GENETICS

Suppression of trinucleotide repeat expansion in spermatogenic cells in Huntington's disease

In K. Cho[1,2,3,4,5] · Charles A. Easley IV[2,3,4] · Anthony W. S. Chan[1,2,6]

Abstract

Trinucleotide repeats (TNRs) are dispersed throughout the human genome. About 20 loci are related to human diseases, such as Huntington's disease (HD). A larger TNR instability is predominantly observed in the paternal germ cells in some TNR disorders. Suppressing the expansion during spermatogenesis can provide a unique opportunity to end the vicious cycle of genetic anticipation. Here, using an in vitro differentiation method to derive advanced spermatogenic cells, we investigated the efficacy of two therapeutic agents, araC (cytarabine) and aspirin, on stabilizing TNRs in spermatogenic cells. Two WT patient-derived induced pluripotent stem cell (iPSC) lines and two HD hiPSC lines, with 44 Q and 180 Q, were differentiated into spermatogonial stem cell-like cells (SSCLCs). Both HD cell lines showed CAG tract expansion in SSCLC. When treated with araC and aspirin, HD1 showed moderate but not statistically significant stabilization of TNR. In HD2, 10 nM of aspirin and araC showed significant stabilization of TNR. All cell lines showed increased DNA damage response (DDR) gene expression in SSCLCs while more genes were significantly induced in HD SSCLC. In HD1, araC and aspirin treatment showed general suppression of DNA damage response genes. In HD2, only *FAN1*, *OGG1*, and *PCNA* showed significant suppression. When the methylation profile of HD cells was analyzed, *FAN1* and *OGG1* showed significant hypermethylation after the aspirin and araC treatment in SSCLC compared to the control. This study underscores the utility of our in vitro spermatogenesis model to study and develop therapies for TNR disorders such as HD.

Keywords Huntington's disease · Trinucleotide repeats · Spermatogenesis · Drug discovery · Cytarabine (AraC) · Aspirin

Introduction

Trinucleotide repeats (TNRs) are dispersed throughout the human genome, and about 20 loci are related to human diseases, such as Huntington's disease (HD), fragile X syndrome (FXS), and spinocerebellar ataxia (SCA) [1–3]. HD is a monogenic neurodegenerative disease characterized by progressive brain atrophy in the striatum, cortex, and other brain regions. Individuals with more than 35 CAG repeats (polyglutamine repeats; polyQ) in exon 1 of the huntingtin (*HTT*) gene will develop HD [4]. Furthermore, the age of onset is inversely correlated to the size of the CAG repeat. TNRs are unstable and are prone to changes (i.e., expansions and contractions). Although peripheral tissues and central nervous tissues, such as the brain, show TNR expansions in the various animal models and human patients [5–7], a larger TNR instability is predominantly observed in the paternal germ cells in TNR disorders, including HD and SCA. The expansion in germ cells results in genetic anticipation where paternal progeny gets expanded TNR [8–11]. Male

✉ In K. Cho
  inkicho@uga.edu

1 Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA

2 Division of Neuropharmacology and Neurologic Diseases, Emory National Primate Research Center, Emory University, Atlanta, GA, USA

3 Department of Environmental Health Sciences, College of Public Health, University of Georgia, Athens, GA, USA

4 Regenerative Bioscience Center, University of Georgia, Athens, GA, USA

5 Environmental Health Science and Regenerative Bioscience Center, College of Public Health, University of Georgia, Edgar L. Rhodes Center for Animal and Dairy Science RM 432, 425 River Rd, Athens, GA 30602, USA

6 Present Address: Center of Scientific Review (CSR), National Institutes of Health, Bethesda, USA

spermatogenesis is a multistep process involving DNA replication, ligation, proof-reading, and repair, providing ample opportunities for errors, such as CAG expansion. However, it is still unknown whether these CAG repeats can be stabilized to reduce the expansion in humans with small molecules. TNR expansion in rodents occurs in the post-meiotic stage [12], while expansions occur in pre-meiotic and post-meiotic sperm cells in humans [13]. The main factor influencing TNR instability is the number of repeats, which makes HD patients with larger number of polyQ repeats more vulnerable to environmental insult and susceptible to changes [14].

With the recent development of stem cell technologies, various cell types can be derived to investigate the underlying mechanism of diseases in a dish. Our team has developed and utilized in vitro directed differentiation method to derive advanced spermatogenic cells using the protocol based on our recent success in deriving haploid spermatids from human PSCs. In recent developments, directed in vitro differentiation of deriving spermatogenic cells have improved [15–17] and used to access developmental toxicity and environmental toxicants [18–21]. However, among the "gold standards" in evaluating in vitro-derived germ cells, the goal for demonstrating the faithful meiosis is the ability to fertilize oocytes and produce embryos [22]. In a recently published article, we have demonstrated that upon intracytoplasmic sperm injection, rhesus oocytes were successfully fertilized by the in vitro differentiated spermatids, and some developed into blastocytes [23]. Also, we have demonstrated that non-human primate (NHP) embryonic stem cell (nhpESC) derived spermatogonial stem cell-like cells (SSCLCs) can replicate the CAG repeat expansion in vitro [24]. We have demonstrated that the daily expansion rate of SSCLC was more significant than the expansion rates of the extended culture of nhpESC [24]. Suppression of expansion or even inducing contraction of TNR in paternal germline will provide an opportunity to eradicate the devastating effect of genetic anticipation in TNR disorders. Recent reports have suggested that TNR instabilities can be promoted by environmental stresses [25], hormonal disruption [26], and chemical exposures [27]. Gomes-Pereira and Monckton showed chronic exposure to chemical treatment could alter CAG*CTG repeat size in *Dmt*-D myotonic dystrophy mouse cell lines [27]. Of eight chemicals tested, araC and aspirin resulted in large contractions, and caffeine showed the largest expansion [27]. Therefore, we selected araC and aspirin as contracting candidates and caffeine as an expansion control.

Here, we utilized human induced pluripotent stem cell (hiPSC) to investigate the efficacy of two therapeutic agents, araC (cytarabine) and aspirin, on stabilizing TNR in spermatogenic cells using our in vitro SSCLC directed differentiation in HD. The differentiation process resulted in expanded CAG repeats in spermatogenic cells. Moreover, treating cells during the differentiation process with subclinical dosages of araC and aspirin stabilized CAG expansion in spermatogenic cells. This study demonstrates that this platform can be used to discover new therapeutic targets, provide a method to investigate the mechanisms involved in germline TNR expansion in different cell types, and has further applications in other TNR disorders with the paternal origin of TNR expansion such as dentatorubral-pallidoluysian atrophy (DRPLA), oculopharyngeal muscular dystrophy (OPMD), SCA, spinomuscular bulbar atrophy (SMBA), and myotonic dystrophy (DM).

Results

Two WT hiPSC cells (BJ-WT1 and ND41658-WT2) and two HD hiPSC cells (ND38547-HD1 with 44 Q and ND36999-HD2 with 180 Q) were differentiated into spermatogonial stem cell-like cells (SSCLCs) in vitro following the established protocol [16] (Fig. 1 and Supplementary Information Fig. 1). Differentiation efficiency of all four cell lines was evaluated by RNA-seq (Fig. 1a, Fig. 2), RT-PCR (Fig. 1b), and immunofluorescence (Fig. 1c, Supplementary Information Fig. 1). When RNA-seq data of day 10 samples (SSCLCs) were compared to iPSC, spermatogenic specific markers such as *VASA*, *ACR*, *GFRA1*, *ZBTB16*, and *MEIOB* were induced in day 10 samples compared to iPSC (Fig. 1a) ($n = 3$ for iPSC and $n = 6$ for SSCLC). Quantitative real-time polymerase chain reaction (qRT-PCR) analysis of selected genes (*OCT4*, *NANOG*, *VASA*, *DAZL*, *ZBTB16*, *GFRA1*, *ACR*, and *TNP1*) showed the significant downregulation of pluripotent stem cell markers (*OCT4* and *NANOG*, $p < 0.00005$ for both) and significant upregulation of spermatogenic specific transcripts (*VASA*, *DAZL*, *ZBTB16*, *GFRA1*, *ACR*, and *TNP1*) ($n = 3$ for each cell lines). The expression of VASA, ZBTB16, PIWIL2, DAZL, and ACR were confirmed with immunofluorescence while no OCT4 expression was detected (Fig. 1c, Supplementary Information Fig. 1). Hierarchical clustering of RNA-sequence analysis of differentially expressed (DE) genes from iPSC and SSCLC showed clustering of iPSCs and SSCLCs (Fig. 2a). Ensemble of Gene Set Enrichment Analyses (EGSEA) of significantly induced genes showed spermatogenesis (M5951) as the third-highest ranked hallmark gene signatures (number of genes = 22/135, $p = 0.036$, average log fold change (FC) = 3.4) (Fig. 2b). A total of 6133 genes were significantly differentially expressed when SSCLCs were compared to iPSCs (Fig. 2a, c). Genes involved in spermatogenesis and specifically associated with testis development were among those significantly induced genes (Fig. 2c).

To evaluate whether in vitro differentiation process can recapitulate the TNR expansion as in in vivo spermatogenesis, genomic DNA was extracted from iPSC and SSCLC,
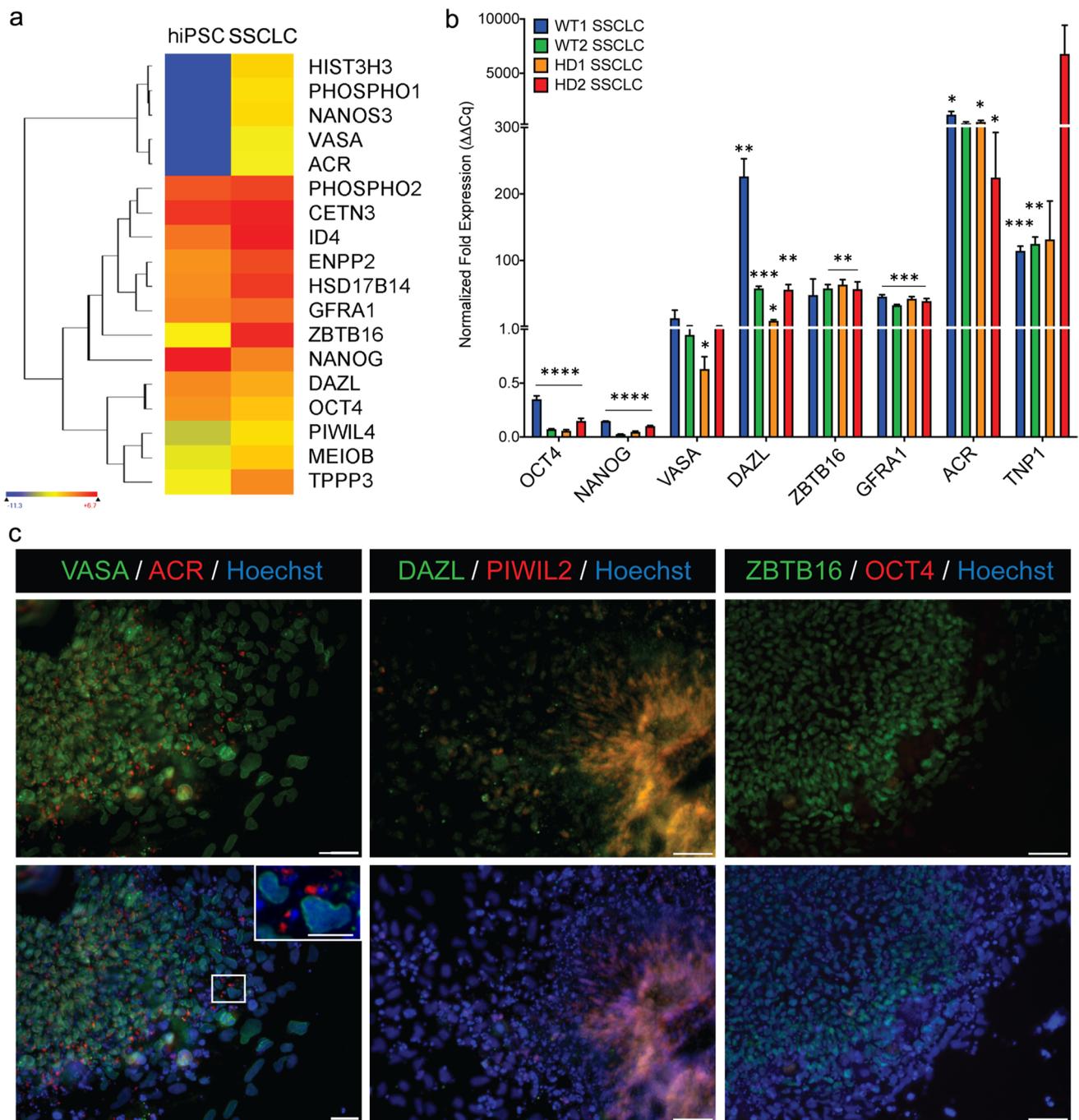
Fig. 1 Differentiation confirmation of hiPSC to SSCLC. **A** Raw read counts (RPKM) of hiPSC compared to day 10 sample showing higher expression of some of the spermatogenesis-specific transcripts: *VASA*, *ACR*, *GFRA1*, and *ZBTB16*. **B** Quantitative real-time polymerase chain reaction (qRT-PCR) showing downregulation of pluripotency markers expressions such as *OCT4* and *NANOG* for all cell lines (for all $p < 0.00005$), and upregulation of spermatogonial stem cells like cells markers such as *VASA*, *DAZL*, *ZBTB16*, *GFRA1*, *ACR*, and *TNP1* ($n = 3$). **C** Representative immunofluorescence showing expression of VASA, ACR, DAZL, PIWIL2, and ZBTB16 but no OCT4 (scale $= 100$ μm, insert scale $= 50$ μm)

and the size of TNRs was analyzed. Both wild-type cells (WT1 and WT2) did not change CAG size before and after the differentiation process (Fig. 3a, b, Supplementary Information Fig. 2). However, after the differentiation process,

HD2 showed significant expansion in CAG repeat sizes in SSCLC compared to iPSC (Fig. 3d; Fig. 4b, d; Supplementary Information Fig. 3). Although HD1 showed a slight increase in CAG repeat size in large allele, the change was
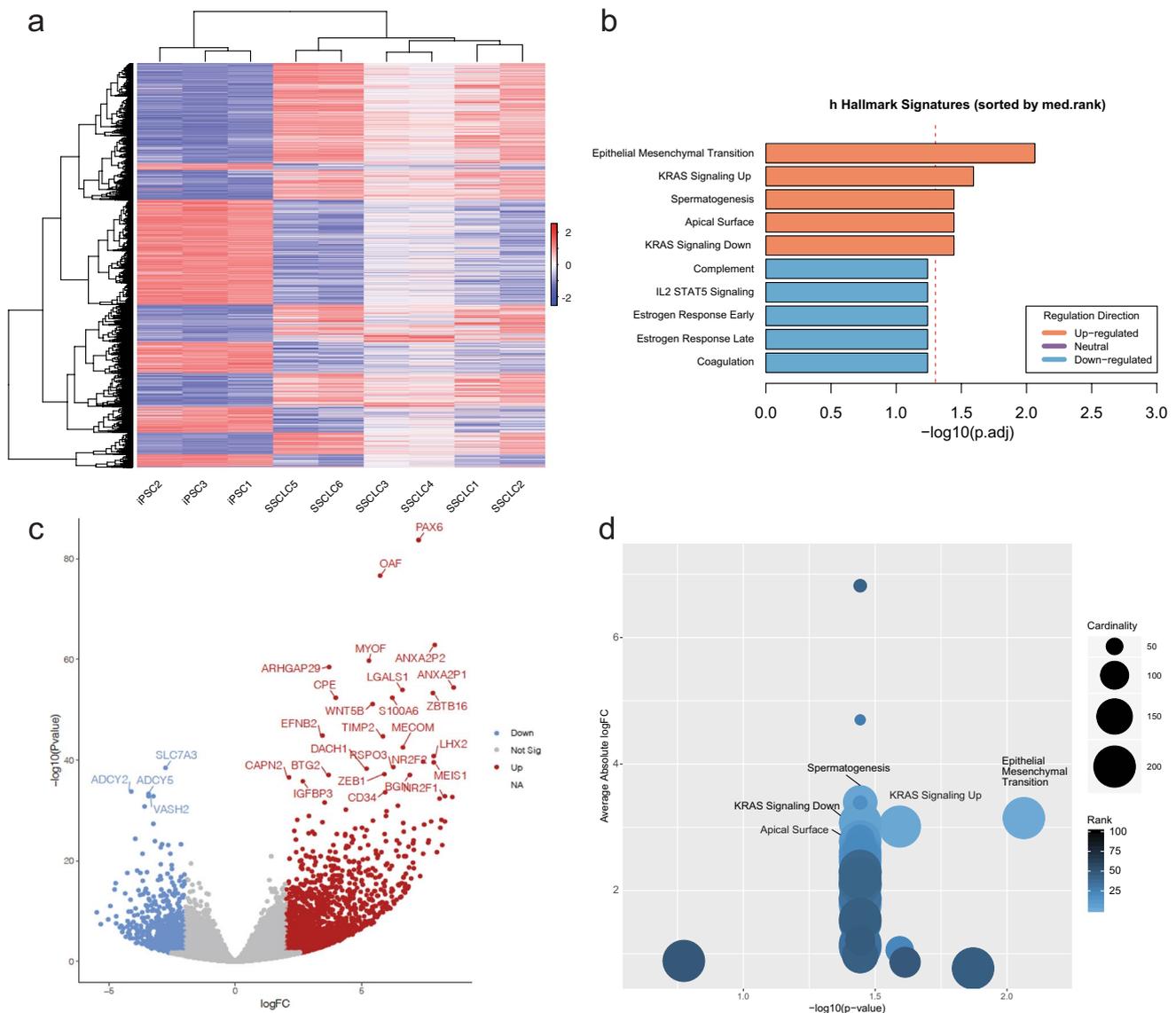
Fig. 2 RNA-seq analysis of differentially expressed genes of iPSC and SSCLC. **A** Hierarchical clustering and the sample-to-sample cluster of differentially expressed genes showing clustering of iPSCs and SSCLCs. **B** Gene set enrichment analysis of both upregulated and downregulated genes showing spermatogenesis as one of the top enriched hallmark signatures (22/135, $p = 0.036$) while downregula-tion of estrogen response hallmark signatures (49/200, $p = 0.054$). **C** Volcano plot of differentially expressed genes showing induced expression of a spermatogenic gene such as *PAX6*, *TIMP2*, *MECOM*, and *ZBTB16*. **D** Gene set enrichment analysis (GSEA) of induced genes after the differentiation shows spermatogenesis as one of the hallmark signatures enriched after the differentiation

not significant with both curve fit analysis (Fig. 4a, $\bar{x} = 38.88$ Q to $\bar{x} = 42.17$ Q, $p = 0.132$) and expansion index analysis (Fig. 4c, p = 0.850). HD2 showed a significant increase in CAG repeat size after the differentiation (Fig. 4b, d). With curve-fit analysis, HD2 showed significant increase of large allele (Fig. 4b, from $\bar{x} = 238$ Q to $\bar{x} = 250.8$ Q, $p = 0.00838$). Also, expansion index increased significantly for small allele ($p = 0.0102$) and larger allele ($p = 0.00193$).

Then, we evaluated whether treating cells with chemicals at iPSC state or during SSCLC differentiation can modulate CAG repeat sizes. Both iPSCs and cells undergoing SSCLC differentiation were treated with 200 mM caffeine, 10 pM and 10 nM aspirin, and 10 pM and 10 nM araC for 10 days. Although slight increase of CAG repeat sizes were observed when cells were treated with 200 mM caffeine (Figs. 5 and 6), no statistically significant changes were observed in all cells (Figs. 5 and 6). In HD1, no statistically significant changes were observed with aspirin and araC treatments (Fig. 6, Supplementary Information Fig. 4). HD2 iPSC treated with 10 pM and 10 nM of araC showed a significant decrease in CAG repeat size compared to 200 mM caffeine group (Fig. 6c, p = 0.0187 and $p = 0.0161$ respectively). HD2
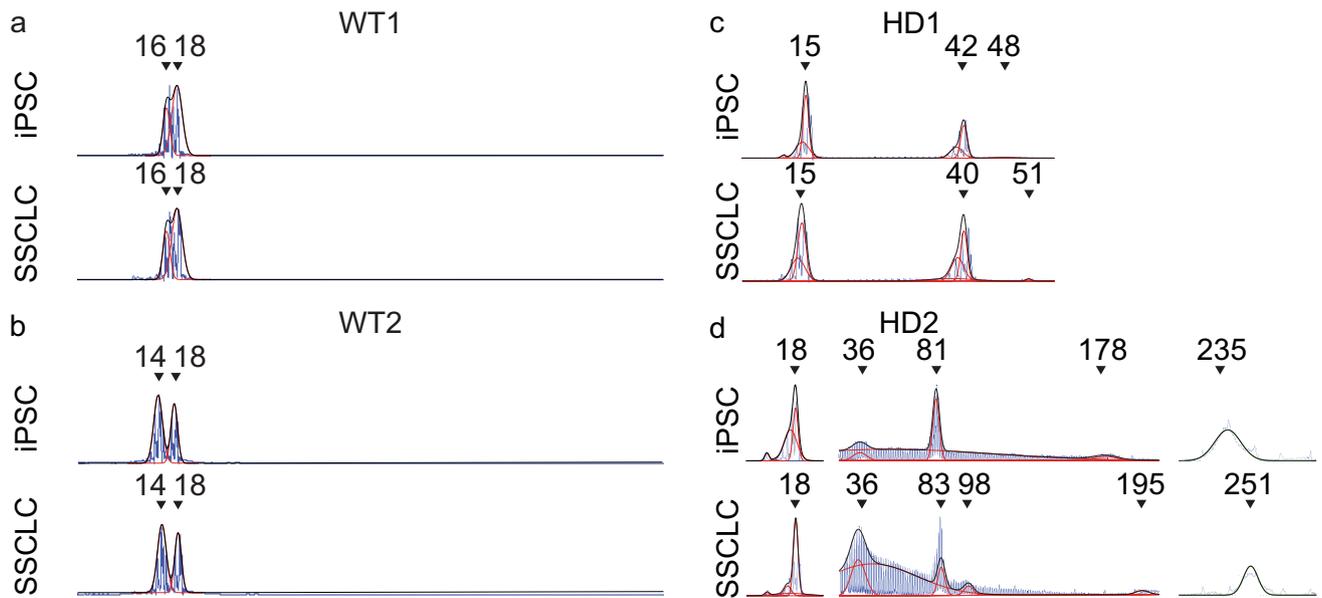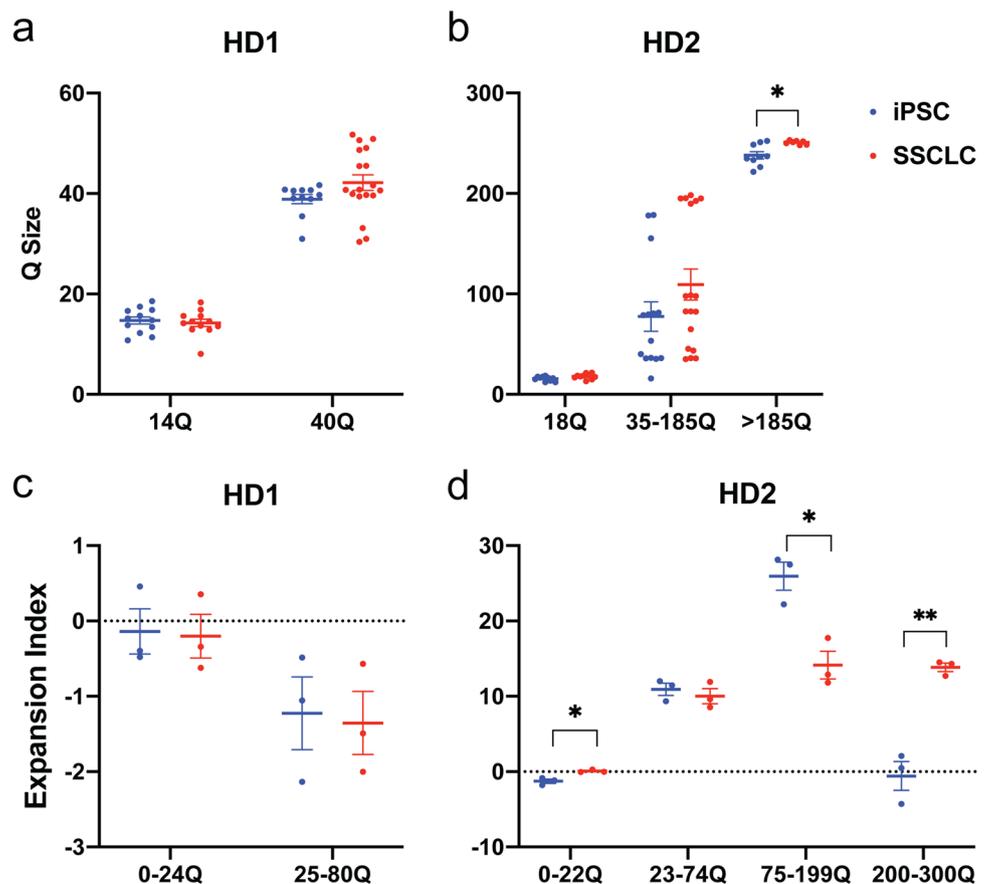
Fig. 3 Representative electropherograms of four cell lines used in this study before (iPSC) and after the differentiation (SSCLC). **A**, **B** Both wild-type cell lines did not change CAG size before and after the differentiation process. **C** HD1 showed a slight increase in the more significant allele of the CAG repeat after the differentiation from 48 to 51 Q, while 15 Q did not change. **D** HD2 showed a general increase in CAG repeat sizes for intermediate alleles and larger alleles, while the smallest allele did not change. The intermediate allele size changed from 81 to 83 Q and 98 Q and 178 Q to 195 Q. The larger allele changed from 235 to 251 Q (Y-axis represents relative fluorescence intensity, RFU)

Fig. 4 TNR changes during differentiation. **A** Curve fit analysis of HD1 before and after the differentiation showing no significant change for both small allele ($\bar{x}=14.72$ Q to $\bar{x}=14.23$ Q, $p=0.64$) and larger allele ($\bar{x}=38.88$ Q to $\bar{x}=42.17$ Q, $p=0.132$). **B** Curve fit analysis of HD2 showed increase of 2 Q of the 18 Q allele from $\bar{x}=15.62$ Q to $\bar{x}=17.88$ Q, which was not statistically significant ($p=0.0534$). The intermediate allele increased 31.85 Q from $\bar{x}=77.54$ Q to $\bar{x}=109.4$ Q but was not significant ($p=0.154$). The largest allele significantly increased after the differentiation from $\bar{x}=238$ Q to $\bar{x}=250.8$ Q ($p=0.00838$). **C** The expansion index showed no changes of small allele ($p=0.885$) and large allele ($p=0.850$) for HD1 after the differentiation. **D** The expansion index showed significant increase of small allele ($p=0.0102$), decrease in larger intermediate allele ($p=0.0107$), and the increase in the larger allele ($p=0.00193$)
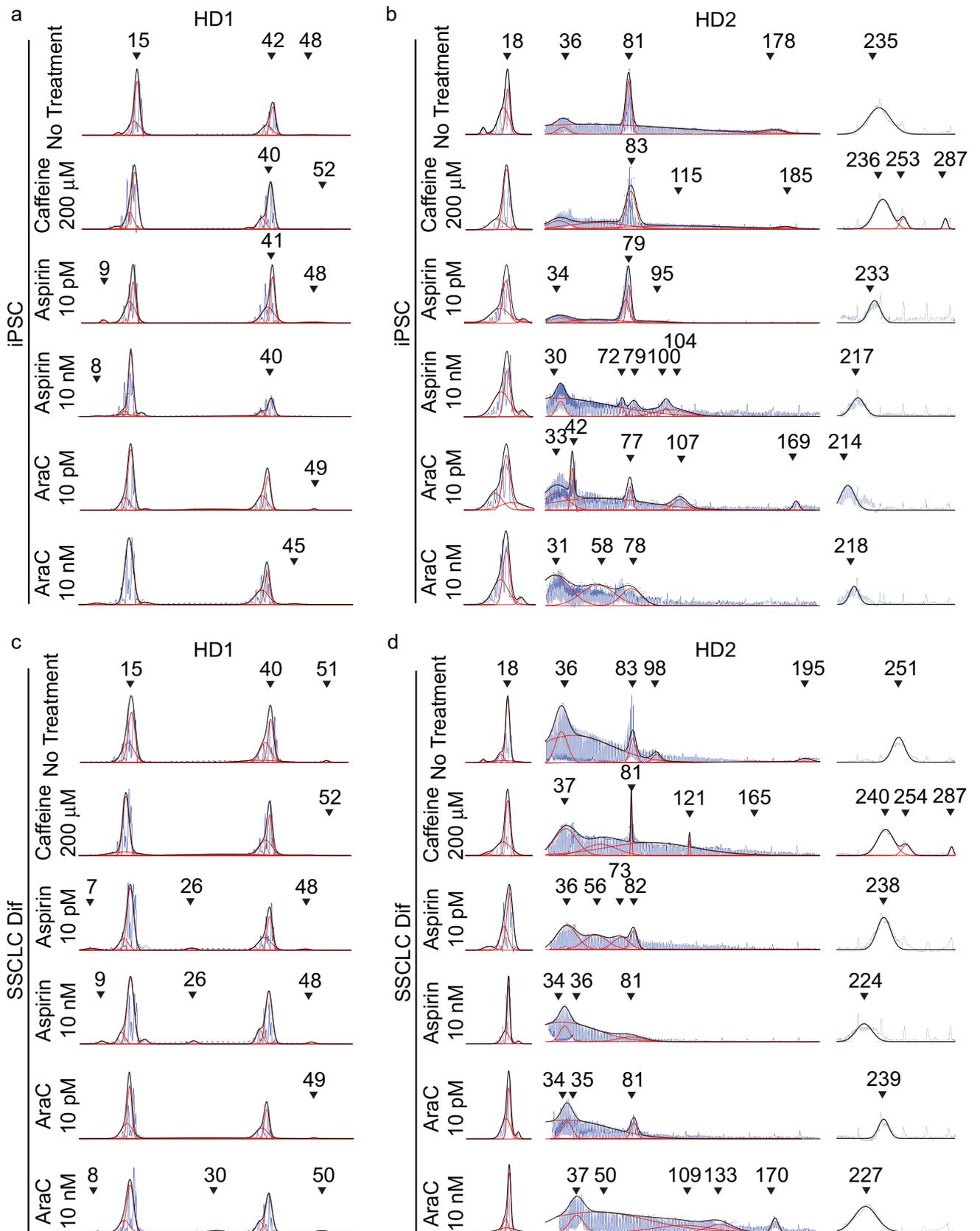
◄**Fig. 5** Representative electropherograms of iPSC and SSCLC chemical treated cells. **A** Representative electropherograms showing the impact of caffeine, aspirin, and araC treatment in iPSC. Caffeine slightly increased 48 Q to 52 Q while aspirin 10 nM treatment showed a decrease in 42 Q to 40 Q, and no 48 Q allele was observed. AraC treatment showed a slight decrease in 48 Q to 45 Q. **B** Chemical treatment of HD2 at iPSC state showed a slight increase of larger allele with caffeine treatment from 235 to 236 Q, 253 Q, and 287 Q. Aspirin treatment showed a general decrease in repeat size. In contrast, the higher concentration of aspirin showed a more significant decrease in CAG repeat size. AraC treatment showed a decrease in CAG repeat size, especially for the intermediate allele. At the same time, the large allele also showed a decrease in the repeat size from 235 to 218 Q. **C** Chemical treatment during SSCLC differentiation of HD1 showed a general decrease with aspirin and araC treatment. At the same time, a slight increase in caffeine was observed. **D** HD2 treated with caffeine showed a general increase in CAG repeat size, while aspirin and araC showed a general decrease in CAG repeat size (*Y*-axis represents relative fluorescence intensity, RFU)

SSCLC treated with 10 pM and 10 nM of aspirin showed significant decrease in intermediate allele (Fig. 6d, $\bar{x} = 109.4$ Q to $\bar{x} = 60.74$ Q and 61.39 Q, $p = 0.015$ and 0.0273 respectively). Both aspirin and araC significantly stabilized larger allele compared to no treatment group. The aspirin treatment showed significant stabilization from $\bar{x} = 250.8$ Q to $\bar{x} = 234.2$ Q for 10 pM and $\bar{x} = 236.8$ Q for 10 nM ($p < 0.001$ and 0.00454 respectively). SSCLC treated with 10 nM of araC showed significant stabilization of larger allele from $\bar{x} = 250.8$ Q to $\bar{x} = 212.5$ Q ($p < 0.001$) but not for 10 pM treatments ($\bar{x} = 244.7$ Q, $p = 0.0736$). Based on expansion index analysis, small allele from all HD2 SSCLC showed significantly increased expansion index compared to iPSC (Supplementary Information Fig. 4b). AraC 10 nM treatment of iPSC showed significantly lower expansion index compared to caffeine and aspirin (10 pM and 10 nM) treatment group (Supplementary Information Fig. 4b). Compared to all treatment groups before the differentiation, SSCLC control group showed increased expansion index after the differentiation for the larger allele (Supplementary Information Fig. 4b, 200–300 Q, purple circles). With chemical treatments, both aspirin and araC treatment showed significant decrease in expansion index compared to no treatment control in SSCLC for all concentration groups (Supplementary Information Fig. 4b). Also, higher concentration of araC (10 nM) showed significant decrease in expansion index compared to lower 10 pM concentration of araC (Supplementary Information Fig. 4b, $p < 0.001$). The larger intermediate alleles (75–199 Q) showed significant decrease in expansion index when SSCLC no treatment group was compared to both control and caffeine treatment group (Supplementary Information Fig. 4b). Both aspirin concentration of 10 pM and 10 nM showed reduced expansion index compared to the control (Supplementary Information Fig. 4b). SSCLC treated with 10 nM araC increased expansion index compared to the control (Supplementary

Information Fig. 4b) and both concentration of aspirin of 10 pM and 10 nM (Supplementary Information Fig. 4b).

To investigate mechanisms involved in the changes in CAG repeat size during the differentiation process and with the chemical treatment, the gene expression of a panel of genes involved in the DNA damage response (DDR) pathway was analyzed utilizing RT-PCR (Supplementary Information Figs. 5, 6). At iPSC state, both WT and HD showed similar DDR gene expression (Supplementary Information Fig. 5a). In SSCLC, HD generally showed induced expression of DDR genes (Supplementary Information Figs. 5b, 6a). Since the presence of expanded CAG repeat track in *HTT* might affect the DDR gene expression in SSCLC, the expression of DDR genes were separately compared in WT and HD cells (Supplementary Information Fig. 5c, d). In both WT and HD, generally DDR gene expressions were higher in SSCLC (Supplementary Information Fig. 5c, d). However, *DDB2* and *PCNA* expressions were significantly lower in WT SSCLC than WT iPSC (Supplementary Information Fig. 5c). In normal tissues, the expression of *HTT* is elevated in the brain, skin, and testis [28]. Both WT and HD cells show significant upregulation in *HTT* expression in SSCLC compared to iPSC (Supplementary Information Fig. 7). In SSCLC, treating cells during differentiation with 10 pM and 10 nM of araC significantly decreased *HTT* expression in HD SSCLC compared to WT SSCLC (Supplementary Information Fig. 7). When gene expressions of selected genes involved in DNA damage response (DDR) pathways were compared to no treatment group, WT iPSC showed induced expression of *ATR* and *FAN1* when it was treated with 200 mM caffeine while aspirin and araC treatment suppressed *DDB2* expression (Fig. 7a). WT iPSC showed induced expression of *DDB2* with caffeine, aspirin, and araC treatment, while aspirin suppressed *OGG1* expression and araC induced *PCNA* expression (Fig. 7b). HD1 iPSC showed suppressed DDR gene expression with 200 mM caffeine treatment (Fig. 7c). Treated with aspirin and araC, HD1 iPSC showed suppressed *OGG1* expression, and *ATR* expression was suppressed with aspirin treatment, while *ATM* and *MSH3* were suppressed with araC treatment (Fig. 7c). In HD1 SSCLC, DDR genes showed generally suppressed expression with aspirin and araC treatments (Fig. 7d). HD2 iPSC responded differently from WT iPSC and HD1 iPSC with caffeine, aspirin, and araC treatment (Fig. 7e). In HD2 SSCLC, except *OGG1*, most DDR genes did not change expression when treated with caffeine, and aspirin and araC suppressed *OGG1* and *FAN1* expression, and araC suppressed *ERCC5* while only 10 nM of araC suppressed *PCNA* and *RAD51* expression (Fig. 7f).

In human spermatogenesis, DNA methylation remodeling occurs continuously during the developmental process [29], and oxidative DNA damage, environmental exposure, and nutrition can change DNA methylation of sperm [18,
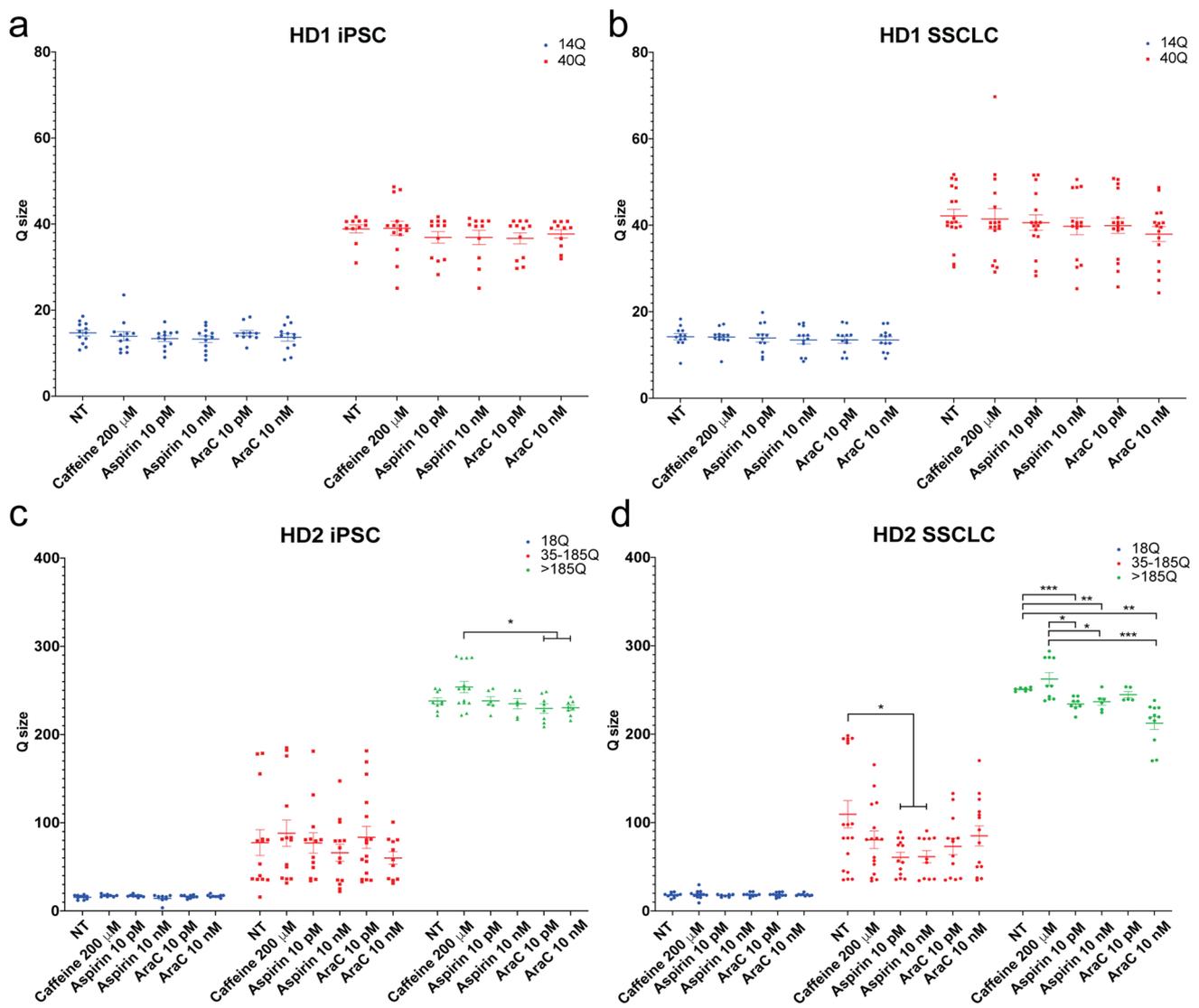
Fig. 6 Curve-fit data analysis of repeat sizes. **A** Chemical treatment of HD1 iPSC did not significantly change the CAG repeat size. **B** Although aspirin and araC treatment of HD1 SSCLC generally reduced the CAG repeat size of the larger allele, the changes were not statistically significant. **C** Chemical treatment of HD2 iPSC did not show a significant reduction in CAG repeat size with aspirin and araC treatment. However, when compared to caffeine treated cells, both concentrations of araC 10 pM and 10 nM treated cells showed significantly reduced size of CAG repeat size ($p = 0.0187$ and $p = 0.0161$, respectively). **D** Both concentrations of aspirin treatment (10 pM and 10 nM) during SSCLC differentiation stabilized intermediate allele size when compared to control group ($\bar{x} = 109.4$ Q to $\bar{x} = 60.74$

Q and 61.39 Q, $p = 0.015$ and 0.0273 respectively). Both aspirin and araC significantly stabilized larger alleles compared to the no treatment group. The aspirin treatment showed significant stabilization from $\bar{x} = 250.8$ Q to $\bar{x} = 234.2$ Q or 10 pM and $\bar{x} = 236.8$ Q for 10 nM ($p < 0.001$ and $= 0.00454$ respectively). SSCLC treated with 10 nM of araC showed significant stabilization of larger allele from $\bar{x} = 250.8$ Q to $\bar{x} = 212.5$ Q ($p < 0.001$) but not for 10 pM treatments ($\bar{x} = 244.7$ Q, $p = 0.0736$). For all groups, the mean and the standard error with individual data points were plotted. Multiple unpaired t-tests were conducted to calculate the statistical significance. $n = 3$, *$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$, ****$p < 0.00005$

30–32]. To further investigate how the chemicals might have regulated those genes, methylation profiles of five genes that either showed a similar trend as TNR expansion suppression or have previously been reported to be associated with TNR expansion were analyzed by the pyrosequencing. *APEX1*, *BRCA1*, and *DDB2* were selected based on the gene expression that correlates with CAG repeat

stabilization, and *OGG1* and *FAN1* have been associated with TNR expansion [33–36]. In both iPSC and SSCLC, WT showed significant decreased in methylation with 10 nM aspirin treatment (Fig. 8a). Compared to all iPSC and WT SSCLC, HD SSCLC showed a significant increase in *APEX1* methylation (Fig. 8a). Compared to HD iPSC no treatment group and WT SSCLC no treatment group,
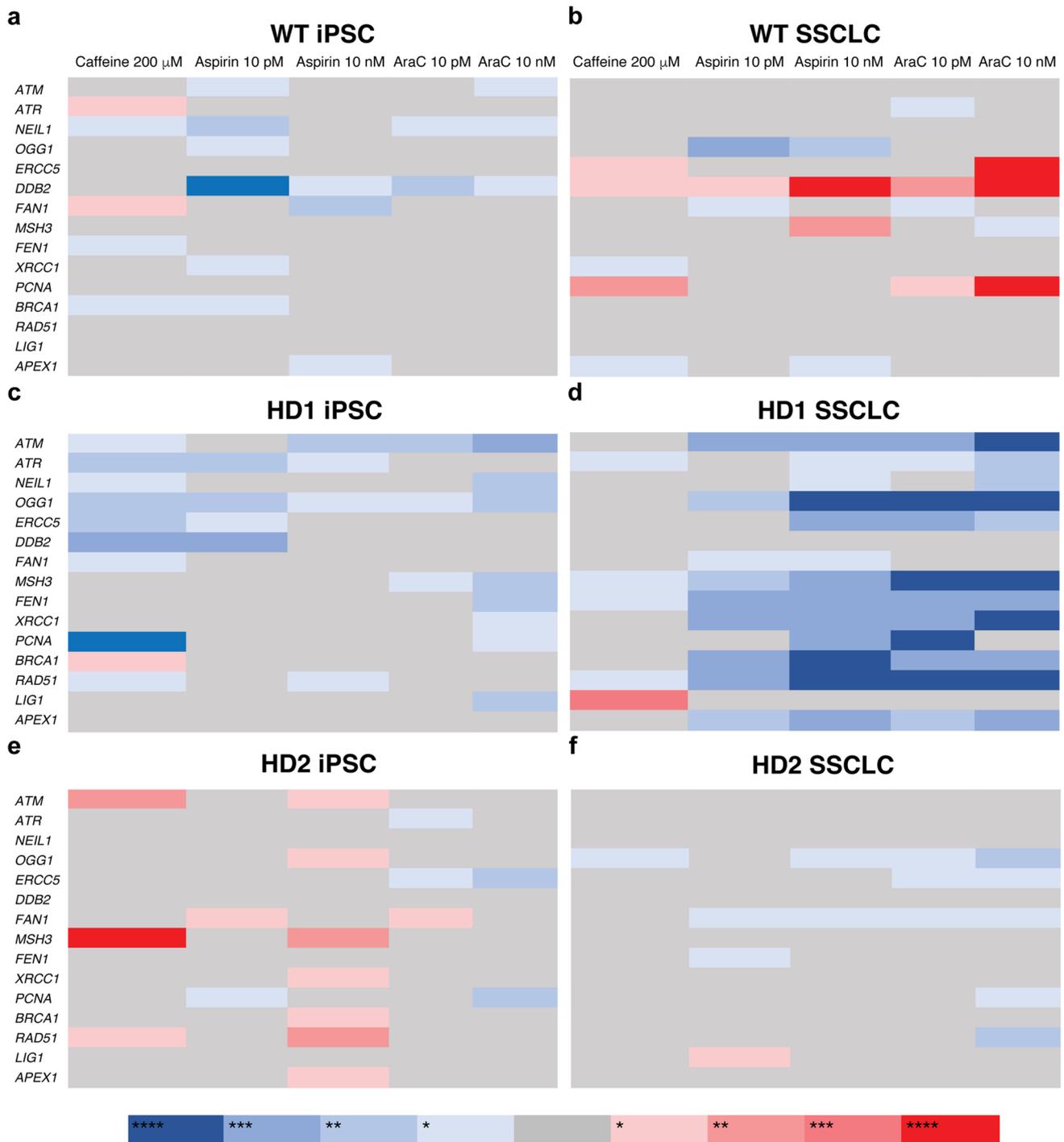
Fig. 7 A summary plot of DDR gene expression relative to control. **a** Caffeine treatment induced expression of *ATR* (*p*=0.0477) and *FAN1* (*p*=0.00687) while aspirin and araC treatment suppressed *DDB2* expression. **b** Caffeine, aspirin, and araC treatment induced expression of *DDB2* while aspirin suppressed *OGG1* expression, and araC induced *PCNA* expression. **c** In general, caffeine suppresses DDR gene expression except for *BRCA1* (*p*=0.0052), and aspirin and araC suppressed *OGG1* expression. *ATR* expression was suppressed with aspirin treatment, while *ATM* and *MSH3* were suppressed with araC treatment. **d** Both aspirin and araC treatment suppressed most of DDR genes while caffeine induced expression of *LIG1* in SSCLC

(*p*<0.001). **e** Caffeine and 10 nM of aspirin treatment induced *ATM*, *MSH3*, and *RAD51* in HD2 iPSC, while *ATR*, *ERCC5*, and *PCNA* expressions were suppressed in the cells treated with araC. Also, 10 nM of aspirin induced most of the DDR genes. **f** Except for *OGG1*, most DDR genes did not change expression when treated with caffeine. SSCLC treated with aspirin and araC suppressed *OGG1* and *FAN1* expression, and araC suppressed *ERCC5* while only 10 nM of araC suppressed *PCNA* and *RAD51* expression. Multiple unpaired t-tßests were conducted to calculate the statistical significance (*n*=6 to 12 based on the sample availability). **p*<0.05, ***p*<0.005, ****p*<0.0005, *****p*<0.00005
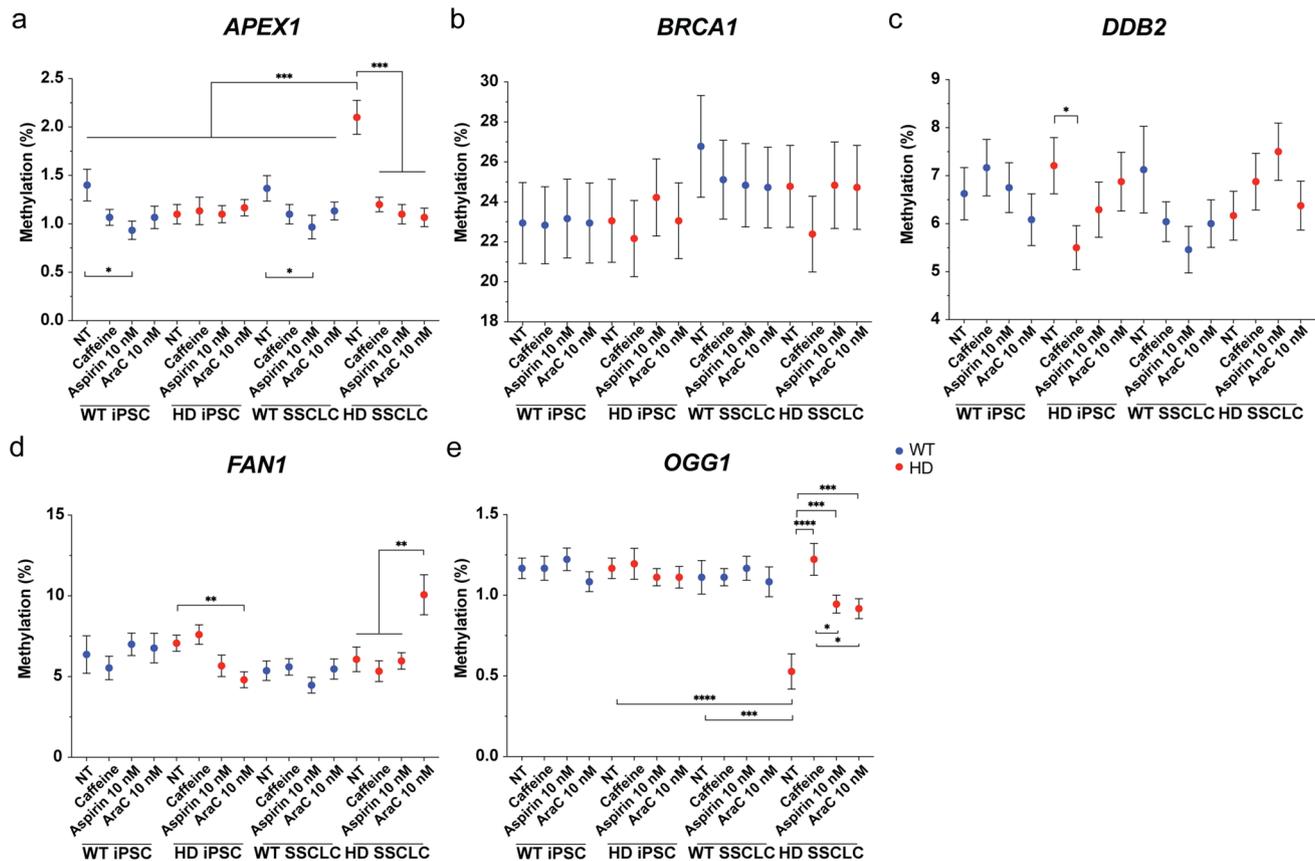
Fig. 8 Differential methylation profile of selected genes involved in DDR. **A** In both iPSC and SSCLC, WT showed a significant decreased in methylation with 10 nM aspirin treatment ($p = 0.0165$ and $p = 0.0295$). Compared to all iPSC and WT SSCLC, HD SSCLC showed a significant increase in *APEX1* methylation. Compared to the HD iPSC no treatment group and WT SSCLC no treatment group, HD SSCLC showed a significant increase in methylation ($p < 0.001$ and $p = 0.00142$). Also, treatment with caffeine, aspirin, and araC all significantly decrease the methylation ($p < 0.001$, $p < 0.001$, and $p < 0.001$ respectively). **B** Methylation of *BRCA1* did not change after SSCLC differentiation and after the chemical treatment. **C** Methylation of *DDB2* was significantly decreased in HD iPSC treated with caffeine than the no-treatment group ($p = 0.0264$).

D HD iPSC treated with araC showed a significant decrease in methylation compared to the no treatment group ($p = 0.0020$). Treating HD SSCLC with araC significantly increased methylation of *FAN1* compared to all other groups especially compared to no treatment group ($p = 0.00802$). **E** Compared to HD iPSC no treatment, HD SSCLC no treatment showed a significant decrease in methylation ($p < 0.001$). Also, HD SSCLC showed significantly lower methylation in *OGG1* compared to WT SSCLC ($p < 0.001$). Also, HD SSCLC no treatment showed lower methylation compared to caffeine, aspirin, and araC treatment group ($p < 0.001$, $p = 0.00109$, and $p = 0.00272$). HD SSCLC treated with caffeine showed significantly higher methylation than aspirin and araC treated groups ($p = 0.0165$ and $p = 0.0104$)

HD SSCLC showed a significant increase in methylation of *APEX1* (Fig. 8a). Also, treatment with caffeine, aspirin, and araC all significantly decrease the methylation (Fig. 8a). When HD1 and HD2 were analyzed separately, HD1 and HD2 showed a significant increase in methylation at *APEX1* (Supplementary Information Fig. 8). Both *BRACA1* and *DDB2* did not show a difference in methylation between cell type and chemical treatments (Fig. 8b, c). However, *DDB2* showed a decreased methylation in HD iPSC with 200 mM caffeine treatment compared to the no treatment group (Fig. 8c). *FAN1* showed no differential methylation in WT iPSC and WT SSCLC (Fig. 8d). HD iPSC treated with 10 nM of araC showed a significant

decrease in methylation compared to the no treatment group (Fig. 8d). However, in HD SSCLC, SSCLC treated with 10 nM of araC showed a significant increase in methylation of *FAN1* compared to no treatment, 200 mM caffeine treated, and 10 nM aspirin-treated cells (Fig. 8d). Both HD1 and HD2 SSCLC showed a similar trend as they both showed increased *FAN1* methylation, but HD2 SSCLC showed a more significant change than HD1 (Supplementary Information Fig. 8). Both WT and HD showed no change in *OGG1* methylation in iPSC (Fig. 8e). However, HD SSCLC showed a significant decrease in methylation after the differentiation (Fig. 8e), and both HD1 and HD2 showed similar methylation changes after the differentiation (Supplementary Information Fig. 8). Also, HD

SSCLC showed significantly lower methylation in *OGG1* than WT SSCLC (Fig. 8e). HD SSCLC with no treatment showed lower methylation compared to caffeine, aspirin, and araC treatment group (Fig. 8e). HD SSCLC treated with caffeine showed significantly higher methylation than aspirin and araC treated groups (Fig. 8e).

Discussion

In Huntington's disease, CAG repeat expansion occurs in male germlines [8–11]. Due to the limitation of availability and ethical concerns, a robust study to dissect the mechanisms involved in TNR instability in human spermatogenesis has been challenging. Understanding the mechanisms of CAG repeat expansion and developing drugs or therapies that can stabilize or even induce contraction of CAG repeat size in male germ cells will provide an opportunity to eradicate the devastating effect of genetic anticipation in trinucleotide repeat expansion disorders (TREDs). Mouse and other model organisms have been instrumental in discovering processes involved in TNR expansion. Still, differences in kinetic, biological, and developmental differences in the spermatogenesis process [37], limited availability, and ethical usage of human tissues leave room for developing a better system to study the paternal expansion of CAG repeat in spermatogenesis. To overcome the current limitations, we have developed an in vitro stem cell model to replicate TNR instability in spermatogenesis.

We have reported CAG repeat expansion in HD monkey lymphocytes and sperm [5] and monkey spermatogenic cells derived from embryonic stem cells in vitro [24]. The current study expands our previous study by utilizing HD patient-derived induced pluripotent stem cell (iPSC) to replicate TNR instability observed in sperm of HD patients that can provide a platform to study TNR instability and the underlying mechanism of genetic anticipation in human germ cells. As we have demonstrated in monkey ESC [24], we confirmed the expansion of CAG repeat in human iPSC-derived SSCLC in HD cells (Fig. 3, Supplementary Information Fig. 9). Although HD1 with the initial CAG repeat size of 14.72 Q and 38.88 Q showed a moderate but not statistically significant increase in CAG repeat size after the differentiation, HD2 with the initial CAG repeat size of 15.62 Q and 238 Q showed a statistically significant expansion after the differentiation (Fig. 4). These findings were consistent with previous studies in rodents [38–40], monkey [5, 24], and human [41–44], where larger CAG repeats are more unstable and prone for expansion.

In this study, we expanded our previous study on monkey SSCLC by treating differentiating SSCLC with two known TNR stabilizing agents, aspirin and araC [27].

While we showed HD2 iPSC was compared to SSCLC, both aspirin and araC prevented further expansion of CAG repeat size in HD2 SSCLC (Fig. 6d, Supplementary Information Fig. 4b). The moderate changes observed in HD1 might be also important because loss and gain of a single CAG repeat could delay or expedite onset by an average of 3 years in 40–50 Q range [45].

To investigate possible mechanisms involved in CAG repeat size change observed in this study, we evaluated the gene expressions of a selected panel of DNA damage response (DDR) genes by quantitative real-time PCR. As a preliminary study, changes in gene expressions in the selected panel of genes were analyzed to infer the involvement of a specific DDR pathway although further analysis is needed to relate those changes to steady-state levels of protein. In general, SSCLC showed induced expression of DDR genes in both WT and HD (Supplementary Information Fig. 5c, d). *OGG1*, 7,8-dihydro-8-oxoguanine-DNA glycosylase, initiates BER by recognizing and removing 8-oxoG from the opposite of C in DNA [46], which has been associated with the CAG expansion [34, 47]. The *OGG1* knockout (*OGG1*$^{-/-}$) R6/1 mice showed suppression of age-dependent CAG repeat expansion [34]. Since increased CAG repeat size does not necessarily induce 8-oxo-G lesions, the authors of the study suggested that the 8-oxo-G lesions are favored within CAG sites, or *OGG1* binds to CAG repeat DNA tertiary structures [34]. Another gene involved in the BER pathway, *APEX1*, was also induced in HD SSCLC (Supplementary Information Fig. 5b). APEX1, apurinic/apyrimidinic endonuclease 1, has been reported to incise the 5′end of an abasic site in DNA and TNR hairpin loop [48, 49]. Also, *APEX1* can stimulate *OGG1, FEN1*, and *LIG1*, which can initiate BER [50–53]. Recent studies have associated oxidative stress with the expansion of TNR [34, 54–57].

Two genes, *FAN1* and *MSH3*, have recently been the focus of the HD research due to recent genome-wide association studies (GWAS) identifying those two genes as genetic modifiers of HD [58]. Increased expression of both *FAN1* and *MSH3* in SSCLC compared to iPSC will be interesting to investigate further in future studies. Although *FAN1* is involved in suppressing somatic repeat expansion, in germ cells, *FAN1* knockout has no effect on repeat expansion in fragile X disorders [59]. *MSH3* induces repeat instability in somatic cells in both HD and DM models [60–63], and in the germline, *Msh3* knockout suppressed expansion, and in DM, even promoted contraction [64, 65]. Also, *MSH3* expression levels are reported to be rate-limiting in the expansion mechanism [64, 65], but the contribution of *MSH3* expression level on repeat expansion in human spermatogenesis remains to be explored.

HD SSCLC specific induced expression of *NEIL1*, *ERCC5*, *XRCC1*, and *BRCA1* might be involved in further expansion of the already expanded CAG repeat. Of those

genes, *Neil1* has been identified as a genetic modifier of the intergenerational repeat instability [66]. Recent studies have suggested crosstalk between BER and MMR [61, 67]. Although MMR plays a more critical role in somatic CAG repeat expansion in mouse models, the interaction between factors involved in BER or other DDR pathways is not well understood in human germline cells. These targets can be modified with siRNA or CRISPR-Cas9 in our in vitro culture model to investigate their function in CAG repeat expansion during human spermatogenesis.

HTT expression was induced in SSCLC, but *HTT* expression did not significantly change by neither aspirin nor araC treatments (Supplementary Information Fig. 7). Therefore, it is unlikely that TNR stabilization by aspirin or araC has resulted from the suppressed expression of *HTT*.

Aspirin was suspected that the induction of expression of MMR proteins, such as MLH1, MSH2, MSH6, and PMS2, by aspirin facilitates apoptosis [68]. The concentration of aspirin used in this study is well below the physiological concentration found in patients, which might not have been enough to induce dosage-response in this study (20 mM) [69]. Except for HD2 iPSC and WT SSCLC, aspirin suppressed most DDR gene expression (Fig. 7). In all SSCLC, the expression of *OGG1* and *FAN1* was significantly suppressed with aspirin treatment (Fig. 7b, d, e). Since *OGG1* is mainly involved in BER by removing 8-oxoG, the antioxidant activity of aspirin might explain the stabilization effect of aspirin on TNR. In human spermatozoa, the truncated BER pathway is functional, containing only OGG1 protein [70]. Reactive oxygen species (ROS) is necessary for basic sperm function [71, 72]. Guanine is the most susceptible to oxidation, and ROS endogenously generates 8-oxoG. The CAG sites could be more prone to 8-oxoG lesions, and CAG repeats tertiary structures might bind better with OGG1 [34]. The oxidation-excision-expansion cycle that is present and accumulates in the brain might also be presented in spermatogenesis.

Except in WT SSCLC, araC suppressed most of DDR gene expression as well although the mechanism of action is different from aspirin (Fig. 7). Cytosine arabinoside (araC, cytarabine, 1-b-D-arabinofuranosylcytosine) is a cytosine analog nucleoside that competes with deoxycytidine triphosphate (dCTP) and blocks the DNA replication process [73]. Also, araC inhibits DNA polymerase δ, the primary polymerase involved in the DNA mismatch repair [74], while allowing translesion DNA synthesis (TLS) polymerase η [75]. AraC inhibits the DNA synthesis stage of MMR [27], and it is possible that araC [76] treatment have changed the preference between replication DNA polymerases and translesion DNA polymerase and mediate TNR repair when hairpin or loop structure is formed during DNA replication. Like aspirin, the suppression of *FAN1* in SSCLC might suggest that oxidative stress response is more involved in

CAG repeat expansion in spermatogenic cells rather than MMR identified as genetic modifiers involved in CAG repeat expansion in fibroblast and blood samples.

HD SSCLC showed a significant increase in *APEX1* methylation after the differentiation, and HD2 showed a more significant increase than HD1 (Fig. 8a, Supplementary Information Fig. 8). In contrast, *OGG1* showed a significant decrease in methylation in HD after the differentiation (Fig. 8d). In BER, APEX1 displaces OGG1 from apurinic/apyrimidinic (AP) sites generated by the OGG1 glycosylase activity allowing the base excision repair to proceed [77], which suggest induced activity of OGG1 in SSCLC cause induced BER activity, and restoring methylation back to the WT range of both *APEX1* and *OGG1* by aspirin and araC treatments could have helped to suppressed TNR expansion in SSCLC. The methylation of *FAN1* was only significantly induced in HD SSCLC when treated with araC (Fig. 8d), and only HD2 SSCLC showed induced *FAN1* methylation with araC treatment (Supplementary Information Fig. 8), which correlates to gene expression profile (Fig. 7f). Although the involvement of FAN1 in somatic repeat expansion is well-documented, FAN1 did not show any effect on repeat expansion in germ cells in a fragile X mouse model [59]. However, in fragile X, germline expansion occurs mainly through the female lineage [3], and *Fan1* has been reported to be involved in non-canonical mechanisms to control somatic CAG instability in the mice [78]. Therefore, genomic context, cell-type, and human-specific functions of *FAN1* involvement in human germline expansion need further investigations.

Our data suggest that in vitro derived SSCLC can replicate dynamic CAG repeat insatiability and genetic anticipation. Studies have identified genetic modifiers that might be involved in germline TNR instability, including *NEIL1* [66], *MSH2* [79], *MSH3* [64, 65], *MSH6* [64], *FEN1* [80], *DNMT1* [81], and *CBP* [82]. However, none of the studies have utilized the human system to assess the impact of those genetic modifiers in germline TNR instability. The majority of polyglutamine protein products function in DNA repair, and the growing number of expanded repeat diseases suggest shared genetic modifiers [83–86]. Many TREDS shares similar characteristics such as pathogenic repeat threshold, negative correlation between age of onset and repeat size, somatic and germline expansion, and similar genetic modifiers [87]. These facts suggest that underlying mechanisms of CAG expansion might share the same mechanism, and our platform can be used to investigate germline TNR instability in other TREDs.

As a monogenic dominant disorder, HD can be accurately diagnosed through preimplantation genetic testing (PGT). The recent development of sequencing technologies allows accurate diagnosis of monogenic disorders (PGT-M), chromosomal structural rearrangements (PGT-SR), and

aneuploidy (PGT-A) [88]. Also, the improved efficiency and safety of PGT guarantee the prevention of transmission of HD. If the patient does not wish to undergo pre-symptomatic testing, preimplantation exclusion genetic testing can be conducted to prevent HD transmission. Exclusion testing is indirect, and it is based on haplotype genetic markers from the grandparents. However, even the embryo that is inherited haplotypes from the affected grandparent only has a 50% risk of carrying CAG expansion which can lead to serious moral and ethical objections [89]. Moreover, because CAG expansion occurs during spermatogenesis, people who do not have family members with HD can still have children with HD. With the rapid technological advances in genetic testing, it is now possible for people to find out their HTT genotypes, and people with reduced penetrance range (36–39 CAG repeats) and intermediate repeats (27–25 CAG repeats) can take a precautionary approach my suppressing TNR expansion.

Methods

hiPSC culture

To investigate if in vitro spermatogenesis can replicate CAG instability in hiPSCs, two WT hiPSC lines (BJ and ND41658) and two HD iPSCs (ND38547 and ND36999) were used for in vitro spermatogenesis using established protocol [16]. Two HD cell lines carry different CAG repeat sizes: ND38547 carries 44 Q (HD1), and ND36999 carries 180 Q (HD2). BJ was gift from C.A.E. and all other cell lines were purchased through NINDS Stem Cell Repository. hiPSC lines were maintained in mTeSR™ Plus (STEM-CELL Technologies, Canada) following the instruction provided by the manufacturer.

Directed differentiation of hiPSC to spermatogenic cells

hiPSCs were differentiated to advanced spermatogenic cells by following our established protocol [90–93]. Briefly hiPSCs were cultured for 10 days in human spermatogonia stem cells (SSC) medium containing the following: minimum essential medium (MEM) alpha (Invitrogen), 0.2% bovine serum albumin, 5 mg/ml insulin, 10 mg/ml transferrin, 60 mM putrescine, 2mML-glutamine (Invitrogen), 50 mM b-mercaptoethanol, 1 ng/ml human basic FGF (hbFGF; BD Biosciences), 20 ng/ml GDNF (R&D Systems), 30 nM sodium selenite, 2.36 mM palmitic acid, 0.21 mM palmitoleic acid, 0.88 mM stearic acid, 1.02 mM oleic acid, 2.71 mM linoleic acid, 0.43 mM linolenic acid, 10 mM HEPES, and 0.5X penicillin/streptomycin (Invitrogen). The medium was changed every other day.

Chemical exposures during the spermatogenesis process

During the 10-day differentiation process, cells were challenged with different concentrations of chemicals to determine the impact on CAG instability during the spermatogenesis process (Supplementary Information Fig. 10). The concentrations were based on the previous publications demonstrating physiological concentrations (urinary and blood): 10 pM and 10 nM for araC (Sigma) and aspirin (Sigma) [90–93] and 200 mM of caffeine [27]. The chemicals were added when the differentiation of iPSC media was changed.

Immunocytochemistry

Cells were fixed using 4% paraformaldehyde (PFA) for 45 min and permeabilized using 0.2% Triton-X solution for 10 min. Cells were blocked for 1 h at room temperature using 1% normal horse serum and 5% BSA in PBS. After the initial blocking, primary antibodies were added and incubated overnight at 4 °C. Primary antibodies were washed away by washing cells with PBS three times, and secondary antibodies were added for 1 h at room temperature. The vendor and dilution information can be found in Supplementary Table 3. Images were captured by Olympus BX63 and using CellSens imaging software.

RNA-seq data analysis

A total of 100,000 cells were harvested and washed with PBS twice to remove any debris. Cell pellets were snap-frozen and submitted to Yerkes National Primate Research Center Genomics Core for further analysis. Briefly, total RNA was isolated using Qiagen RNeasy Mini kit, and the quality control was conducted on 4200 Bioanalyzer Capillary electrophoresis (Agilent). The total RNA (10 ng) was used for mRNA amplification using Clontech Smarter V4 chemistry following the manufacturer's protocol. Amplified mRNA was fragmented and barcoded using Illumina's Nextera XT kits. Amplified Libraries were validated by Agilent 3200 Tapestation, and quantification was conducted on a Qubit Fluorimeter. The libraries were normalized, pooled, and clustered on Illumina HiSeq3000/4004 Flowcells using the Illumina cBOT. The prepared libraries were sequenced on an Illumina HiSeq3000 system in 101-base single-read reactions. The raw (FASTQ) files were uploaded to the Galaxy web platform, and data analysis was conducted on the public server at usegalaxy.org [94]. The sequence was mapped using Spliced Transcripts Alignment to a Reference (STAR) [95] (v2.6.0b-1) utilizing the latest release of human Ensembl genome assembly (hg38). Read counts were measured using featureCount

[96] (v2.0.1). Differential expression analyses of two cell types were performed using DESeq2 method [97] (Galaxy Version 2.11.40.6 + galaxy1). The resulting *p*-values were adjusted with Benjamini and Hochberg method (FDR) with the threshold set at 0.05. Gene set enrichment (GSE) was conducted using the Ensemble of Genes Set Enrichment Analysis (EGSEA) (Galaxy Version 1.10.0) [98]. Differential expression analysis data and normalized read counts were used to generate graphs in RStudio (1.4.1717) running on R Package (R-4.1.0).

qRT-PCR

Briefly, for quantification of mRNA expression, mRNA was extracted using Trizol® (Invitrogen) followed by DNA digestion using Turbo DNA-free™ Kit. RNA sample (500 ng) was reversed transcribed using iScript™ cDNA Synthesis Kit (Bio-Rad) and quantified using SYBR Green Supermix (Bio-Rad) using primers described in Supplementary Table 2 on CFX96 Real-Time Detection System (Bio-Rad). PCR conditions for SYBR Green assays are as followed: Initial 95 °C for 30 s, followed by 39 cycles of 95 °C for 10 s and 55 °C for 20 s with melt curve analysis.

DNA extraction

DNA from cell pellets were extracted using the DNeasy Extraction Kit (Qiagen), and the concentration and purity of each sample were determined using NanoDrop™ 2000 (Thermo Fisher, Waltham, MA).

Assessment of CAG instability

CAG instability was be analyzed by GeneScan analysis. Day 0 was used to establish the baseline for TNR instability in spermatogenesis for each cell line. TNR instability was assessed by PCR as previously described [8]. Briefly, DNA was extracted from cell pellets by adding lysis buffer (0.5% Tween-20, 0.1 mg proteinase K, 1X PCR buffer (Takara)). The reaction was incubated at 56 °C for 45 min followed by 95 °C for 10 min. PCR reaction was carried out using 50 ng of DNA. Two primers, forward Hu4F, 5′-FAM-atg-gcgaccctggaaaagctgatgaa, and reverse HD5R, 5′-cggctgag-gcagcagcggctgt, which flanks the CAG repeat tract, were used to amplify the CAG repeat region. PCR conditions as followed: 98 °C for 5 min, followed by 40 cycles of 96 °C for 5 min, 67 °C for 45 s, and 72 °C for 1.5 min, and 72 °C for 10 min. Each reaction contained $MgCl_2$ (2 nM), 1% of Betain, and 0.05 U Taq polymerase (Takara). PCR products were run on 2% agarose gel to verify the product presence,

and 10 uL of the product was submitted for GeneScan analysis at Emory Integrated Genomics Core. The samples were run on ABI 3130xl Genetic Analyzer (ABI). The precise size of the product was calculated against GeneScan™ 1000 ROX™ dye Size Standard (ABI). The sizing of PCR fragments was conducted using GeneMarker Software Version 2.7.0 (Softgenetics®).

Data analysis using curve fitting method

The data were analyzed using GeneMarker® (SoftGenetics); allele data was imported into MatLab (MathWorks). We employed a previously described method [99], which has been used in multiple publications [100, 101]. Briefly, imported masked data were analyzed using ipf.m function in MatLab optimizing for error (less than 10%) and overall fitness R^2 value (higher than 0.95). Prolonged in vitro culture has been reported to induce TNR contraction [102–104] and expansion [105–109]. Also, in order to minimize the PCR amplification bias toward smaller alleles, and possible DNA synthesis errors known as "PCR stutter," all peaks that cross the threshold were analyzed in this study. The consecutive normal Gaussian distribution was used to fit the data maximizing for the coverage area of the raw data set. Later, electrograms and curve-fit results were combined using Adobe Photoshop and Adobe Illustrator (Adobe). Expansion index was calculated by modifying instability index [110] following the equation: $\Sigma\left(\left(\frac{\text{peak height}}{\Sigma\text{peak heights}}\right)(\Delta TNR \text{ from the reference allele})\right)$. This represents the instability of a sample and its tendency toward expansions (i.e., positive values) or contractions (i.e., negative values). Expansion index close to 0 indicates low instability, while values between 1 and 5 show medium instability and values over 5 show high instability. For each cell type (iPSC and SSCLC), no treatment group was used to define the reference alleles. The numerical changes in TNR sizes are presented in Supplemental Table 1. Expansion index lacks the ability to separate the continuous and periodic expansion of trinucleotide repeats. Therefore, both the curve-fit method and expansion index methods were used in this study. For the linear regression, the Pearson correlation coefficient (R^2) and statistical significance (*p*-value) were calculated using GraphPad Prism (GraphPad). For statistical significance, a *p*-value of less than 0.05 and an R^2 value of larger than 0.95 were used.

Pyrosequencing

The methylation profiles of *APEX1*, *BRCA1*, *DDB2*, *FAN1*, and *OGG1* were assessed with pyrosequencing. Isolated DNA was bisulfite-treated with EpiTect Fast DNA

Bisulfite Kit (Qiagen). Bisulfite-treated DNA samples were assessed on the PyroMark Q48 System using Pyro-Mark CpG primers (Qiagen) described in Supplementary Table 4.

Ethics approval

This study was approved by the Institutional Review Board (IRB) and the Biosafety Committee of Emory University.

Conclusion

Investigating CAG repeat instability in spermatogenesis is challenging due to biological differences among different animal models and the limited availability of human HD testicular tissues [1, 13, 38, 111–114]. We have developed an in vitro model to replicate CAG repeat instability utilizing patient-derived human pluripotent stem cells to address such challenges and provide unique insights. This proof-of-concept study demonstrates an in vitro system to study TNR instability in human spermatogenesis. In this study, we demonstrate CAG repeat expansion in the in vitro spermatogenesis system and suppressed expansion of TNR during spermatogenesis by aspirin and araC treatment. However, we only have analyzed the gene expressions of a panel of selective genes and DNA methylation changes; thus, a further study is needed to address how these changes at the genomic level can impact at the protein level. Our study presents an alternative to study CAG repeat instability and provides a platform to evaluate novel therapeutics that can be used to stabilize or even induce contraction in HD spermatogenic cells.

Declarations

References

1. Budworth H, McMurray CT. A brief history of triplet repeat diseases. Methods Mol Biol. 2013;1010:3–17.
2. Orr HT, Zoghbi HY. Trinucleotide repeat disorders. Annu Rev Neurosci. 2007;30:575–621.
3. McMurray CT. Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet. 2010;11(11):786–99.
4. Bates GP, Dorsey R, Gusella JF, Hayden MR, Kay C, Leavitt BR, et al. Huntington disease. Nat Rev Dis Primers. 2015;1:15005.
5. Clever F, Cho IK, Yang J, Chan AWS. Progressive polyglu-tamine repeat expansion in peripheral blood cells and sperm of transgenic Huntington's disease monkeys. J Huntingtons Dis. 2019;8(4):443–8.
6. Telenius H, Kremer B, Goldberg YP, Theilmann J, Andrew SE, Zeisler J, et al. Somatic and gonadal mosaicism of the Hunting-ton disease gene CAG repeat in brain and sperm. Nat Genet. 1994;6(4):409–14.
7. Mouro Pinto R, Arning L, Giordano JV, Razghandi P, Andrew MA, Gillis T, et al. Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's dis-ease and in spinocerebellar ataxia type 1. Hum Mol Genet. 2020;29(15):2551–67.
8. Putkhao K, Kocerha J, Cho IK, Yang J, Parnpai R, Chan AW. Pathogenic cellular phenotypes are germline transmissible in a transgenic primate model of Huntington's disease. Stem Cells Dev. 2013;22(8):1198–205.
9. Kovtun IV, Therneau TM, McMurray CT. Gender of the embryo contributes to CAG instability in transgenic mice

containing a Huntington's disease gene. Hum Mol Genet. 2000;9(18):2767–75.

10. Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. Nat Genet. 1997;15(2):197–200.

11. Wheeler VC, Auerbach W, White JK, Srinidhi J, Auerbach A, Ryan A, et al. Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. Hum Mol Genet. 1999;8(1):115–22.

12. Kovtun IV, McMurray CT. Trinucleotide expansion in haploid germ cells by gap repair. Nat Genet. 2001;27(4):407–11.

13. Yoon SR, Dubeau L, de Young M, Wexler NS, Arnheim N. Huntington disease expansion mutations in humans can occur before meiosis is completed. Proc Natl Acad Sci U S A. 2003;100(15):8834–8.

14. Wheeler VC, Persichetti F, McNeil SM, Mysore JS, Mysore SS, MacDonald ME, et al. Factors associated with HD CAG repeat instability in Huntington disease. J Med Genet. 2007;44(11):695–701.

15. Xu H, Yang M, Tian R, Wang Y, Liu L, Zhu Z, et al. Derivation and propagation of spermatogonial stem cells from human pluripotent cells. Stem Cell Res Ther. 2020;11(1):408.

16. Easley CAt, Phillips BT, McGuire MM, Barringer JM, Valli H, Hermann BP, et al. Direct differentiation of human pluripotent stem cells into haploid spermatogenic cells. Cell Rep. 2012;2(3):440–6.

17. Sosa E, Chen D, Rojas EJ, Hennebold JD, Peters KA, Wu Z, et al. Differentiation of primate primordial germ cell-like cells following transplantation into the adult gonadal niche. Nat Commun. 2018;9(1):5339.

18. Greeson KW, Fowler KL, Estave PM, Kate Thompson S, Wagner C, Clayton Edenfield R, et al. Detrimental effects of flame retardant, PBB153, exposure on sperm and future generations. Sci Rep. 2020;10(1):8567.

19. Steves AN, Turry A, Gill B, Clarkson-Townsend D, Bradner JM, Bachli I, et al. Per- and polyfluoroalkyl substances impact human spermatogenesis in a stem-cell-derived model. Syst Biol Reprod Med. 2018;64(4):225–39.

20. Steves AN, Bradner JM, Fowler KL, Clarkson-Townsend D, Gill BJ, Turry AC, et al. Ubiquitous flame-retardant toxicants impair spermatogenesis in a human stem cell model. iScience. 2018;3:161–76.

21. Easley CAt, Bradner JM, Moser A, Rickman CA, McEachin ZT, Merritt MM, et al. Assessing reproductive toxicity of two environmental toxicants with a novel in vitro human spermatogenic model. Stem Cell Res. 2015;14(3):347–55.

22. Handel MA, Eppig JJ, Schimenti JC. Applying "gold standards" to in-vitro-derived germ cells. Cell. 2014;157(6):1257–61.

23. Khampang S, Cho IK, Punyawai K, Gill B, Langmo JN, Nath S, et al. Blastocyst development after fertilization with in vitro spermatids derived from non-human primate embryonicstem cells. F&S Science. 2021. https://doi.org/10.1016/j.xfss.2021.09.001.

24. Khampang S, Parnpai R, Mahikul W, Easley CA, Cho IK, Chan AWS. CAG repeat instability in embryonic stem cells and derivative spermatogenic cells of transgenic Huntington's disease monkey. J Assist Reprod Genet. 2021. https://doi.org/10.1007/s10815-021-02106-3.

25. Chatterjee N, Lin Y, Santillan BA, Yotnda P, Wilson JH. Environmental stress induces trinucleotide repeat mutagenesis in human cells. Proc Natl Acad Sci U S A. 2015;112(12):3764–9.

26. Dowsing AT, Yong EL, Clark M, McLachlan RI, de Kretser DM, Trounson AO. Linkage between male infertility and trinucleotide repeat expansion in the androgen-receptor gene. Lancet. 1999;354(9179):640–3.

27. Gomes-Pereira M, Monckton DG. Chemically induced increases and decreases in the rate of expansion of a CAG*CTG triplet repeat. Nucleic Acids Res. 2004;32(9):2865–72.

28. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics. 2014;13(2):397–406.

29. Bajrami E, Spiroski M. Genomic Imprinting. Open Access Maced J Med Sci. 2016;4(1):181–4.

30. Tunc O, Tremellen K. Oxidative DNA damage impairs global sperm DNA methylation in infertile men. J Assist Reprod Genet. 2009;26(9–10):537–44.

31. Cui X, Jing X, Wu X, Yan M, Li Q, Shen Y, et al. DNA methylation in spermatogenesis and male infertility. Exp Ther Med. 2016;12(4):1973–9.

32. Lombo M, Fernandez-Diez C, Gonzalez-Rojo S, Herraez MP. Genetic and epigenetic alterations induced by bisphenol A exposure during different periods of spermatogenesis: from spermatozoa to the progeny. Sci Rep. 2019;9(1):18029.

33. Goold R, Flower M, Moss DH, Medway C, Wood-Kaczmar A, Andre R, et al. FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat. Hum Mol Genet. 2019;28(4):650–61.

34. Kovtun IV, Liu Y, Bjoras M, Klungland A, Wilson SH, McMurray CT. OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. Nature. 2007;447(7143):447–52.

35. Genetic Modifiers of Huntington's Disease C. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell. 2015;162(3):516–26.

36. Kim KH, Hong EP, Shin JW, Chao MJ, Loupe J, Gillis T, et al. Genetic and functional analyses point to FAN1 as the source of multiple Huntington disease modifier effects. Am J Hum Genet. 2020;107(1):96–110.

37. Ehmcke J, Wistuba J, Schlatt S. Spermatogonial stem cells: questions, models and perspectives. Hum Reprod Update. 2006;12(3):275–82.

38. Simard O, Gregoire MC, Arguin M, Brazeau MA, Leduc F, Marois I, et al. Instability of trinucleotidic repeats during chromatin remodeling in spermatids. Hum Mutat. 2014;35(11):1280–4.

39. Usdin K, House NC, Freudenreich CH. Repeat instability during DNA repair: Insights from model systems. Crit Rev Biochem Mol Biol. 2015;50(2):142–67.

40. Neto JL, Lee JM, Afridi A, Gillis T, Guide JR, Dempsey S, et al. Genetic contributors to intergenerational CAG repeat instability in Huntington's disease knock-in mice. Genetics. 2017;205(2):503–16.

41. Cannella M, Maglione V, Martino T, Ragona G, Frati L, Li GM, et al. DNA instability in replicating Huntington's disease lymphoblasts. BMC Med Genet. 2009;10:11.

42. MacDonald ME, Barnes G, Srinidhi J, Duyao MP, Ambrose CM, Myers RH, et al. Gametic but not somatic instability of CAG repeat length in Huntington's disease. J Med Genet. 1993;30(12):982–6.

43. Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SA, et al. De novo expansion of a (CAG)n repeat in sporadic Huntington's disease. Nat Genet. 1993;5(2):168–73.

44. Ranen NG, Stine OC, Abbott MH, Sherr M, Codori AM, Franz ML, et al. Anticipation and instability of IT-15 (CAG)n repeats in parent-offspring pairs with Huntington disease. Am J Hum Genet. 1995;57(3):593–602.

45. Wright GEB, Black HF, Collins JA, Gall-Duncan T, Caron NS, Pearson CE, et al. Interrupting sequence variants and age of onset in Huntington's disease: clinical implications and emerging therapies. Lancet Neurol. 2020;19(11):930–9.

46. van der Kemp PA, Thomas D, Barbey R, de Oliveira R, Boiteux S. Cloning and expression in Escherichia coli of the OGG1 gene of Saccharomyces cerevisiae, which codes for a DNA glycosylase that excises 7,8-dihydro-8-oxoguanine and 2,6-diamino-4-hydroxy-5-N-methylformamidopyrimidine. Proc Natl Acad Sci U S A. 1996;93(11):5197–202.

47. Kovtun IV, Johnson KO, McMurray CT. Cockayne syndrome B protein antagonizes OGG1 in modulating CAG repeat length in vivo. Aging (Albany NY). 2011;3(5):509–14.

48. Xu M, Lai Y, Torner J, Zhang Y, Zhang Z, Liu Y. Base excision repair of oxidative DNA damage coupled with removal of a CAG repeat hairpin attenuates trinucleotide repeat expansion. Nucleic Acids Res. 2014;42(6):3675–91.

49. Marenstein DR, Wilson DM 3rd, Teebor GW. Human AP endonuclease (APE1) demonstrates endonucleolytic activity against AP sites in single-stranded DNA. DNA Repair (Amst). 2004;3(5):527–33.

50. Hill JW, Hazra TK, Izumi T, Mitra S. Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair. Nucleic Acids Res. 2001;29(2):430–8.

51. Vidal AE, Hickson ID, Boiteux S, Radicella JP. Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: bypass of the AP lyase activity step. Nucleic Acids Res. 2001;29(6):1285–92.

52. Dianova II, Bohr VA, Dianov GL. Interaction of human AP endonuclease 1 with flap endonuclease 1 and proliferating cell nuclear antigen involved in long-patch base excision repair. Biochemistry. 2001;40(42):12639–44.

53. Beaver JM, Lai Y, Xu M, Casin AH, Laverde EE, Liu Y. AP endonuclease 1 prevents trinucleotide repeat expansion via a novel mechanism during base excision repair. Nucleic Acids Res. 2015;43(12):5948–60.

54. Xu M, Gabison J, Liu Y. Trinucleotide repeat deletion via a unique hairpin bypass by DNA polymerase beta and alternate flap cleavage by flap endonuclease 1. Nucleic Acids Res. 2013;41(3):1684–97.

55. Liu Y, Prasad R, Beard WA, Hou EW, Horton JK, McMurray CT, et al. Coordination between polymerase beta and FEN1 can modulate CAG repeat expansion. J Biol Chem. 2009;284(41):28352–66.

56. Liu Y, Wilson SH. DNA base excision repair: a mechanism of trinucleotide repeat expansion. Trends Biochem Sci. 2012;37(4):162–72.

57. Lai Y, Xu M, Zhang Z, Liu Y. Instability of CTG repeats is governed by the position of a DNA base lesion through base excision repair. PLoS ONE. 2013;8(2):e56960.

58. Lee J-M, Correia K, Loupe J, Kim K-H, Barker D, Hong EP, et al. CAG repeat not polyglutamine length determines timing of Huntington's disease onset. Cell. 2019;178(4):887-900.e14.

59. Zhao XN, Usdin K. FAN1 protects against repeat expansions in a fragile X mouse model. DNA Repair (Amst). 2018;69:1–5.

60. Flower M, Lomeikaite V, Ciosi M, Cumming S, Morales F, Lo K, et al. MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. Brain. 2019. https://doi.org/10.1093/brain/awz115.

61. Lai Y, Budworth H, Beaver JM, Chan NL, Zhang Z, McMurray CT, et al. Crosstalk between MSH2-MSH3 and polbeta promotes trinucleotide repeat expansion during base excision repair. Nat Commun. 2016;7:12465.

62. Kantartzis A, Williams GM, Balakrishnan L, Roberts RL, Surtees JA, Bambara RA. Msh2-Msh3 interferes with Okazaki fragment processing to promote trinucleotide repeat expansions. Cell Rep. 2012;2(2):216–22.

63. van den Broek WJ, Nelen MR, Wansink DG, Coerwinkel MM, te Riele H, Groenen PJ, et al. Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. Hum Mol Genet. 2002;11(2):191–8.

64. Dragileva E, Hendricks A, Teed A, Gillis T, Lopez ET, Friedberg EC, et al. Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiol Dis. 2009;33(1):37–47.

65. Foiry L, Dong L, Savouret C, Hubert L, te Riele H, Junien C, et al. Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. Hum Genet. 2006;119(5):520–6.

66. Mollersen L, Rowe AD, Illuzzi JL, Hildrestrand GA, Gerhold KJ, Tveteras L, et al. Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. Hum Mol Genet. 2012;21(22):4939–47.

67. Guo J, Gu L, Leffak M, Li GM. MutSbeta promotes trinucleotide repeat expansion by recruiting DNA polymerase beta to nascent (CAG)n or (CTG)n hairpins for error-prone DNA synthesis. Cell Res. 2016;26(7):775–86.

68. Goel A, Chang DK, Ricciardiello L, Gasche C, Boland CR. A novel mechanism for aspirin-mediated growth inhibition of human colon cancer cells. Clin Cancer Res. 2003;9(1):383–90.

69. Awtry EH, Loscalzo J. Aspirin. Circulation. 2000;101(10):1206–18.

70. Smith TB, Dun MD, Smith ND, Curry BJ, Connaughton HS, Aitken RJ. The presence of a truncated base excision repair pathway in human spermatozoa that is mediated by OGG1. J Cell Sci. 2013;126(Pt 6):1488–97.

71. Aitken RJ, Smith TB, Jobling MS, Baker MA, De Iuliis GN. Oxidative stress and male reproductive health. Asian J Androl. 2014;16(1):31–8.

72. Kothari S, Thompson A, Agarwal A, du Plessis SS. Free radicals: their beneficial and detrimental effects on sperm function. Indian J Exp Biol. 2010;48(5):425–35.

73. Grant S. Ara-C: cellular and molecular pharmacology. Adv Cancer Res. 1998;72:197–233.

74. Longley MJ, Pierce AJ, Modrich P. DNA polymerase delta is required for human mismatch repair in vitro. J Biol Chem. 1997;272(16):10917–21.

75. Rechkoblit O, Johnson RE, Buku A, Prakash L, Prakash S, Aggarwal AK. Structural insights into mutagenicity of anticancer nucleoside analog cytarabine during replication by DNA polymerase eta. Sci Rep. 2019;9(1):16400.

76. Ishikura H, Sawada H, Okazaki T, Mochizuki T, Izumi Y, Yamagishi M, et al. The effect of low dose Ara-C in acute nonlymphoblastic leukaemias and atypical leukaemia. Br J Haematol. 1984;58(1):9–18.

77. Kuznetsova AA, Kuznetsov NA, Ishchenko AA, Saparbaev MK, Fedorova OS. Pre-steady-state fluorescence analysis of damaged DNA transfer from human DNA glycosylases to AP endonuclease APE1. Biochim Biophys Acta. 2014;1840(10):3042–51.

78. Loupe JM, Pinto RM, Kim KH, Gillis T, Mysore JS, Andrew MA, et al. Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by Mlh1 knock-out. Hum Mol Genet. 2020;29(18):3044–53.

79. Savouret C, Brisson E, Essers J, Kanaar R, Pastink A, te Riele H, et al. CTG repeat instability and size variation timing in DNA repair-deficient mice. EMBO J. 2003;22(9):2264–73.

80. Spiro C, McMurray CT. Nuclease-deficient FEN-1 blocks Rad51/BRCA1-mediated repair and causes trinucleotide repeat instability. Mol Cell Biol. 2003;23(17):6063–74.

81. Dion V, Lin Y, Hubert L Jr, Waterland RA, Wilson JH. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. Hum Mol Genet. 2008;17(9):1306–17.

82. Jung J, Bonini N. CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. Science. 2007;315(5820):1857–9.

83. Massey TH, Jones L. The central role of DNA damage and repair in CAG repeat diseases. Dis Model Mech. 2018;11(1). https://doi.org/10.1242/dmm.031930.

84. Jones L, Houlden H, Tabrizi SJ. DNA repair in the trinucleotide repeat disorders. Lancet Neurol. 2017;16(1):88–96.

85. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol. 2016;79(6):983–90.

86. Maiuri T, Suart CE, Hung CLK, Graham KJ, Barba Bazan CA, Truant R. DNA damage repair in Huntington's disease and other neurodegenerative diseases. Neurotherapeutics. 2019;16(4):948–56.

87. Donaldson J, Powell S, Rickards N, Holmans P, Jones L. What is the pathogenic CAG expansion length in Huntington's disease? J Huntingtons Dis. 2021;10(1):175–202.

88. De Rycke M, Berckmoes V. Preimplantation genetic testing for monogenic disorders. Genes (Basel). 2020;11(8). https://doi.org/10.3390/genes11080871.

89. Sermon K, De Rijcke M, Lissens W, De Vos A, Platteau P, Bonduelle M, et al. Preimplantation genetic diagnosis for Huntington's disease with exclusion testing. Eur J Hum Genet. 2002;10(10):591–8.

90. Rahman MS, Kwon WS, Lee JS, Yoon SJ, Ryu BY, Pang MG. Bisphenol-A affects male fertility via fertility-related proteins in spermatozoa. Sci Rep. 2015;5:9169.

91. Meeker JD, Ehrlich S, Toth TL, Wright DL, Calafat AM, Trisini AT, et al. Semen quality and sperm DNA damage in relation to urinary bisphenol A among men from an infertility clinic. Reprod Toxicol. 2010;30(4):532–9.

92. Padmanabhan V, Siefert K, Ransom S, Johnson T, Pinkerton J, Anderson L, et al. Maternal bisphenol-A levels at delivery: a looming problem? J Perinatol. 2008;28(4):258–63.

93. Calafat AM, Ye X, Wong LY, Reidy JA, Needham LL. Exposure of the U.S. population to bisphenol A and 4-tertiary-octylphenol: 2003–2004. Environ Health Perspect. 2008;116(1):39–44.

94. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 2016;44(W1):W3–10.

95. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

96. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.

97. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

98. Alhamdoosh M, Ng M, Wilson NJ, Sheridan JM, Huynh H, Wilson MJ, et al. Combining multiple tools outperforms individual methods in gene set enrichment analyses. Bioinformatics. 2017;33(3):414–24.

99. Mollersen L, Rowe AD, Larsen E, Rognes T, Klungland A. Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice. PLoS Genet. 2010;6(12):e1001242.

100. Zhao XN, Usdin K. Timing of expansion of fragile X premutation alleles during intergenerational transmission in a mouse model of the fragile X-related disorders. Front Genet. 2018;9:314.

101. Zhao X, Zhang Y, Wilkins K, Edelmann W, Usdin K. MutL-gamma promotes repeat expansion in a fragile X mouse model while EXO1 is protective. PLoS Genet. 2018;14(10):e1007719.

102. Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, et al. Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell. 2009;35(1):82–92.

103. Khristich AN, Armenia JF, Matera RM, Kolchinski AA, Mirkin SM. Large-scale contractions of Friedreich's ataxia GAA repeats in yeast occur during DNA replication due to their triplex-forming ability. Proc Natl Acad Sci U S A. 2020;117(3):1628–37.

104. Callahan JL, Andrews KJ, Zakian VA, Freudenreich CH. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. Mol Cell Biol. 2003;23(21):7849–60.

105. Ditch S, Sammarco MC, Banerjee A, Grabczyk E. Progressive GAA.TTC repeat expansion in human cell lines. PLoS Genet. 2009;5(10):e1000704.

106. Du J, Campau E, Soragni E, Jespersen C, Gottesfeld JM. Length-dependent CTG.CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. Hum Mol Genet. 2013;22(25):5276–87.

107. Gomes-Pereira M, Fortune MT, Monckton DG. Mouse tissue culture models of unstable triplet repeats: in vitro selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates. Hum Mol Genet. 2001;10(8):845–54.

108. Claassen DA, Lahue RS. Expansions of CAG.CTG repeats in immortalized human astrocytes. Hum Mol Genet. 2007;16(24):3088–96.

109. Lee JM, Zhang J, Su AI, Walker JR, Wiltshire T, Kang K, et al. A novel approach to investigate tissue-specific trinucleotide repeat instability. BMC Syst Biol. 2010;4:29.

110. Lee JM, Zhang J, Su AI, Walker JR, Wiltshire T, Kang K, et al. A novel approach to investigate tissue-specific trinucleotide repeat instability. BMC Syst Biol. 2010;4(29):1–16.

111. Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, Roloff TC, et al. Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. Nat Struct Mol Biol. 2010;17(6):679–87.

112. Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR. Distinctive chromatin in human sperm packages genes for embryo development. Nature. 2009;460(7254):473–8.

113. Kovtun IV, Goellner G, McMurray CT. Structural features of trinucleotide repeats associated with DNA expansion. Biochem Cell Biol. 2001;79(3):325–36.

114. Siklenka K, Erkek S, Godmann M, Lambrot R, McGraw S, Lafleur C, et al. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. Science. 2015;350(6261):aab2006.