencing analysis of DNA methylation of the *FMR1* promoter sequence (in 11 CpG positions) in the methylated pFMR1-CGG(240)-EGFP plasmid, and the corresponding genomic CpG positions in FXS- and WT-iPSCs (positions (-456) to (-409) from the start site of *FMR1* translation).

(D and E) Transient transfection of haploid human ESCs with either an unmethylated (bottom middle panel) or methylated (bottom right panel) construct, followed by flow cytometry analysis 48 h post transfection.



(legend on next page)



Therefore, to induce the epigenetic repression of the p*FMRI*-(240)CGG-EGFP construct, we used *in vitro* methylation using the recombinant CpG methyltransferase M.SssI. DNA methylation of the construct was validated by its digestion with the methylation-sensitive McrBC restriction enzyme and by bisulfite-pyrosequencing analysis (Figures 2B and 2C). *In vitro* methylation using M.SssI efficiently silenced p*FMRI*-(240)CGG-EGFP expression following transient transfection, with >10-fold enrichment of the GFP-positive cell fraction between the cultures transfected with unmethylated and methylated constructs (Figures 2D and 2E). Interestingly, pre-treatment with the demethylating agent 5-azadC was associated with higher GFP fluorescence upon transfection with the methylated construct, suggesting that the depletion of DNA-methylation machinery interferes with the epigenetic silencing of the reporter plasmid (Figure S3B).

Using an LOF genome-wide library to screen for genes involved in *FMRI* inactivation

Next, we applied our assay to the CRISPR-Cas9 haploid hESCs library, which contains 178,896 different gRNA constructs, targeting 18,166 genes (Yilmaz et al., 2018). Forty-eight hours following transfection with the methylated p*FMRI*-(240)CGG-EGFP construct, library cells were harvested and sorted to GFP(+) and GFP(-) populations (Figure 2A). The abundance of different gRNAs represented in both populations was assessed by the amplification of the sgRNA-containing genomic DNA segment and high-throughput sequencing. Following the mapping of the reads to the sgRNA sequences, an enrichment score was assigned to each gene by calculating the log₂ fold change (FC) of its sgRNA counts between GFP(+) (n = 4) and GFP(-) (n = 3) populations to allow for the identification of genes predicted to be involved in silencing the expression of the methylated p*FMRI*-(240)CGG-EGFP.

Analysis of the significantly enriched genes revealed several functionally related gene groups (Figures 3A–3C; Tables S3 and S4): First, a subset of the candidate genes was related to chromatin regulation and transcriptional repression, identified either by the Epifactors database (Medvedeva et al., 2015), as being listed in databases of transcription factors or as associated with chromatin-related GO annotations (Lambert et al., 2018). Functional-annotation analysis revealed a significant association of the top enriched genes with GO terms related to chromatin regulation (Figure 3C; Table S2), and a canonical-pathway analysis demonstrated a significant association with the reactome of RNA polymerase II (FDR < 0.0001). Specifically, genes included in the Epifactor database were significantly enriched in the candidate list compared with their representation in the library (7.6% versus 3.8%, p value = 0.02 using Fisher's exact test; Figure 3D). Interestingly, another subset of the enriched genes was related to several metabolic pathways, specifically the mitochondrial respiration pathway, including four different subunits that assemble the succinate dehydrogenase complex (Figures 3E and 3F). Finally, some enriched genes were categorized as growth-restricting genes in hPSCs (Figure 3A) (Yilmaz et al., 2018). Although cell-cycle-related genes may influence the epigenetic landscape, these genes might be overrepresented in the screen because their disruption confers a growth advantage under selection pressure. Other enriched categories might be related to confounding factors as transfection efficiency (e.g., the internalization and intracellular trafficking of the plasmid DNA). To filter out genes that were overrepresented due to selection pressure, we excluded genes identified as growth restricting in hESCs (FDR < 0.05, CRISPR score > 1). This led to the establishment of a candidate list of 155 genes predicted to be involved in maintaining gene repression, 28 of which

Figure 3. Screening for genes involved in *FMRI* inactivation using CRISPR-Cas9 library in haploid hESCs

(A) Volcano plot showing the median log₂ FC (x axis) and -log(FDR) (y axis) of the mutants included in the library, calculated based on the distribution of the normalized sgRNA read counts between the GFP(+) and GFP(-) populations (n = 4 and n = 3 independent experiments, respectively). Marked in orange are genes defined as enriched (log₂ FC > 0.5 and FDR < 0.05). Representative genes of the main groups among the enriched genes are indicated in green (genes included in the "Epifactors" database), blue (metabolic factors), and red (growth-restricting genes).

(B) Flowchart demonstrating the analysis pipeline for defining the candidate genes predicted to be involved in *FMRI* inactivation.

(C) Enriched GO terms (analyzed by GSEA, FDR q < 0.05) for the top significant genes (average log₂ FC of gRNA abundance > 0.5 and FDR < 0.05, in genes expressed in haploid hESCs [TPM > 0.1]). Two groups of enriched terms: terms associated with chromatin regulation (top) and terms associated with mitochondrial function (bottom).

(D) Percentage of the significantly enriched genes included in the Epifactors database, compared with their representation in the library (p = 0.02 using Fisher's exact test).

(E and F) log₂ FC of enriched epigenetic (E) or mitochondrial (F) factors in the m*FMRI*-(240)CGG-GFP screen, compared with their log₂ FC values, also known as CRISPR score, in a previous essentiality screen in hESCs (Yilmaz et al., 2018).

(G) Representative genes for different functional groups within the candidate gene list. Black lines connecting the gene names indicate protein-protein interactions identified by STRING analysis.



were previously shown to have a role in chromatin regulation (Figure 3E; Table S1). Among these genes were transcriptional co-repressors (e.g., *ZNF217*, *ZFP90*, *CTBP2*), chromatin remodeling factors (e.g., *SMARCD1*), and RNA polymerase II transcription-initiation factors (e.g., *TAF8*) (Figure 3G). The enrichment of both epigenetic and mitochondrial factors among the GFP-positive population was not correlated with their association with growth restriction in a previous essentiality screen in haploid hPSCs (Figures 3E and 3F) (Yilmaz et al., 2018).

Characterization of genes predicted to be involved in *FMR1* epigenetic silencing

Next, we aimed to validate the effect of disrupting selected candidate genes on the expression of the methylated p*FMR1*-(240)CGG-EGFP construct. Using transduction with CRISPR-Cas9 and two sgRNAs per gene, the selected genes were mutated in haploid hESCs. Mutant cultures were then transfected with both unmethylated and methylated p*FMR1*-(240)CGG-EGFP constructs, and the GFP fluorescent population was compared between the two cultures (Figure 4A). This way, we could isolate the epigenetic effect of candidate-gene disruption and exclude confounding factors as transfection efficiency or translational control. Haploid hESCs infected with lentiviral vectors targeting the transcriptional regulators *SMARCD1* and *ZNF217*, as well as the metabolic factor *C6orf57*, showed the highest levels of methylated-construct expression compared with control haploid hESCs, demonstrating a significant increase in relative GFP fluorescence upon methylated-construct transfection, reflecting the disinhibition of DNA-methylation-mediated silencing (Figure 4B).

Finally, as transiently transfected promoters might not be subjected to the same regulatory mechanisms that operate on the endogenous *FMR1* promoter, we infected FXS-iPSCs with lentiviral constructs containing Cas9 and sgRNAs targeting the candidate genes. RT-PCR analysis of the mutated samples revealed some increase above basal *FMR1* transcription levels following the disruption of *SMARCD1*, *ZNF217*, and *C6orf57* (Figure 4C). However, the effect was lower than the reactivation effects observed with *DNMT1* perturbation or demethylating treatment. Although none of the mutated samples reached the *FMR1* expression levels associated with *DNMT1* perturbation, the relative increase in *FMR1* mRNA in the mutant samples suggests an interaction of these genes with the endogenously silenced *FMR1* locus. DNA-methylation analysis of the *FMR1* promoter in the mutated samples identified an overall decrease in DNA methylation in the *C6orf57* mutant, but it did not reach statistical significance ($p = 0.09$; Figure S3C). This could reflect the lower sensitivity of the pyrosequencing assay in a heterogeneous population. To further characterize the regulatory effect of the identified genes, we

performed a global transcriptional analysis of the *ZNF217*- and *SMARCD1*-mutated samples. Upon comparison of up- and downregulated genes ($p < 0.001$ and $|FC| > 1$), the mutated samples showed a bias toward a positive effect on gene expression, supporting the repressive function of these genes (Figure 4D). Analysis of the enriched gene sets among the upregulated genes following *SMARCD1* or *ZNF217* disruption identified overlapping regulators of both gene groups, pointing to common regulatory pathways for these target genes (Figure 4E).

DISCUSSION

In this study, we have used genome-wide CRISPR-Cas9 screening to identify genes involved in *FMR1* inactivation. The characterization of the transcriptional outcomes of *DNMT1* disruption in FXS-iPSCs revealed *FMR1* as one of the most upregulated genes in the mutated samples and the *FMR1*-adjacent genomic region as the most enriched locus in the transcriptional response to the loss of *DNMT1*. The ability of *DNMT1* disruption alone to fully reproduce the effect size of the demethylating agent 5-azadC, which binds the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B (Oka et al., 2005), shows that the reactivating effect of this compound is mediated primarily through DNMT1, with a lesser or no role for DNMT3A and DNMT3B in maintaining *FMR1* inactivation in PSCs. As the cytotoxic effect of 5-azadC was previously shown to be mediated primarily by DNMT3A and DNMT3B (Oka et al., 2005), this finding calls for the use of selective DNMT1 inhibitors that will be able to provide robust *FMR1* reactivation with reduced toxicity compared with 5-azadC.

The substantial overexpression of genes located in Xq27-28 following *DNMT1* disruption reflects the inherent epigenetic features of this chromosomal region, which appears to be tightly regulated by cellular DNA-methylation levels. The association of *FMR1* activation with the induction of the adjacent testis-specific genes is of particular interest: first, *FMR1* itself is known to be particularly highly expressed in both adult and fetal testis (Bakker et al., 2000). Second, in carriers of full-mutation *FMR1* alleles, the CGG repeats become unstable during spermatogenesis, leading to active, pre-mutation-length *FMR1* alleles in the sperm of fragile X patients (Reyniers et al., 1993). The co-activation of the expanded *FMR1* locus and the adjacent male gamete-specific expression program might thus point to the link between male gamete differentiation and *FMR1* epigenetic regulation. The activation of testis-specific regions during spermatogenesis may be associated with the recruitment of transcriptional activators and chromatin remodeling enzymes, therefore altering the epigenetic landscape of the Xq27-28 chromosomal region. These results

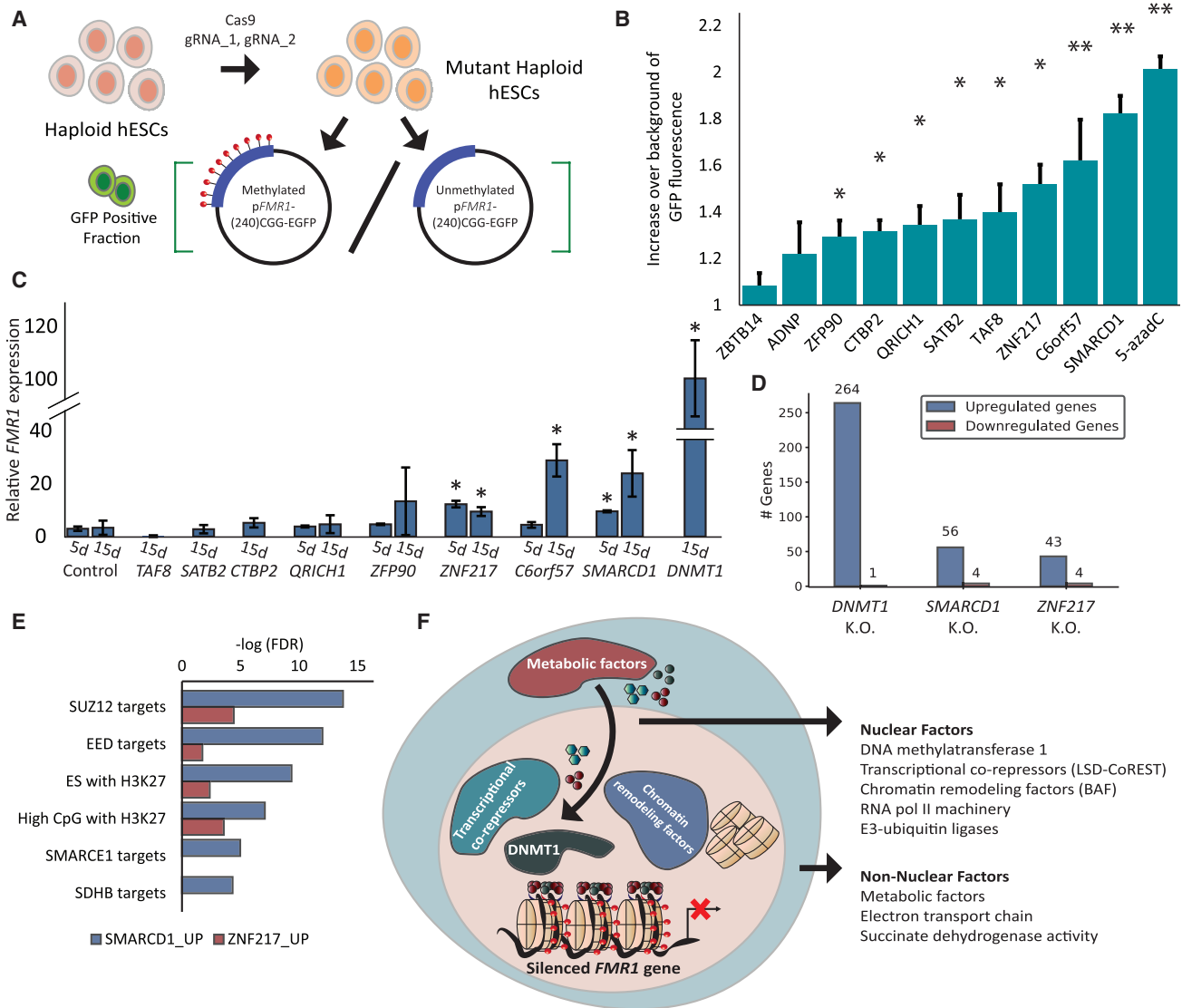


Figure 4. Verification of candidate hit genes predicted to be involved in *FMR1* inactivation

(A) Experimental workflow. Mutant haploid hESCs were transfected with either a methylated or unmethylated pFMR1-(240)CGG-EGFP construct and analyzed by flow cytometry.

(B) Analysis of the mutants of candidate genes reveals higher relative GFP fluorescence following transfection with methylated pFMR1-(240)CGG-EGFP, relative to samples infected with Cas9 without sgRNA. Bars indicate the normalized ratio of GFP fluorescence between the cell cultures transfected with methylated and unmethylated constructs. Data are presented as mean values \pm SEM, and statistical tests were performed with 3 independent experiments.

(C) Analysis of heterogeneous mutant populations of FXS-iPSCs for candidate genes *SMARCD1*, *C6orf57*, *ZNF217*, *ZFP90*, *CTBP2*, *SATB2*, *QRICH1*, and *TAF8*. Statistical tests were performed from at least 3 independent experiments. Error bars represent SEM.

(D) Bar plot showing the number of the significantly up- (blue) and downregulated (red) genes for the *DNMT1*, *SMARCD1*, and *ZNF217* samples ($p < 0.001$, $|\log \text{FC}| > 1$).

(E) Enriched gene sets among the genes upregulated following either *SMARCD1* or *ZNF217* disruption.

(F) Schematic representation summarizing a proposed model of *FMR1* epigenetic regulation following CGG repeat expansion. * $p < 0.05$.

call for the analysis of the epigenetic status of full-mutation alleles during spermatogenesis.

As *DNMT1* disruption resulted in only partial reactivation of *FMR1* expression, to levels comparable with

5-azadC treatment, we turned to search for additional mechanisms that contribute to *FMR1* inactivation. For this aim, we developed a genetic screening platform that provides a traceable readout of the transcriptional output



of an exogenous methylated *FMRI* promoter followed by an expanded CGG repeat tract. This platform allowed us to utilize haploid hPSCs for the study of *FMRI* silencing, although these cells harbor a normal range of CGG repeats.

Transient transfection of reporter plasmids was previously used for studying the relationship between CGG repeat length and the regulation of gene expression (Sølvsten and Nielsen, 2011). As exogenous plasmid DNA is assembled into nucleosome structures and becomes associated with the transcription machinery (Smith and Hager, 1997), this system allows for the analysis of an interplay of different layers of transcriptional regulation that might be relevant to FXS pathophysiology. Although the transcriptional regulation of episomal vectors is different than their endogenous counterparts, this system presented several important advantages: first, since DNA methylation of CpG sites is known to be affected by the methylation status of adjacent sequences, transient transfection allows us to avoid the locus-specific effects of random genomic integration. Second, transient vectors rarely carry mammalian replication origins, which allowed us to test the epigenetic regulation of *FMRI* independently of cell replication, which is of interest because neurons, the disease-relevant cell type in FXS, do not proliferate.

While transfection with a plasmid containing full-mutation-length CGG repeats was not sufficient to repress EGFP expression, *in vitro* methylation of the construct efficiently silenced the reporter gene, establishing a clear and traceable phenotype suitable for large-scale screening. Using our screen, we established a list of 155 candidate genes potentially important for *FMRI* regulation. As expected, among the enriched genes, there was a substantial fraction of epigenetic regulators. *ZNF217*, an enriched gene selected for further validation, is a member of the LSD-CoREST complex, which is known to repress the expression from promoters by the recruitment of C-terminal binding proteins (CtBPs), one of which was also enriched in our screen (*CTBP2*). *SMARCD1*, another candidate gene, is a subunit of the SWI/SNF complex, which is involved in transcriptional regulation by chromatin remodeling. In hESCs, *SMARCD1* knockdown resulted in chromatin de-condensation with reduced heterochromatin foci (Alajem et al., 2015). Moreover, *SMARCD1* knockdown led to an aberrant gene expression profile, including failure in silencing the pluripotency network upon differentiation. We also note that one of the enriched transcriptional regulators, *ZBTB14*, was previously found to repress the transcription of CGG-repeat-containing elements (Orlov et al., 2007).

Besides the enrichment of epigenetic factors, analysis of the candidate genes list revealed enrichment of several metabolic pathways. This observation may be explained by the well-known influence of metabolic enzymes on the epigenome, mainly by catalyzing the production and

degradation of metabolites that function as substrates, cofactors, or inhibitors of chromatin-modifying enzymes (Kaelin and McKnight, 2013). *C6orf57* plays a role in the assembly of succinate dehydrogenase (SDH) complex. Besides *C6orf57*, the SDH complex subunits *SDHC*, *SDHA*, and *SDHAF2* were also found to be enriched in our screen. Inactivating *SDH* mutations were observed in several types of tumors, leading to succinate accumulation and inhibition of 2-oxoglutarate-dependent dioxygenases that are responsible for histone demethylation (Cervera et al., 2009). That *SDH* disruption led to enhanced expression of the methylated *FMRI* promoter construct points to the importance of epigenetic regulation by this metabolic pathway and may reflect its differential influence on both histone and DNA methylation.

Overall, our results suggest a complex model of epigenetic regulation that merges different levels of cellular organization: factors involved directly in transcription regulation and nucleosome assembly but also non-nuclear factors that determine cellular metabolism (Figure 4F). This model suggests that the *FMRI* locus regulation might be tightly linked to developmental time point, cell identity, and metabolic state. In addition, the increased sensitivity of Xq27-28 genes to *DNMT1* disruption highlights the contribution of regional and cell-type-specific epigenetic features in FXS pathogenesis.

The analysis of the consequences of candidate-gene disruption on the methylated *pFMRI*-(240)CGG-EGFP expression demonstrated the disinhibition of the methylated construct in selected mutants. Targeting of *SMARCD1*, *ZNF217*, and *C6orf57* in FXS-iPSCs was associated with some increase above the basal levels of *FMRI* expression, suggesting the interaction of these genes with the inactive *FMRI* locus. The fact that disruption of no single factor besides *DNMT1* resulted in robust *FMRI* expression and that even *DNMT1* mutagenesis induced only partial levels of *FMRI* mRNA may reflect the robustness of endogenous *FMRI* silencing compared with the exogenous construct, possibly due to functional redundancy of different repressive mechanisms. It is possible that two or more factors should be targeted to achieve robust *FMRI* expression or that the screen was not fully comprehensive in mimicking the physiological context of *FMRI* inactivation.

Collectively, this work demonstrates the utility of LOF genetic screening in the study of genetic and epigenetic diseases. We use a system of an exogenous *pFMRI*-(240)CGG-EGFP reporter as a model for *FMRI* silencing and highlight gene networks associated with transcriptional regulation of *FMRI* expression. Our results prompt further investigation regarding the involvement of epigenetic and metabolic factors in the induction and maintenance of *FMRI* inactivation and present novel targets for *FMRI* reactivating therapy.



EXPERIMENTAL PROCEDURES

Cell lines

Throughout the study, we used the following cell lines: the FXS-iPSCs male cell line A52 (Vershkov et al., 2019) and haploid hpESCs - hPES10 (Sagi et al., 2016). CSES7 and CSES9 hESCs and their derivatives were used as a reference for wild-type expression (Biancotti et al., 2010). 293T cells from R. Weinberg (Whitehead Institute) were used for the construction of lentiviral constructs.

In vitro DNA methylation

The pFMR1-n(CGG)-EGFP construct (Sølvsten and Nielsen, 2011) was methylated using an M.SssI CpG methyltransferase (NEB) according to the manufacturer's recommendations. The S-adenosylmethionine (SAM) concentration was adjusted to 640 μ M. DNA was extracted using ethanol precipitation.

CRISPR-Cas9 LOF library

For the genome-wide screen, we used a CRISPR-Cas9-based genome-wide LOF library of haploid hESCs previously established in our lab (Yilmaz et al., 2018). Briefly, haploid-enriched ESC cultures of the hpES10 cell line were infected with a lentivirus CRISPR-Cas9 genome-wide library at a multiplicity of infection of 0.3. Infected cells were then selected with puromycin (Sigma) and cultured for about fifteen doublings before harvesting and freezing. The library was then thawed and cultured at 37°C with 5% CO₂ in feeder-free conditions using matrigel-coated plates (Corning) and mTeSR1 medium (STEMCELL Technologies) supplemented with 10 μ M ROCK inhibitor (Y27632, Stemgent) for 1 day after thawing or splitting.

Library transfection

Ten μ M ROCK inhibitor (Y27632) was added 1 h prior to trypsinization and transfection. Library cells were harvested using TrypLE Select Enzyme solution (Thermo Fisher Scientific, cat no. 12563029) into a single-cell state. Following the first centrifugation, cells were re-suspended twice with DMEM/F12 and centrifuged again before incubation with the plasmid DNA. Transfection of the methylated pFMR1-CGG(240)-EGFP construct was performed using the Xtreme Gene9 reagent (Roche) using standard conditions. Forty-eight h following transfection, cells were washed with PBS, harvested using TrypLE Select, re-suspended in PBS supplemented with 15% fetal bovine serum (FBS), filtered through a 70 μ m cell strainer (Corning), and sorted using BD FACSAria III to GFP(+) and GFP(-) populations. Transfection efficiency was assessed using the simultaneous transfection of control haploid hESCs with an unmethylated pFMR1-CGG(240)-EGFP construct.

Data and code availability

The accession number for the CRISPR-Cas9 library sequencing data reported in this paper is in GEO: GSE182551. The accession number for the RNA sequencing (RNA-seq) data reported in this paper is in GEO: GSE182391.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.stemcr.2022.03.011>.

AUTHOR CONTRIBUTIONS

D.V. and N.B. designed the experiments and interpreted the data. D.V. performed the experiments with a significant contribution from O.Y. and analyzed the data. A.L.N. provided the pFMR1-(240)CGG-EGFP plasmid. A.Y. assisted with the establishment of the LOF library in haploid hESCs. D.V. and N.B. wrote the manuscript with input from all other authors. N.B. supervised the study.

CONFLICT OF INTERESTS

N.B. is CSO of NewStem, Ltd.

ACKNOWLEDGMENTS

The authors would like to thank Tamar Golan-Lev and Aviva Petcho for their help with tissue cultures and Elyad Lezmi for critically reading the manuscript. The study was partially supported by the Azrieli Foundation by the Rosetrees Trust and by the ISF-Israel Precision Medicine Partnership (IPMP) Program (no. 3605/21). N.B. is the Herbert Cohn Chair in Cancer Research.

Received: September 9, 2021

Revised: March 16, 2022

Accepted: March 17, 2022

Published: April 14, 2022

REFERENCES

- Alajem, A., Biran, A., Harikumar, A., Sailaja, B.S., Aaronson, Y., Livyatan, I., Nissim-Rafinia, M., Sommer, A.G., Mostoslavsky, G., Gerbasi, V.R., et al. (2015). Differential association of chromatin proteins identifies BAF60a/SMARCD1 as a regulator of embryonic stem cell differentiation. *Cell Rep.* 10, 2019–2031.
- Bakker, C.E., de Diego Otero, Y., Bontekoe, C., Raghoe, P., Luteijn, T., Hoogeveen, A.T., Oostra, B.A., and Willemsen, R. (2000). Immunocytochemical and biochemical characterization of FMRP, FXR1P, and FXR2P in the mouse. *Exp. Cell Res.* 258, 162–170.
- Biancotti, J.-C., Narwani, K., Buehler, N., Mandefro, B., Golan-Lev, T., Yanuka, O., Clark, A., Hill, D., Benvenisty, N., and Lavon, N. (2010). Human embryonic stem cells as models for aneuploid chromosomal syndromes. *Stem Cells* 28, 1530–1540.
- Cervera, A.M., Bayley, J.-P., Devilee, P., and McCreath, K.J. (2009). Inhibition of succinate dehydrogenase dysregulates histone modification in mammalian cells. *Mol. Cancer* 8, 89.
- Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.-Y., Gerhardt, J., Disney, M.D., and Jaffrey, S.R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* 343, 1002–1005.
- Epsztejn-Litman, S., and Eiges, R. (2019). Monitoring for epigenetic modifications at the *FMR1* locus. *Methods Mol. Biol.* 1942, 29–48.
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The transcriptome and DNA methylation landscapes of human primordial germ cells. *Cell* 161, 1437–1452.
- Kaelin, W.G., and McKnight, S.L. (2013). Influence of metabolism on epigenetics and disease. *Cell* 153, 56–69.



- Kumari, D., and Usdin, K. (2010). The distribution of repressive histone modifications on silenced *FMR1* alleles provides clues to the mechanism of gene silencing in fragile X syndrome. *Hum. Mol. Genet.* *19*, 4634–4642.
- Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. *Cell* *172*, 650–665.
- Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C., et al. (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* *47*, 469–478.
- Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I.V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T., and Drabløs, F. (2015). EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database* *2015*, bav067.
- Oka, M., Meacham, A.M., Hamazaki, T., Rodić, N., Chang, L.-J., and Terada, N. (2005). De novo DNA methyltransferases Dnmt3a and Dnmt3b primarily mediate the cytotoxic effect of 5-aza-2'-deoxycytidine. *Oncogene* *24*, 3091–3099.
- Orlov, S.V., Kuteykin-Teplyakov, K.B., Ignatovich, I.A., Dizhe, E.B., Mirgorodskaya, O.A., Grishin, A.V., Guzhova, O.B., Prokhortchouk, E.B., Guliy, P.V., and Perevozchikov, A.P. (2007). Novel repressor of the human *FMR1* gene – identification of p56 human (GCC)_n-binding protein as a Krüppel-like transcription factor ZF5. *FEBS J.* *274*, 4848–4862.
- Reyniers, E., Vits, L., De Boulle, K., Van Roy, B., Van Velzen, D., de Graaff, E., Verkerk, A.J., Jorens, H.Z., Darby, J.K., and Oostra, B. (1993). The full mutation in the *FMR-1* gene of male fragile X patients is absent in their sperm. *Nat. Genet.* *4*, 143–146.
- Sagi, I., Chia, G., Golan-Lev, T., Peretz, M., Weissbein, U., Sui, L., Sauer, M.V., Yanuka, O., Egli, D., and Benvenisty, N. (2016). Derivation and differentiation of haploid human embryonic stem cells. *Nature* *532*, 107–111.
- Smeets, H.J., Smits, A.P., Verheij, C.E., Theelen, J.P., Willemsen, R., van de Burgt, I., Hoogeveen, A.T., Oosterwijk, J.C., and Oostra, B.A. (1995). Normal phenotype in two brothers with a full *FMR1* mutation. *Hum. Mol. Genet.* *4*, 2103–2108.
- Smith, C.L., and Hager, G.L. (1997). Transcriptional regulation of mammalian genes *in vivo*. *J. Biol. Chem.* *272*, 27493–27496.
- Sølvsten, C., and Nielsen, A.L. (2011). *FMR1* CGG repeat lengths mediate different regulation of reporter gene expression in comparative transient and locus specific integration assays. *Gene* *486*, 15–22.
- Vershkov, D., and Benvenisty, N. (2017). Human pluripotent stem cells in modeling human disorders: the case of fragile X syndrome. *Regen. Med.* *12*, 53–68.
- Vershkov, D., Fainstein, N., Suissa, S., Golan-Lev, T., Ben-Hur, T., and Benvenisty, N. (2019). *FMR1* reactivating treatments in fragile X iPSC-derived neural progenitors *in vitro* and *in vivo*. *Cell Rep.* *26*, 2531–2539.
- Yanovsky-Dagan, S., Mor-Shaked, H., and Eiges, R. (2015). Modeling diseases of noncoding unstable repeat expansions using mutant pluripotent stem cells. *World J. Stem Cells* *7*, 823.
- Yilmaz, A., Peretz, M., Sagi, I., and Benvenisty, N. (2016). Haploid human embryonic stem cells: half the genome, double the value. *Cell Stem Cell* *19*, 569–572.
- Yilmaz, A., Peretz, M., Aharony, A., Sagi, I., and Benvenisty, N. (2018). Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat. Cell Biol.* *20*, 610–619.
- Zhou, Y., Kumari, D., Sciascia, N., and Usdin, K. (2016). CGG-repeat dynamics and *FMR1* gene silencing in fragile X syndrome stem cells and stem cell-derived neurons. *Mol. Autism* *7*, 42.